Imaging Biomarkers

Development and Clinical Integration

Luis Martí-Bonmatí
Angel Alberich-Bayarri
Editors

Springer

Imaging Biomarkers

Luis Martí-Bonmatí
Angel Alberich-Bayarri
Editors

Imaging Biomarkers

Development and Clinical Integration

Springer

Editors
Luis Martí-Bonmatí
La Fe University
and Polithecnic Hospital
University of Valencia
Valencia
Spain

Angel Alberich-Bayarri
La Fe University
and Polithecnic Hospital
University of Valencia
Valencia
Spain

ISBN 978-3-319-43502-2 ISBN 978-3-319-43504-6 (eBook)
DOI 10.1007/978-3-319-43504-6

Library of Congress Control Number: 2016957713

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland
The registered company is Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book started so many years ago. On those times, most researchers on medical imaging were involved in how to obtain the best images to achieve accurate diagnosis in clinical medicine. Detecting lesions and knowing the nature of what was found and how these findings impact treatment were the hallmarks of radiology for years. Our contribution was to initiate quantitative approaches to improve the way radiology was performed. In some way, measurements were not usual in the radiology departments, and, even, they were considered an inappropriate way to describe lesions when compared with the art of reporting images.

In this book on imaging biomarkers, we tried to describe how quantitative imaging changes the way we should interact with other physicians and patients from the first idea to the final clinical application. Imaging biomarkers might generate spatially and time-resolved in vivo maps of the distribution of some relevant disease hallmarks in individual patients.

The different chapters will deal with the shift in paradigm from descriptive signs and radiological findings toward a precision medicine where imaging plays a major role in the detection, grading, and localization of tissue abnormalities. The different international initiatives for the promotion of imaging biomarkers will be presented, together with the definition of their stepwise pipeline development. All the different parts to be considered when dealing with biomarkers will be precisely defined, starting from the biological basis and clinical question to be answered through the acquisition modalities and protocols definition, image preparation, and analysis. Those aspects related to how and where the measurements should be obtained, their main biases and sources of uncertainty, the reproducibility of the results and their validation, and the appropriate reporting to efficiently communicate the results are key aspects of this monograph. Ultimately, main clinical indications, pitfalls, and big data aspects of imaging biomarkers will be also presented. Some examples of imaging biomarkers in neurodegenerative diseases, brain and breast tumors, lung and cardiac diseases, and musculoskeletal and diffuse liver disorders will highlight their clinical relevance.

We hope that this text will be helpful for readers with very different backgrounds, who need to quantify and analyze imaging data to critically answer interesting clinical questions. Although this book is not intended to be a manual, it concentrates ideas and concepts relevant to most radiologists and biomedical engineers looking for fundamental concepts in radiomics and virtual biopsies. It is our understanding that in silico computational models obtained

from medical images and patient-specific processing have a great promise and expectation toward more effective and enduring disease ontology, signatures, and therapies.

We are indebted to many people who allowed this book to come into existence, mainly the authors and researchers whom we interacted with all over these years, and of course our families, who allowed us to spend time and energy in this beautiful and precise project.

Valencia, Spain
July 2016

Luis Martí-Bonmatí
Angel Alberich-Bayarri

Contents

1 The Shift in Paradigm to Precision Medicine in Imaging: International Initiatives for the Promotion of Imaging Biomarkers

Siegfried Trattnig

1.1 Background

Qualitative, subjective interpretations of medical images, which have been the backbone of image interpretation for the past century, provide useful information to the treating physician. This procedure will continue for some time into the future, since the alternative method of automatic reading of images by artificial intelligence is not developed enough to replace the trained and experienced observer with his/her ability to interpret and judge during image reading sessions. Nevertheless, subjective and, therefore, qualitative interpretations are observer dependent and highly variable, and variability inevitably degrades outcomes in healthcare in general. Extracting objective, quantitative results from medical images is one way to reduce the variability associated with medical image interpretation and, thus, obtain relevant and useful quantitative results, which will improve patient outcomes.

During the past two decades, advances in medical imaging technology have offered the possibility to extract high-resolution anatomic, physiologic, functional, biochemical, and metabolic information from clinical images, all of which reflect the molecular composition of the healthy or diseased tissue of organs imaged in the human body. With appropriate calibration, most of these imaging technologies can provide quantitative information about specific properties of the tissues being imaged.

The *computed tomography (CT)* signal is linearly proportional to atomic density and has high spatial resolution. It can provide accurate distance measurements (e.g., tumor dimensions or volumetry [1]), basic tissue characterization, quantitative physiologic information related to perfusion [2, 3] or necrosis [4], angiographic information, and quantitative functional information using dynamic contrast techniques.

The *positron emission tomography (PET)* signal is linearly proportional to atomic decay events and has high sensitivity. If the physiology of the labeled substance being detected is known, it can be related to the PET signal magnitude of the molecular substrates of interest. [5] PET radiopharmaceuticals can be developed to assay a wide variety of physiologic or molecular functions. CT provides attenuation correction data for PET imaging with PET/CT, although contrast-enhanced CT will also give diagnostic information [6].

Magnetic resonance imaging (MRI) signals are complex: they are quantitatively, but nonlinearly, related to T1 and T2 relaxation phenomena, as well as proton density distribution and/or other physiologic and functional features, such as

S. Trattnig
Department of Biomedical Imaging and Image-guided Therapy, Center of Excellence for High Field MR, Vienna, Austria
e-mail: siegfried.trattnig@meduniwien.ac.at

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_1

flow, permeability, diffusion, and cell density, among others. Nevertheless, with appropriate calibration and standardization, MRI techniques can be devised in which the signal response is a quantitative indicator of tissue structure and function [7–11].

Ultrasound data are also complex, but, again, with appropriate calibration and standardization, quantitative information about attenuation, refraction, reflection, bulk tissue properties, and shear wave speed of the tissues being imaged can be calculated. In addition, quantitative distance, elasticity, and Doppler flow measurements can be extracted from ultrasound signals [12].

Therefore, these imaging methods can be considered useful as imaging biomarkers and are comparable to biomarkers from laboratory assays. The term "quantitative imaging" has recently been formally defined as "the extraction of quantifiable features from medical images for the assessment of normality, or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal" (QIBA).

Fundamentally, quantitative imaging involves the measurement of some variables using medical images, and, most meaningfully, quantitative imaging is based on an underlying biophysical model for the tissue of interest. It involves the measurement of image-related properties that are analyzed according to that model. This results in an unbiased, quantitative estimate of a physical or biological parameter [13].

In general, a biomarker represents a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes [14], or biological responses to a therapeutic intervention. Measurements of anatomical, physiological, functional, biochemical, and metabolic characteristics of the body through medical imaging are referred to as imaging biomarkers and have gained increasing attention from the medical community in the last several years. Imaging biomarkers are becoming increasingly used in clinical research for drug development and clinical decision-making [15].

Imaging biomarkers include the development, standardization, and optimization of imaging acquisition protocols, data analyses, display methods, and reporting structures. These are requirements for the validation of accurately obtained image-derived metrics with anatomically and physiologically relevant parameters in different oncologic, as well as non-oncologic diseases, in different organs of the human body. These requirements are also necessary for the monitoring of treatment response. Therefore, the use of such metrics will play a crucial ever-increasing role in research and patient care [2].

The molecular bases of health and disease have become increasingly well understood in the past 20 years. In fact, the molecular characterization of disease has revealed that each patient is likely to have a unique combination of genotypic and phenotypic profiles for a specific disease. Healthcare delivery is now focused on trying to determine the most appropriate therapy for any patient's molecularly unique version of a particular disease [16]. This concept is referred to as *precision medicine*, and it has recently been identified as a national priority in the USA, with a large initiative to explore and implement such practices that would result in the need for tests that can provide objective, reproducible information for clinical research and practice [17].

Biomarkers in general—both specimen and imaging—play an increasingly important role in healthcare by providing more information about molecular processes. Specimen biomarkers are typically obtained from body fluids or a tissue sample (biopsy). The main disadvantage, however, is that, in most cases, the sample represents only a small portion of the normal or diseased tissue of interest, and, therefore, spatial sampling bias is a serious problem. Imaging scans, on the other hand, typically cover a broad segment of tissue, or even the entire patient, and potentially provide more comprehensive information about heterogeneity. Another disadvantage of specimen biomarkers from biopsies is that they represent a single time point, whereas imaging data can be obtained dynamically and longitudinally in the individual patient without the need for repeated invasive procedures. For these reasons, imaging biomarkers are of considerable interest in evidence-based clinical decision-making, as well

as for therapeutic development and treatment monitoring. Whereas for in vitro laboratory specimen assays, standard terminology and methods have become established to describe, evaluate, and validate assays for medical applications for many years, the same approach to standardization for imaging data has only begun to occur in an organized way in the past few years.

It is expected that the added value of objective, quantitative imaging biomarkers in both research and clinical applications will increase as healthcare initiatives place increased pressure on imaging physicians to provide evidence-based decisions and care. The demand for quantitative results from imaging studies will increase as treatment decisions are driven by such results. Fortunately, contemporary computing power and image processing methods are such that quantitative results can be obtained from current digital medical images at little or no increased cost, which will allow efficient integration of quantitative imaging into the radiologist's workflow as an added value, with minimal or no increased cost to the healthcare system.

However, one of the basic requirements for establishing the link between quantitative imaging results and improved patient outcomes is the availability of robust, reliable, accurate, reproducible, and validated quantitative results from clinical trials. There are still substantial hurdles to the achievement of reproducible quantitative imaging measures in clinical practice, but particularly in clinical trials. There are several sources of bias and variance in the quantitative results obtained from clinical images, which can be grouped into three main categories: the image acquisition hardware, software, and procedures, the measurement methods used, and the reader variability. Thus, the entire chain involved in producing a clinical image must be approached from a systematic perspective. Considerable work is needed to validate specific metrics, improve standardization across vendor platforms, and educate imaging physicians about the reliability of quantitative imaging. Validating an imaging biomarker requires identifying and characterizing all of the sources of bias and variance that could affect the end measurement.

The process of acquiring a clinical imaging scan is complex; therefore, the goals of standardization and improved reproducibility require coordinated activity among imaging device and software manufacturers, regulatory organizations, healthcare providers, academic institutions, groups using imaging in clinical trials, and professional societies. An additional challenge in the efforts to improve reproducibility in quantitative clinical imaging is the continual technological advancement that occurs in medical imaging hardware and software. Thus, the quantitative accuracy of medical images must be continually reassessed, and the standardization requirements have to be periodically updated.

Examples for such a collaborative activity that attempts to address these issues are the Quantitative Imaging Biomarkers Alliance (QIBA) in the USA and the European Imaging Biomarkers Alliance (EIBALL) in Europe.

1.2 Quantitative Imaging Biomarkers Alliance (QIBA)

The *Quantitative Imaging Biomarkers Alliance (QIBA)* was organized in 2007 by the RSNA to unite researchers, healthcare professionals, and industry stakeholders in the advancement of quantitative imaging [18]. QIBA draws from the very successful precedents set by the DICOM and Integrating the Healthcare Enterprise (IHE) efforts, but has been adapted to the needs of imaging science. Strategic guidance supporting the development, qualification, and deployment of imaging biomarkers will lead to improved standardization of imaging tests, proof of imaging test performance, and greater use of imaging to predict tissue biologic behavior and to monitor therapy response.

QIBA's mission is to improve the value and practicality of quantitative imaging biomarkers by reducing variability across devices, patients, and time. The QIBA initiative involves (1) stakeholder collaboration to identify needs, barriers, and solutions to develop and test consistent, reliable, valid, and achievable quantitative imaging results across imaging platforms,

clinical sites, and time and (2) accelerating the development and adoption of hardware and software standards needed to achieve accurate and reproducible quantitative results from imaging methods.

The forum created by QIBA for an organized and effective cooperative effort among key participants has advanced through the generous efforts of volunteer members from academia, the medical device, pharmaceutical and other business sectors, and government. QIBA exemplifies a collaborative Model for Partnership and Leadership. The QIBA structure has been developing and evolving over the past seven years and is now widely recognized and respected by industry, academia, and government agencies. This structure is having a positive impact on imaging in clinical trials and clinical care.

QIBA participants span a wide range of expertise, including, but not limited to, clinical practice, clinical research, physics, engineering, statistics, marketing, senior management, regulatory, pharmaceutical, and computer science. The structure of QIBA explicitly includes the imaging device industry, which allows for precompetitive cooperation across all vendors to achieve standardization of quantitative outputs.

From its inception, QIBA established communication with members of the FDA, NIH, and NIST at several levels, and these essential interactions continue. FDA and NIST staff scientists participate in QIBA Committees and other working groups, such as the QIBA Metrology Working Group, ex officio representation on the QIBA Steering Committee, and review QIBA documents for comment.

A cornerstone of the QIBA methodology is to produce a description of a quantitative imaging biomarker in sufficient detail that it can be considered a validated assay, which means that the measurement bias and variability are both characterized and minimized. This is accomplished through the use of the so-called QIBA Profile, which is a document intended for a broad audience, including scanner and third-party device manufacturers (e.g., display stations), pharmaceutical companies, diagnostic agent manufacturers, medical imaging sites, imaging contract research organizations, physicians, technologists, researchers, professional organizations, and accreditation and regulatory authorities.

A QIBA Profile comprises the description of the intended use of, or clinical context for, the quantitative imaging biomarker, a "Claim" of the achievable minimum variability and/or bias, and a description of the image acquisition protocol needed to meet the Claim. Finally, a description of compliance items is needed to meet the Claim.

In a QIBA Profile, the Claim is the central result and describes the quantitative imaging biomarker as a standardized, reproducible assay in terms of technical performance. The Claim is based on peer-reviewed results as much as possible and also represents a consensus opinion by recognized experts in the imaging modality. For example, the QIBA fluorodeoxyglucose (FDG)-PET/CT profile was based on nine original research studies, one meta-analysis, and two multicenter studies that were in the process of being submitted for publication, as well as review by over 100 experts. The draft Profile was 90 pages long and was widely advertised for public comment. Over 100 comments and suggestions were received during the public comment phase, and each comment was resolved by a consensus review, and all comments and resolutions are publicly available, as well as the "Publicly Reviewed" version of the Profile.

QIBA Profiles and test objects are now being referenced as standards by industry and academic investigators and are being implemented in clinical trials. Internationally, QIBA-related activities are being implemented or developed in Europe (ESR, EORTC, EANM), Canada, Japan (JSR), Korea, and Brazil.

There has been a considerable increase in the number of QIBA committees and ad hoc task groups in the past couple of years. The current organizational structure is comprised of a steering committee, coordinating committees, biomarker committees, and task forces.

Recent and current QIBA activities are listed on the QIBA homepage [18].

1.3 European Imaging Biomarkers Alliance (EIBALL)

The European Imaging Biomarkers Alliance (EIBALL) was officially introduced at the ECR conference in March 2015, and its implementation was based on a brainstorming meeting of the QIBA European Task Force on January 13, 2015.

Within the ESR Research Committee, a reorganization was necessary, since there was a need for coordination of all activities in the field of imaging biomarkers, a need to restructure existing resources and bodies, and a need to increase collaboration with QIBA™ and the European Organisation for Research and Treatment of Cancer (EORTC).

The activities of the ESR Subcommittee on Imaging Biomarkers, the ESR Working Group on Personalized Medicine, and the ESR-EORTC Working Group were *merged* into the *European Imaging Biomarkers Alliance (EIBALL).*

The tasks of EIBALL comprise the preparation of an inventory of all activities and projects in the field of quantitative imaging biomarkers where ESR is involved. It should serve as a toolbox that provides information about current projects, as well as the centers involved in clinical trials and educational activities related to Imaging Biomarkers. EIBALL should enhance and coordinate collaborations with EORTC and QIBA and foster the development of quantitative imaging biomarker with the support of the European Society of Molecular and Functional Imaging in Radiology (ESMOFIR) and the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB).

1.4 Organizational Structure and Aspects of EIBALL

The operational work of the alliance should become the responsibility of the European Institute for Biomedical Imaging Research (EIBIR).

The EIBIR representative in the alliance is responsible for reporting about European projects and calls. EIBIR should initiate and establish a network of European Centers of Excellence (core institutions) for the implementation and clinical validation of imaging biomarkers on a European level. ESMOFIR and ESMRMB are responsible for the educational and scientific activities of EIBALL.

All ESR members of the original ESR-EORTC Working Group will create a permanent collaboration link between the two societies.

Three task-specific groups for the collaboration with EORTC, the collaboration with QIBA, and the development of quantitative imaging biomarkers have been formed and should develop their activities.

The chair of the alliance became a member of the QIBA Steering Committee, and QIBA appointed their chair as the QIBA representative to the alliance.

The involvement of further societies, such as the European Society of Oncologic Imaging (ESOI) and the European Society for Hybrid Medical Imaging (ESHI), and/or of nuclear medicine representatives will be discussed in a second step.

1.5 Collaboration of EIBALL with the European Organisation for Research and Treatment of Cancer (EORTC)

EORTC performs many clinical multicenter trials per year, has 150,000 patients in its database, with 50,000 in the follow-up stage, and comprises 2000 collaborators. Organizationally, EORTC is subdivided into 18 disease-oriented groups with their own respective steering committees in which multicenter trials are planned and performed.

The *Imaging Group* (comprising radiologists, nuclear medical physicians, physicists, and imaging scientists) is part of the "Translational Research and Imaging Department" within EORTC and is responsible for imaging protocols in multicenter trials, in collaboration with the disease-specific groups and their steering committees in EORTC.

1.6 The aims of EIBALL in collaboration with EORTC are the following:

Imaging protocols that are acceptable for a number of sites should be established, and a program of quality assurance (QA)/quality control (QC) should be implemented prior to imaging in multicenter trials.

In planned multicenter projects, potential sites should be recruited based on the list of core institutions that were defined in an EIBIR-based survey, which will help to establish a network of European Centers of Excellence (core institutions) for the implementation and clinical validation of imaging biomarkers.

Imaging protocols, including quantitative imaging biomarkers, should be distributed to the members of EIBALL, and an expert consensus should be achieved, which then should be integrated within the EORTC multicenter studies.

1.6.1 EIBIR should provide organizational support for this concept.

All EIBALL members of the original ESR-EORTC Working Group should create a permanent collaboration link between the two societies and foster the implementation of quantitative imaging biomarker protocols into the EORTC multicenter trials.

Collaboration with the Imaging Group of EORTC has to be strengthened, with the need to integrate more radiologists into the Imaging Group.

There is a need to integrate more imaging people, particularly radiologists, into the disease-oriented groups and their respective steering committees of EORTC where multicenter studies are developed and to which imaging protocols for clinical validation of imaging biomarkers should be added.

Training programs should be organized for radiologists, with state-of-the-art oncologic imaging, imaging interpretation, and integration of medical imaging and imaging biomarkers into clinical trials and translational research in imaging. For the training of young imagers in cancer research, three levels of training could be proposed: (1) training in imaging criteria commonly used in clinical trials, (2) training in advanced imaging criteria for clinical trials, and (3) training in the building of specific imaging-related clinical trials. The young researchers targeted for these courses would then become a potential network of imagers available and with the requested knowledge to participate in EORTC-organized clinical trials.

A program of quality assurance/quality control should be implemented prior to imaging in multicenter trials, based on a recent manuscript ("A risk management approach for imaging biomarker-driven clinical trials in oncology" jointly drafted by the EORTC Imaging Group, ESR, NCI, and EIBALL members) [19]. Different levels of QA/QC for imaging in clinical trials should be identified according to the complexity of image acquisition, processing, etc. Each clinical trial submitted to EORTC has a checklist, which does yet not include imaging, but this will be made mandatory.

1.7 Collaboration of EIBALL with QIBA™

The chair of the alliance became member of the QIBA Steering Committee, and QIBA has appointed the chair as the QIBA representative to the alliance recently to enhance collaboration between both alliances.

QIBA collaborates with EIBALL for field tests, which means clinical studies, to test the clinical validity of their technically developed quantitative imaging biomarker profiles via the large multicenter trials performed by EORTC. In the USA, the bureaucratic hurdles for clinical multicenter trials are much higher than in Europe.

In addition, QIBA has great interest in working together with EIBALL for joint quantitative imaging biomarker development, such as the planned contrast-enhanced US (CEUS) Projekt, since it is well aware that there are active and innovative groups of researchers in CEUS in Europe.

References

1. Mokry T, Bellemann N, Muller D, Lorenzo Bermejo J, Klauss M, Stampfl U, Radeleff B, Schemmer P, Kauczor HU, Sommer CM. Accuracy of estimation of graft size for living-related liver transplantation: first results of a semi-automated interactive software for CT-volumetry. PLoS One. 2014;9(10), e110201.
2. Klotz E, Haberland U, Glatting G, Schoenberg SO, Fink C, Attenberger U, Henzler T. Technical prerequisites and imaging protocols for CT perfusion imaging in oncology. Eur J Radiol. 2015;84(12):2359–67.
3. Prezzi D, Khan A, Goh V. Perfusion CT imaging of treatment response in oncology. Eur J Radiol. 2015;84(12):2380–5.
4. Thaiss WM, Sauter AW, Bongers M, Horger M, Nikolaou K. Clinical applications for dual energy CT versus dynamic contrast enhanced CT in oncology. Eur J Radiol. 2015;84(12):2368–79.
5. An YS, Kang DK, Jung YS, Han S, Kim TH. Tumor metabolism and perfusion ratio assessed by 18F-FDG PET/CT and DCE-MRI in breast cancer patients: correlation with tumor subtype and histologic prognostic factors. Eur J Radiol. 2015;84(7):1365–70.
6. Lu W, Wang J, Zhang HH. Computerized PET/CT image analysis in the evaluation of tumour response to therapy. Br J Radiol. 2015;88(1048):20140625.
7. Trattnig S, Welsch GH, Juras V, Szomolanyi P, Mayerhoefer ME, Stelzeneder D, Mamisch TC, Bieri O, Scheffler K, Zbyn S. 23Na MR imaging at 7 T after knee matrix-associated autologous chondrocyte transplantation preliminary results. Radiology. 2010;257(1):175–84.
8. Welsch GH, Apprich S, Zbyn S, Mamisch TC, Mlynarik V, Scheffler K, Bieri O, Trattnig S. Biochemical (T2, T2* and magnetisation transfer ratio) MRI of knee cartilage: feasibility at ultra-high field (7T) compared with high field (3T) strength. Eur Radiol. 2011;21(6):1136–43.
9. Guermazi A, Alizai H, Crema MD, Trattnig S, Regatte RR, Roemer FW. Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. Osteoarthritis Cartilage. 2015;23(10):1639–53.
10. Juras V, Winhofer Y, Szomolanyi P, Vosshenrich J, Hager B, Wolf P, Weber M, Luger A, Trattnig S. Multiparametric MR imaging depicts glycosaminoglycan change in the achilles tendon during ciprofloxacin administration in healthy men: initial observation. Radiology. 2015;275(3):763–71.
11. Valkovic L, Gajdosik M, Traussnigg S, Wolf P, Chmelik M, Kienbacher C, Bogner W, Krebs M, Trauner M, Trattnig S, Krssak M. Application of localized (3)(1) P MRS saturation transfer at 7 T for measurement of ATP metabolism in the liver: reproducibility and initial clinical application in patients with non-alcoholic fatty liver disease. Eur Radiol. 2014;24(7):1602–9.
12. Hudson JM, Williams R, Tremblay-Darveau C, Sheeran PS, Milot L, Bjarnason GA, Burns PN. Dynamic contrast enhanced ultrasound for therapy monitoring. Eur J Radiol. 2015;84(9):1650–7.
13. Carballido-Gamio J, Majumdar S. Atlas-based knee cartilage assessment. Magn Reson Med. 2011;66(2):574–83.
14. Eckstein F, Collins JE, Nevitt MC, et al. Brief Report: Cartilage Thickness Change as an Imaging Biomarker of Knee Osteoarthritis Progression: Data From the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheumatol. 2015;67(12):3184-9.
15. Welsch GH, Mamisch TC, Domayer SE, Dorotka R, Kutscha-Lissberg F, Marlovits S, White LM, Trattnig S. Cartilage T2 assessment at 3-T MR imaging: In vivo differentiation of normal hyaline cartilage from reparative tissue after two cartilage repair procedures - Initial experience. Radiology. 2008;247(1):154–61.
16. Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, Mendelsohn J, Lazar V, Kurzrock R. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. J Clin Oncol. 2015;33(32):3817–25.
17. www.nih.gov/precisionmedicine.
18. www.rsna.org/QIBA.aspx.
19. Liu Y, deSouza NM, Shankar LK, Kauczor HU, Trattnig S, Collette S, Chiti A. A risk management approach for imaging biomarker-driven clinical trials in oncology. Lancet Oncol. 2015;16(16):e622–8.

2 Introduction to the Stepwise Development of Imaging Biomarkers

Luis Martí-Bonmatí

2.1 Introduction to Imaging Biomarkers and Precision Medicine

Since the beginning of radiology, its traditional goal was to provide images of the highest quality to allow the most accurate diagnoses [30]. This goal has changed. Clinical radiologists must now incorporate biological development and preclinical research into a clinical reality. The use of imaging measurements in individual patients improves the main clinical outcomes, such as depicting abnormalities, predicting prognosis, staging diseases, and defining follow-up outcomes. Radiologists may hereby engage patients and physicians into new approaches to facilitate decision-making for personalized care [12].

There are different steps within the healthcare cycle where medical imaging has a clear role in personalized medicine (Fig. 2.1). The way to foster this role is through the promotion and adoption of quantitative imaging features. Although the term biomarker applies to all detection methods used in life sciences to evaluate a biologic parameter, the radiologist's role is centered in imaging biomarkers and radiomics. The impressive developments in digital imaging, producing high-quality and signal-controlled medical images, generate a wide range of useful information allocated within the radiology departments. The remarkable advances in all medical imaging technologies in the last decades have made it possible to obtain anatomic, functional, metabolic, and physiological-related measurements from images acquired in routine clinical settings. Simultaneously, the massive expansion of computing power allows signal modeling and data processing, which generates a large amount of quantitative information that could not be visualized by previous qualitative radiological analysis.

Radiology approaches are based on the interaction of energy and living organisms to analyze tissue noninvasively, revealing properties relevant to detection, diagnosis, prognosis, or response to therapy. The data sets produced by the different imaging modalities are the end point of multiple, interdependent components that may be analyzed qualitatively or quantitatively. By extracting important information from images beyond our usual interpretation, radiologists will participate in the healthcare cycle through mastery of technology, clinical acumen, and attention to patient safety [16].

Imaging seems ideally suited to flourish as a quantitative science. A clinical image is inherently

L. Martí-Bonmatí
Medical Imaging Department and Imaging Research Group (GIBI230), Hospital Universitario y Politécnico La Fe and Research Institute, QUIBIM SL. Hospital Quiron Valencia, Valencia, Spain
e-mail: marti_lui@gva.es

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_2

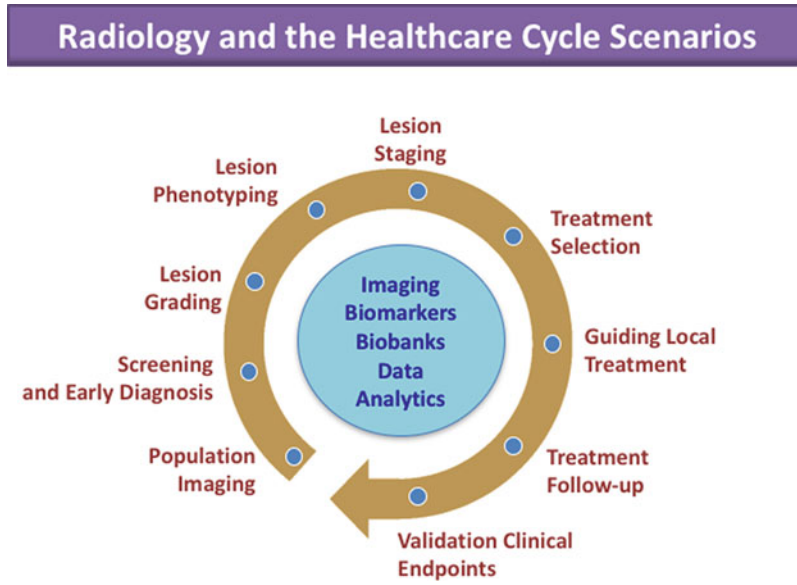


Fig. 2.1 Healthcare cycle and imaging biomarkers' main hallmarks

quantitative, as it is a matrix of numbers [28]. *Quantitative imaging biomarkers* extract and measure objective biological characteristics from any type of medical images [6]. The term radiomics is synonymously used for the development of methods to extract quantitative features which convert single images into mineable high-dimensional data and the subsequent analysis of these data for decision support [10]. The main characteristics of imaging biomarkers are that these tissue properties are resolved both in space, through parametric images, and in time, as response maps. As medical imaging does not destroy the evaluated samples, test-retest evaluations are feasible, allowing the repetition of experiments and measurements as frequently as desired.

Each voxel in a computer-derived image represents both the location and the value of a specific calculated parameter (morphological, biological, response) obtained by the application of mathematical or simulation models to the source images. These synthetic parametric maps represent the new paradigm in clinical radiology and should be considered as virtual biopsies, showing different morphological and biopathological abnormalities. These quantitative images assess the presence and the degree of a condition relative to normal findings. Imaging biomarkers are subrogated, spatially and temporally resolved, in vivo biopsies. Imaging biomarkers provide information related to the individual patient's biological situation and clinical problems. It should be emphasized that imaging biomarker analyses are subrogated to the different underlying processes, strictly showing correlations but not cause-effect conditions [10]. Even more, the diagnostic information derived from these imaging-based measurements is often nonspecific, and the knowledge of the molecular or biological mechanisms remains implicit or unknown.

Biomarkers can be classified as prognostic, if accuracy of patient diagnosis or prognosis is improved; predictive, if the most beneficial treatment can be defined; response, when the beneficial outcomes can be shown after treatment; and monitoring, to detect relapse or toxicity [6].

Technology and scientific biological discoveries have changed the way radiology is performed today. The multidisciplinary interaction between medicine and computer science, which falls within the field of *biomedical engineering*, focuses on the disease's hallmarks that should be evaluated and on the proper way to do it in every clinical scenario. New acquisition techniques and new mathematical models are continuously being

developed to accurately simulate the in vivo physiobiological status of a tissue. The final objective is to assure that the microscopic in vitro, macroscopic in vivo, and virtual in silico realities match to enforce the role of imaging in the major decisions of the healthcare cycle (Fig. 2.1). Imaging biomarker must be acquired in quality controlled and technologically stable equipments, to avoid sources of variability in the acquisition process, under safety conditions for the patients. They must be widely available, reproducible, and standardized in their main acquisition parameters and signal analysis and modeling. Finally, they must be validated in terms of precision and clinical efficacy. All these aspects will be discussed in this chapter.

In the era of *personalized and precision medicine*, imaging biomarkers provide specific information that helps in the decision-making process to achieve a definitive diagnosis, select the best treatment, and accurately evaluate treatment response [11]. To do it precisely, image-driven computing and data mining tools are applied to the analysis of biological systems to explore how the disease affects the different human organs and systems. Therefore, research and innovation in radiology involve multidisciplinary knowledge and team networking. Personalized and evidence-based guidelines will require a multidisciplinary approach to integrate relevant scientific and technological advances and provide expert guidance [12].

The radiologist's role in multidisciplinary teams is to evaluate the evidence and interpret the radiology data in the context of other genetic or health-related data. Tailoring therapy to the individual characteristics of the specific disease requires quantitative information and knowledge of the biological signals associated to the specific disease process. To be used in preclinical and clinical decision-making, imaging biomarkers must be appropriately validated for use as surrogate end points for the application at issue. This pathway must fulfill some conditions, including that the biomarker is closely coupled to the target disease; the detection and measurement of the biomarker is accurate, reproducible, and feasible over time; and the measured changes over time are closely linked to the therapeutic effect and the end points [27].

Quantitative imaging biomarkers, as *precision medicine tools*, are expected to improve routine clinical care and speed the development of new treatments. Scientific and regulatory communities have embraced biomarkers as acceptable surrogate end points for clinical trials, fostering their widespread use in medicine [13]. Imaging biomarkers also aim to demonstrate the phenotypic manifestations of diseases, even when the genotype and penetrance are known, contributing to phenome-wide association studies [11].

2.2 Pipeline Development of Imaging Biomarkers: The Hypothesis

Understanding a disease and the way it is actually managed in a critical way is crucial to success in innovative projects. The radiologist and participating physicians need to know which are the facts whose knowledge will help improve early diagnosis, better phenotype the patient, or select the most adequate treatment. These aspects (Fig. 2.1) must be clearly defined, together with their correlation with the different biological and physiological changes, in order to properly assess the role of imaging for each individual healthcare cycle.

Academic institutions and research companies are establishing centers for translational and preclinical research designed to accelerate the discovery and adoption of imaging biomarker technology. Therefore, defining the appropriate pipeline for biomarker development and implementation is critical for success [5, 9, 21, 22].

Implementation of an imaging biomarker has several consecutive steps before it can be used as an innovative information tool in clinical settings. The different pipeline phases in biomarkers' elaboration resemble those of the development in the pharmaceutical business [6, 26]. The definition of the target hallmark, source images, analytical methodology, and type of measurements are essential aspects that must be considered before studying a specific aspect in a given

disease. Integrating an imaging biomarker into clinical practice needs conceptual consistency, technical reproducibility, adequate accuracy, and meaningful appropriateness.

The path to biomarker development, expansion, and subsequent implementation involves a number of consecutive phases described below (Fig. 2.2) [10, 21, 22, 26]. The development of a biomarker involves not only the validation of its relationship with the objective reality to which it is surrogated, either structural or physiological, but also the monitoring of its overall validity. Biomarkers need to follow all the developmental, validation, and implementation phases before clinical approval.

The *proof of concept* tries to demonstrate that a specific biological hallmark or pathological abnormality might be evaluated using imaging and computational techniques. It can be also considered as the hypothesis to be tested and the main clinical objective or treatment improvement related to this innovative idea. As an example, nonalcoholic steatohepatitis is a condition that involves liver steatosis and inflammation, leading to chronic liver disease and cirrhosis. There is a need to measure liver steatosis and inflammation at the same time, as both biomarkers are synergic toward parenchymal fibrosis progression. It seems that a multivariate combination of a multiecho chemical shift-encoded sequence, giving a T1-T2* corrected estimation of fat deposit, together with a diffusion-weighted MR using the intravoxel incoherent motion model, giving an estimation of the amount of inflammation, might be a useful multivariate prognostic biomarker of steatohepatitis early diagnosis and grading [21, 22].

The *proof of mechanism* is needed to define the expected relationship between the extracted virtual imaging parameter and the relevant disease target that is under evaluation, both in magnitude and direction. In this sense, linearity in the relationship can be considered as the ability to

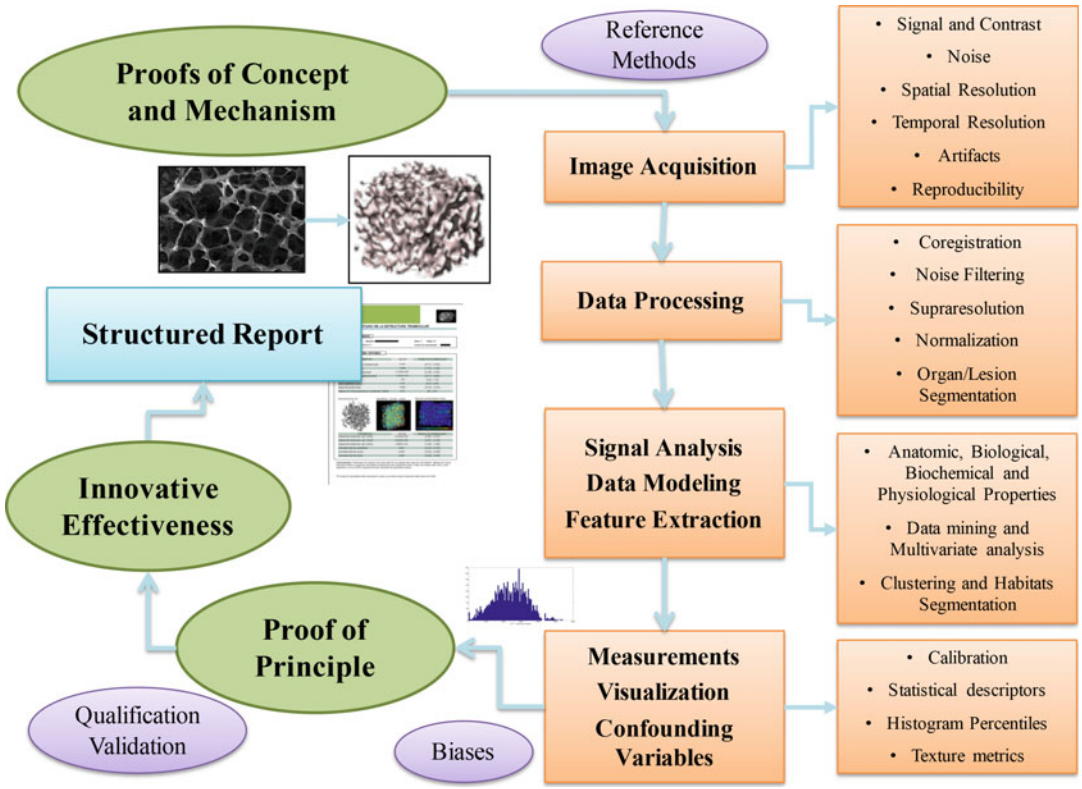


Fig. 2.2 The stepwise development of imaging biomarkers

provide measured quantity values that are directly proportional to the value of the measurand. As an example, if the R2* measurements, corrected by T1 and spectral fat-confounding variables, can measure the liver iron content in an accurate way, even in the presence of large deposits also by initial signal (S_0) decay correction, this biomarker should be used to evaluate the response to treatment in hemochromatosis, as iron depletion will be related to a significant decreased and even disappearance of the liver iron deposits.

Defining both proofs, concept and mechanism, is extremely important as they represent the main hypothesis that has to be proven and tested with the subsequent steps. New methods and measures have to be compared to the known ground truth. *Reference methods* are the approaches or procedures widely recognized as the best available to determine the true state of disease under evaluation. Quantitative reference standards are values, generally accepted as having small measurements of uncertainty, that are used as a basis for comparison by using a reference method. The values obtained through imaging biomarkers need to be compared to known real, or truth, values. The relationships between measured and true values may be proportional, nonproportional, or even nonconstant or nonlinear. In some situations, measuring ranges or intervals should be defined as the range of the measurand in which bias, linearity, and precision are within acceptable bounds, although other ranges will not have this reliable proportion [29]. An example is liver iron concentration in hemochromatosis, as large deposits will hinder an accurate measurement due to the extremely low liver signal.

Defining which are the gold standards to which the biomarkers have to be compared is an extremely important aspect with huge relevance on the final validation process. Most studies agree to use core biopsy and pathology analysis as the gold standard to evaluate most biomarkers. However, pathology has some drawbacks and biases. Pathological studies are an ex vivo analysis, with no information on in vivo dynamic processes, as some biological pathways cannot be evaluated in pathological ex vivo samples. Biopsy is invasive, being associated to morbidity and even mortality. Even more important, biopsy is not feasible for research studies as ethics limits sample repeatability, and follow-up biopsies might not be possible as the biopsied tissue is partially destroyed.

Biopsy has sampling bias due to tissue and disease heterogeneity in distribution and grade, being subject to individual variability assessment with inter- and intra-subject discrepancies. Intrinsic tumor properties, such as the intra- and intertumoral heterogeneity, combined with interpatient heterogeneity, introduce a high grade of complexity in treatment planning. This is the main reason for single-tumor biopsy failure to assess tumor aggressiveness, treatment appropriateness, and tumor resistance. However, imaging biomarkers can evaluate tumor phenotypic heterogeneity when biopsy has limitations in assessing the genetic intratumoral differences.

Pathologic analysis is frequently based on semiquantitative grading and scoring, and not on continuous data, measuring subjective changes and percentages but not quantities, as digital quantitative pathology is still developing. If imaging biomarkers prove to be surrogate findings, they also could provide confirmatory information to support histopathologic findings [10]. Confirmation of disease measurements is often not reliable or even impossible also at autopsy.

In comparison to pathology, clinical outcomes might be more appropriate in some circumstances. Patient's survival or time to progression may be more appropriate end points in cancer-related evaluations as prognostic outcomes. In this way, the relationship between perfusion permeability at the periphery of brain glioblastomas and patient survival seems more relevant than the pathological proof of tumor infiltration [25].

It is critical to recognize that all reference methods have some general biases. Measurement uncertainties can be present as there might not be a single right answer in those cases with a heterogeneous distribution of the biological hallmark. As tumors and lesions are nonhomogeneous in their phenotypic, physiologic components and genomic aspects, histogram-based analysis of the evaluated parameters may be more appropriate

than normal statistical descriptors, such as the mean. Also, researchers have to consider the problem of short-term intrinsic variability, as in vivo biological examinations might be influenced by physiological changes in the subject and the lesion over time. As an example, liver stiffness for fibrotic evaluation in chronic diseases is subject to the patient's fasting state.

2.3 Image Acquisition and Preparation for Analysis

Appropriate source images are essential for the extraction of useful biomarkers. The best image modality and protocol have to be defined for each biomarker and imaging modality.

Image acquisition (Chap. 4) must provide the best and more reproducible images, standardized by the radiological community after image quality and signal stability validation studies. The target organ must be studied with sufficient coverage and spatial resolution. The best compromise between spatial and temporal resolution requirements must be discussed for each biomarker to fit the analysis requirements. Data quality has to be checked regarding signal- and contrast-to-noise ratios, spatial resolution, temporal resolution, artifacts, and reproducibility. Periodic image quality control must be performed to assure data stability and quality over time. There have been multiple efforts to uniform the definition of acquisition and reconstruction standards [10].

Patient anonymization must be accomplished before the images are sent to any storage or computing server outside the hospital network. The pipeline must follow the principle of providing the minimum amount of confidential information, such as patient identifiers, to avoid the possibility of patient identification outside the hospital network. It is, therefore, necessary to accommodate the analysis of imaging data to DICOM image data stripped of identified headers and assigned a de-identified number [15].

To guarantee that the acquired images are optimal for the analysis, *data processing and image preparation* (Chap. 5) tools are needed to improve source image quality before the voxel-wise signal analysis step. General procedures include filters to reduce scattered noise and homogenize signal, as noise and heterogeneous signal distribution is a major confounding factor [17, 18]. Noise reduction techniques improve the entire signal modeling processes and must be always considered in the process of imaging biomarkers' extraction. Artifacts should be also removed if possible as they introduce non-real data that will bias the signal analysis process. In most cases, there is a need to increase the spatial resolution of the acquired images and to enhance the detail of the tissues through super-resolution interpolation algorithms [2, 3]. The highest possible resolution is always recommended, as in trabecular bone 3D virtual model example.

Image registration methods transform different data sets into one spatial coordinate system. In dynamic acquisitions, where data is obtained over finite time periods, image coregistration through voxel repositioning must be performed to guarantee an accurate spatial coherence and ensure that the evaluated anatomical area is coherent in space in all the image series [8]. Registration is needed to be able to analyze and integrate the acquired data. A clear example is the voxel-by-voxel analysis of the dynamic series obtained after the intravenous administration of a contrast agent for tumor response evaluation, as respiratory and vascular movements displace the 3D data sets in every single acquisition.

Image segmentation is used to locate organs and lesions' boundaries, labeling every pixel pertaining to a specific organ or lesion. Organ or lesion segmentation facilitates their analysis and the results visualization. Either a region of interest (ROI) or whole-organ maps (VOI, volume of interest) can be manually selected or extracted by automatic segmentation. Segmentation of images into VOIs such as lesion and normal organ is a crucial step for the subsequent informatics analyses. Manual segmentation by expert readers is often treated as ground truth, although it suffers from high interreader variability and is labor intensive [15]. Segmentation algorithms without user dependence are preferred over manual segmentations to minimize inter-subject variability. Segmentation of normal structures and organs, as

well as well-defined lesions, can now be achieved with full automation. Advance segmentation of different subvolumes or habitats is more complex and requires knowledge of feature distribution. This process will be discussed later in the measurement and visualization section within this chapter.

2.4 Image Analysis and Feature Extraction

After image preparation, *signal analysis and modeling* (Chaps. 6 and 7) procedures have to be implemented to extract with the most appropriate computational processes the targeted identifications and required features from the acquired medical digital images. Static anatomical methods estimate tissue aspects related to the volume and shape of the tissues, topology, and co-occurrence matrix features for texture classification, while dynamic biological analysis assesses the different physical, chemical, and biological hallmarks (Fig. 2.3). As examples, cortical thickness and lung emphysema analyses are static methods, while fat and iron measurements or ADC quantification within the pancreas or perfusion-related D* measurements in the prostate are biomarkers obtained after biological dynamic modeling of the acquired data.

The calculated tissue properties, obtained from each voxel within the image-segmented framework, will demonstrate the spatial distribution of the biomarker by the use of 2D or 3D parametric images. In these color-coded maps, the pixels' brightness represents the value of the specific biomarker in a color scale, showing the distribution of the parameter all over the evaluated

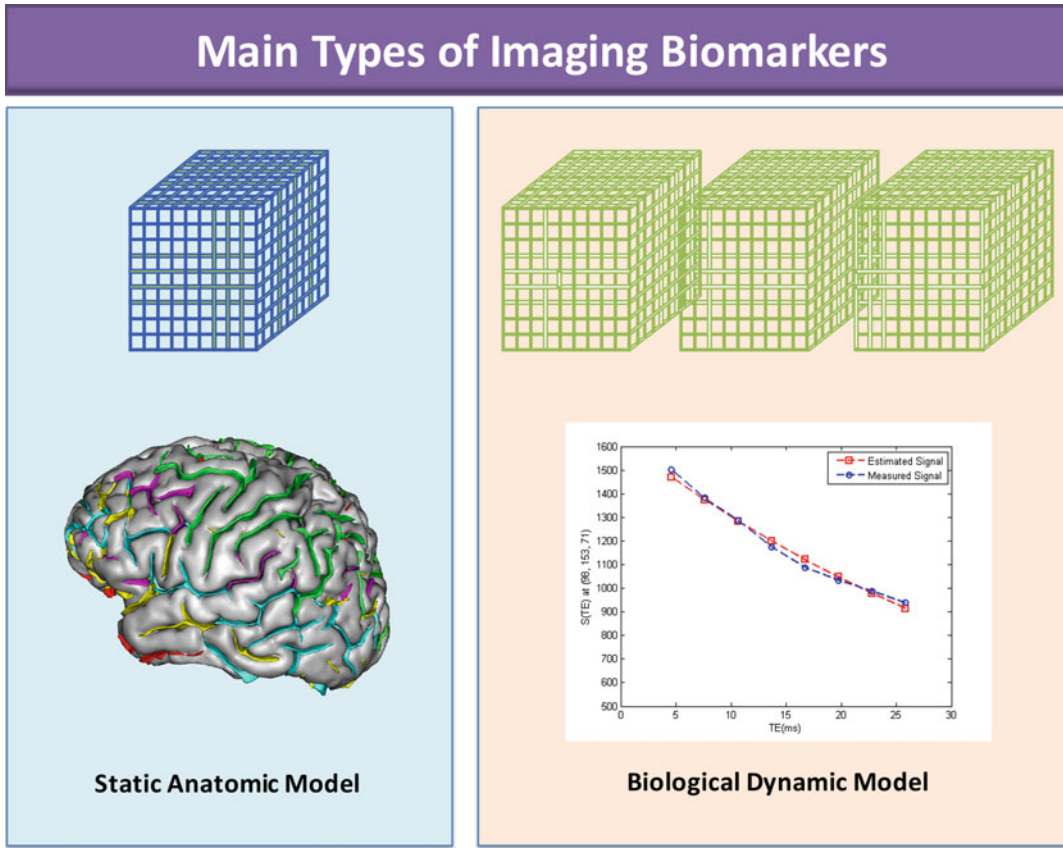


Fig. 2.3 Main types of imaging biomarkers

tissue or organ. One successful representation is to show only those abnormal color-coded voxels overlaying the gray-scale anatomic reference image. The basic aspect of radiomics is the extraction of high-dimension feature data to quantitatively describe different attributes of the volume of interest within an organ, a lesion, or a subregion. Radiomics data are in a mineable form, allowing building descriptive and predictive models relating image features to phenotypes or genetic signatures.

Data mining with imaging biomarkers and radiomics data allows development of classification schemes, or models, to predict outcomes. These alone or in combination with additional information, such as demographic, clinical, liquid biopsy, or genomic data, will improve the clinical value of imaging biomarkers and radiomics analysis [10].

Combining multiple imaging quantitative parameters that reflect different aspects of pathophysiological processes will provide even newer insights into most diseases. Multivariate parametric images allow demonstrating the abnormal combination of biomarkers relevant to the evaluated pathways, reducing the amount and redundancy of the acquired data. The color of each voxel is determined by a multivariate function and shown habitats, reflecting the different microenvironments within the tissue. Some of the most popular multivariate statistical methods are linear regression, discriminant function analysis, and independent or principal component analysis. Statistics and pattern-recognition-based techniques may determine which computationally derived biomarkers, and with which weights, provide the most useful information about the clinical question or outcome being evaluated. The final output of the multiparametric analysis can be considered, mainly if specimen biomarkers are also included, as a nosologic image that shows on a pixel-by-pixel basis the probability of a pathological change or biological condition, expressed within an organ or lesion and relevant to the patient (Fig. 2.4).

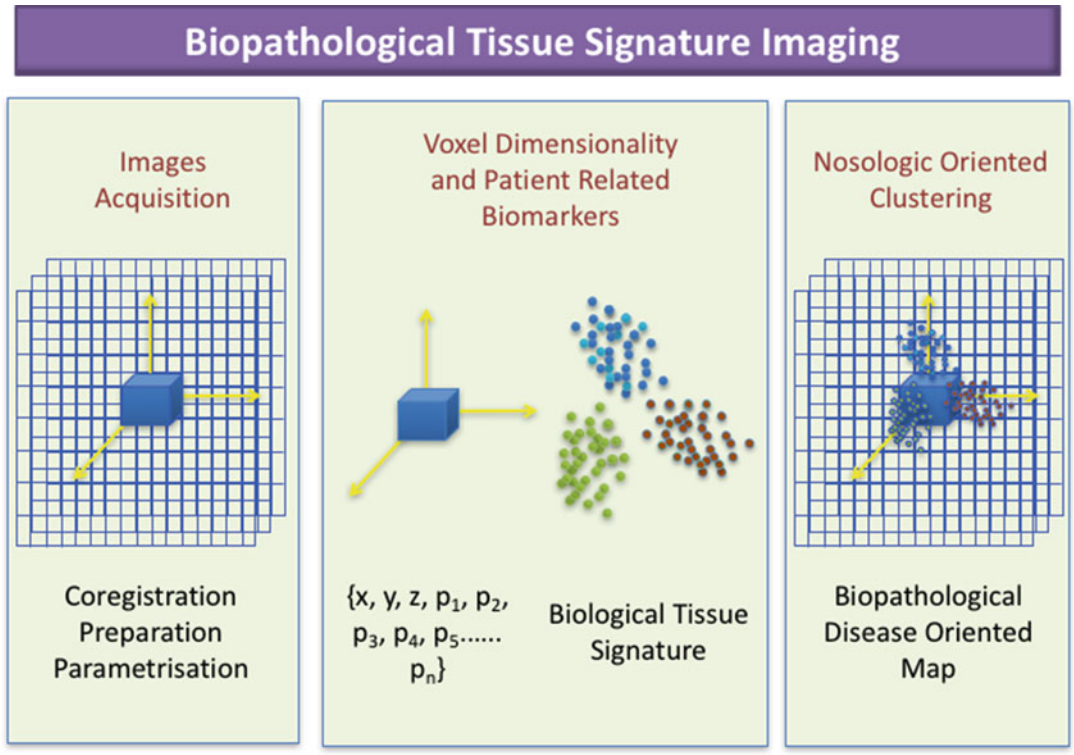


Fig. 2.4 From image acquisition to biopathological multiparametric maps

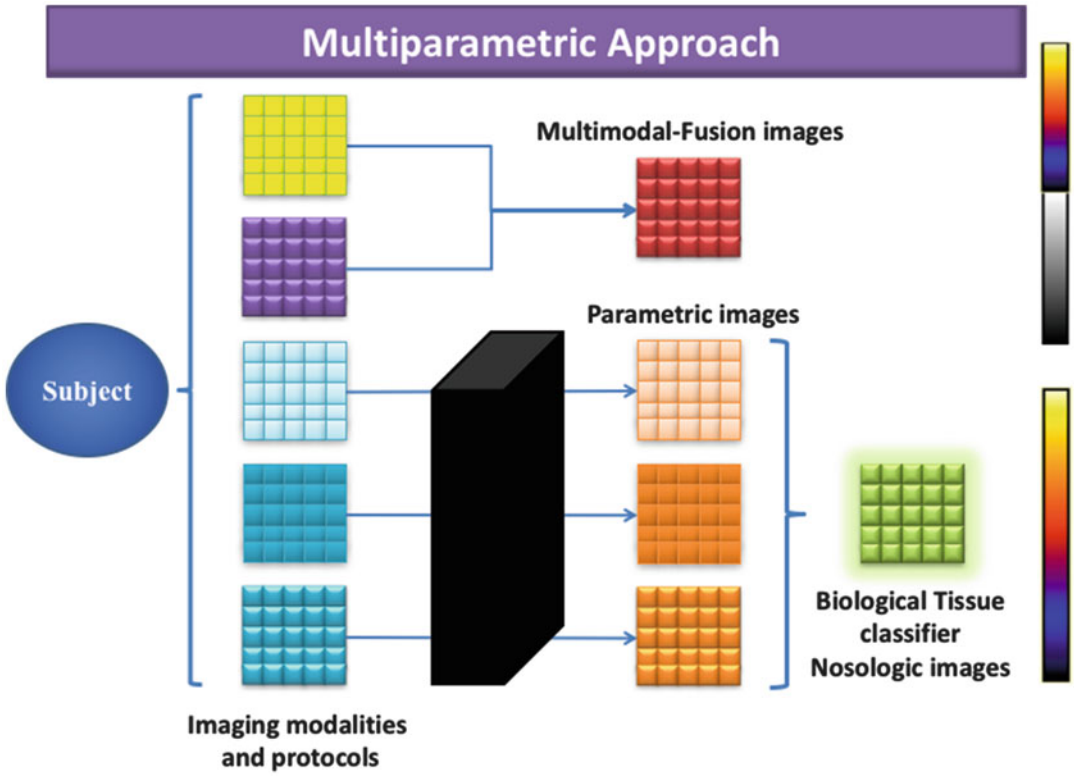


Fig. 2.5 Multiparametric and multimodality approaches

In some cases, these surrogate end point clusters, including both imaging and nonimaging surrogates, may be better in predicting clinical outcomes than single surrogates, as disease outcomes are rarely the result of a single factor entirely encapsulated by one biomarker [26]. *Multivariate*, *multidimensional*, or *multiparametric maps* might demonstrate the disease's hallmark presence and distribution, in a voxel-by-voxel basis, if the answer is complex enough to be represented by a single surrogate property. The basic philosophy is based in process engineering, capturing as much data as possible at the front end and use downstream database mining to identify the features with the highest prognostic value [19].

It is necessary here to clearly define the difference between multimodality and multiparametric analysis. In multimodality imaging, two or more imaging techniques are combined to compensate for the disadvantages of each imaging system while taking advantage of their individual strengths. Combination might be synchronous, at the same time, or metachronous, at different but close time points. In multimodality imaging, the voxel signal is a linear visualization of two different color palettes (i.e., PET-CT and PET-MR). Multiparametric imaging reflects the result of multidimensional data reduction and classifier model techniques applied to relevant parameters on a voxel-by-voxel basis to compensate for the disadvantages of single parameter-isolated analysis. The voxel signal tries to give a nosologic answer to the disease (Fig. 2.5). The combined radiologist's subjective evaluation of two or more parameters and different images cannot be considered a multiparametric approach, although several of these notifications can be found in oncologic imaging papers, such as prostate cancer depiction.

2.4.1 Imaging Biomarker's Metrics

After the parametric and/or multiparametric maps are obtained, the different tissue or lesion characteristics have to be measured and shown. Measurements are an essential part in imaging biomarkers' development and integration, as they assign a number or range to a characteristic of a tissue or lesion to be used in research and clinical practice. Radiologists have to define which is the best way to evaluate the target tissue parameter. Chapter 8 is dedicated to the need for define appropriate measurement strategies of imaging biomarkers.

After identifying the parameters having diagnostic, prognostic, or therapeutic value in a specific disease, there is a need to segment the relevant volumes and extract and qualify the main descriptive features from the volume. Imaging features extracted after identification of the VOI can be obtained from either the entire lesions or from defined detailed subvolumes of interest, known as habitats or clusters reflecting the different physiologic microenvironments [10]. The different habitats within an organ or a lesion can also be selected after data mining from the mathematically obtained multiparametric maps. Habitat properties, also known as agnostic features, attempt to capture lesion heterogeneity through quantitative descriptors.

Segmentation is a critical and challenging component of the pipeline process, as the biomarker data are generated from the segmented volumes and habitats may have indistinct borders. Data mining from the different source images or parametric maps might also extract different habitats, reflecting different microenvironments, to be used as VOI clusters. The histograms obtained from the segmented volumes of interest allow a graphical representation of the distribution of the target parameter. The vertical axis represents the frequency for the observed value of the biomarker, while the horizontal axis represents the different observed values. Descriptive metrics (such as mean, mode, standard deviation, interquartile, kurtosis, entropy, inertia, energy, correlation) must be evaluated in order to select the ones that best demonstrate the tissue target abnormality and best correlate to the basic truth. In most situations, radiologists should avoid the tendency to use mean values as they underestimate the abnormal changes by averaging. Frequently, histogram analysis and interquartile values will show a higher and more appropriate relationship with the biological or clinical end points.

Parametric images, both conventional and multivariate nosologic, provide measurements from either the whole tissue being studied (VOI) or only from those areas considered more representative or abnormal (ROI). Volumetric analysis of the abnormal target parameter distribution, either an organ or a lesion, is preferable as this analysis shows the relevant abnormal parameters selected by thresholding the histogram distribution and not by a subjective visual evaluation of the radiologists, which will bias the measurements. To be used in clinical decision-making, quantitative biomarkers need to have established cutoff points or threshold values.

The analysis of the *heterogeneity* in the spatial distribution of a biomarker provided by its parametric image is also extremely useful. Kurtosis is a statistical parameter that measures the sharpness of the histogram, higher kurtosis values meaning more homogeneous distributions. Entropy is another measure of tissue organization that can be measured. Heterogeneity measurements are being used in imaging cancer evaluation as prognostic variables.

Biases have to be considered before validation studies are planned. All over this process, there is a need to standardize the entire imaging workflow, also for multimodal and multiparametric approaches, in order to achieve a high level of robustness and reproducibility of results in any multicenter study validation. Bias can be considered an estimate of a systematic error that affects and distorts the measurement process, generating wrong values and leading toward misinterpretation of results. Bias describes the difference between the expected value and its true value. Therefore, biased studies lack validity. The different sources of error during the sampling and data processing procedures must be identified and minimized. The aspects of measurement uncertainties are further detailed in Chap. 9.

Phantom studies are critical in the biomarker's development. Some biomarkers measure compounds or structures and need to be calibrated with phantoms to ensure that the degree of correlation between obtained measurements and the property is accurate and stable. Phantoms must be also used to control changes in signal-to-noise ratios, signal uniformity, and spatial distortion. However, phantoms have some limitations as they do not have all of the characteristics of a human target and also might not adequately account for the different measurement distortions that can be induced by the physical properties of living tissues [29]. Other disadvantages and limitations of phantoms include the potential lack of realism compared with the in vivo measurements, as some physiological quantities, such as vessel permeability, are extremely difficult to simulate in a phantom, and the measurement process does not replicate the complexity of the in vivo process. Also, measured properties may vary with time, due to either a temperature dependence of the parameter under study, such as in diffusion-weighted experiments, or by instability of the material over time, as can be seen by fungal attack or chemical decay of the phantom's compounds [31].

2.5 Biomarker Validations

To be clinically useful, research on quantitative biomarkers must adopt common standards and cross-disciplinary, systems-based approaches for biomarker discovery and validation. Qualification is a measure of the use of a biomarker in specific contents, while validation refers to the general performance of the biomarker [6]. Qualification relates to clinical approval, while validation relates to the performance of the test. Validation of imaging biomarkers is challenging, mainly because the disease-related changes in the tissue properties measured at imaging, such as the D and D* components in an IVIM DW MR experiment, are indirectly linked to several structural changes such as necrosis, cellularity, fibrosis, and vascular architecture and are influenced by coexisting factors, such as inflammation, perfusion, permeability, and interstitial pressure [11]. Qualification and validation require the evaluation of the nature and strength of evidence regarding whether a biomarker is associated with the disease and assembly of available evidence demonstrating that interventions targeting the biomarker impact the clinical end points of interest. Chapter 10 is dedicated to the validation phases of imaging biomarkers.

Qualification, *validation*, and *standardization* are crucial parts in the stepwise development of imaging biomarkers. The influence of the center, equipment, technical parameters, and other biases must be scrutinized before the clinical potential can be assessed. Even more, the voxel-measured signal is complex in most clinical situations, as voxel constituents are diverse and might interfere with each other. Extremely important is the consideration that image signal comes from the voxels, and these are complex in their components and physicochemical properties. Therefore, obtained measurements must have a precise relationship (interclass correlation coefficient) with biological reality. The complex components and properties of a voxel introduce bias when the signal is evaluated to represent a physicochemical reality (Fig. 2.6). Before assuming this correlation, a well-defined relationship between the underlying measurand and the biomarker has to be established. Several validations, including technical, biological, and clinical, have to be conducted before a biomarker is introduced in clinical research. Reducing variability across devices, patients, and time is a must for a reliable and reproducible biomarker. Even more, the clinical usefulness in terms of the benefit for the patients and improvements of outcomes has to be proven.

The development and implementation of imaging biomarkers need a consistent and correct use of terminology and methods regarding technical performance and statistical concepts. Technical performance can be considered as the assessment of how a test performs in reference objects or subjects under controlled conditions [29]. *Technical validation* evaluates and controls the different measurements related to image acquisition, processing, analysis, and measurements of the biomarker prototype [20]. This vali-

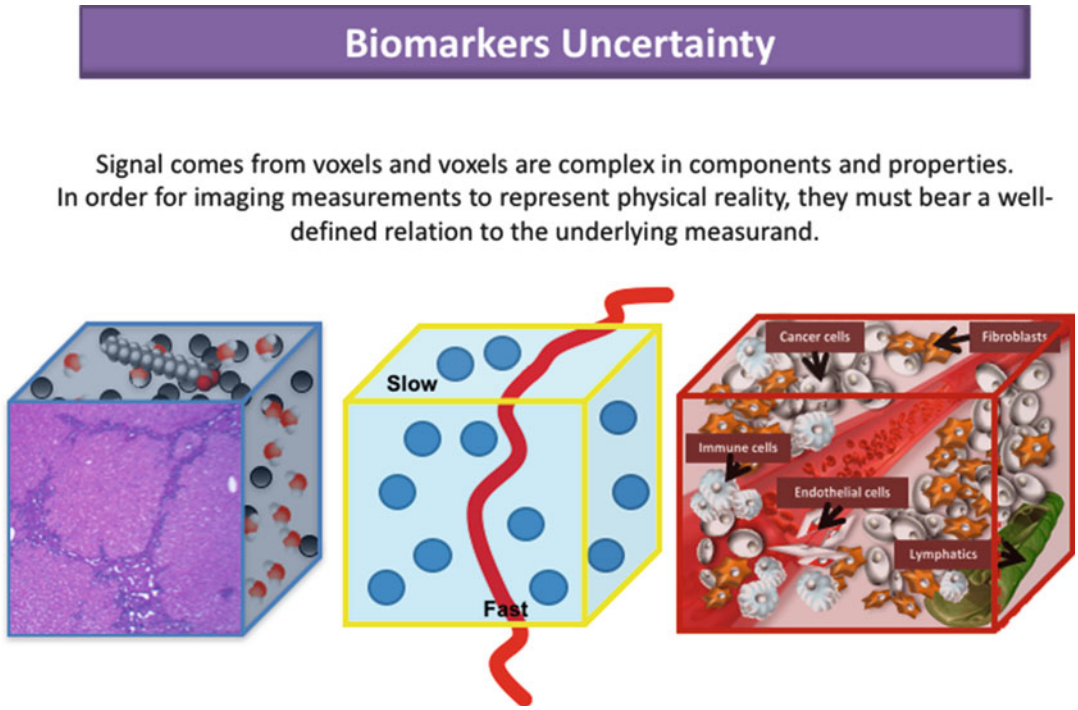


Fig. 2.6 The complexity of the voxel constituents

dation is usually performed initially as a single-center experiment, where iterative definitions on reference quality controls have to be implemented. Typically, experiments deal with the influence and the biases associated with changes in the different technical steps. Results are targeted to find which solutions are more robust and precise and how to control deviations from the optimal correlations. Precision can be defined as the closeness of agreement between measured quantity values obtained by means of replicate measurements under specified conditions. The quality of a diagnosis is based on the accuracy of a given biomarker, but the quality of monitoring of disease severity and treatment response effects depends on the precision of a biomarker [11]. Of relevance, precision is often not constant over the range of values of the measurand that are of clinical interest. Precision profiles for each biomarker over the range of interest must be, therefore, assessed [29].

In any attempt to minimize multisite differences, the initial effort should be focused on matching the sequence parameters and image analysis procedure [31]. Repeatability and reproducibility are types of precision. Repeatability must be considered a measurement of precision with conditions that remain unchanged between replicate measurements. Repeatability studies are typically conducted at a single clinical site with a specific imaging device. Human repeatability tests are usually limited to a few scans performed as test-retest experiment.

In reproducibility studies, the measurement precision is evaluated with conditions that vary between replicate measurements. Reproducibility studies are designed to evaluate different factors that may affect the biomarker measurement's precision, including clinical sites, scanner models or manufacturers, standard procedures, operators, technologists, and radiologists [29]. One of the most important challenges in implementing any quantitative protocol in a multicenter clinical trial is developing a procedure that is sufficiently specific to guarantee the ability of each site to meet the informational requirements while also allowing enough flexibility to accommodate the different scanner makes and models, coils, and

software versions available at the various sites [4]. To control variability, test-retest studies should be conducted at this phase to improve reproducibility by knowing the coefficients of variation and the influence that the different steps have on them. Experimental models for method validation are needed before the test goes for biological and clinical validations. If the results are biased, the new proposed methodologies might be wrongly favored against conventional tests.

After single-center experiments, the tests have to be performed at different centers and with different equipment and conditions. *Multicenter analysis* should consider how the experiment could be safely implemented in different settings, evaluating and controlling how these changes affect not only the measurements but also the correlations with the reference standards. These multicenter analyses are based on iterative reduction of biases and variance across the different scanners and techniques. To avoid deviations between centers, different approaches can be designed. First, optimal acquisition protocols from the single-center study should be defined and shared in order to minimize variability. Also, calibration studies should be performed, if feasible, in order to correct for deviations on raw data and measurements. Computer simulations and adjustments can be also implemented to further improve the reproducibility of the biomarkers in the different settings. A clear final technical specification, with its accuracy and precision characteristics, will be the standard procedure to be implemented for the biological validation studies.

Biological validations are based on the correlation between preclinical disease models or human studies and other reference methods, such as histopathology, immunohistochemistry, invasive measurements, or genomics. These biological validations are experimental studies that select the appropriate reference standards relevant to the biological change to be tested and evaluate the different confounding variables related to target outcome (Fig. 2.6).

Basic biological validation has to be done in single-center experiments, although multicenter evaluations of the influence of the individuals' physiological status on the calculated measurements are also needed. In these studies, the influence of epidemiologic data (such as sex, age, physiological status) and genuine biological variations on the measurements has to be studied. Physiological changes are a source of biological variation. This is the variance measured with an instrument that has perfect reproducibility and represents how much the biology of a tissue varies over a short period of time. One clear example is the relationship of the regional blood flow or blood volume changes in response to cardiac output, while true tumor volume does not change [31].

Clinical validations are initially performed as single-center observational studies in patients where the biomarker is analyzed in terms of how well it works in small and controlled conditions. If the biomarker proves to be accurate in this setting, multicenter observational patient cohorts, both retrospective and prospective, studies and even clinical trials can be started. They are needed to evaluate the biomarkers' acceptance and robustness. Data from clinical trials or large observational studies from imaging biobanks are recommended for validation and approval of imaging tests and quantitative imaging biomarker qualification [6]. Many of the barriers for an optimal developmental pathway for imaging biomarkers relate to the existing regulatory and fiscal environment, mainly when radiotracer agents are involved. The main metrics for clinical validation of broader clinical relevance from a patient's perspective are improvement in functional status, quality of life and morbidity, downstream care and health events, and death [7].

The first single-center analysis is usually performed as a pilot test on a small sample of well-controlled and defined subjects. This first pilot study is also known as the *proof of principle*, whose purpose is to verify that the concept and related methods have the potential of being used. This small initial clinical validation study must be performed before embarking on large-scale multicenter projects and clinical trials. The proof of principle can be considered as a dummy run study practice to evaluate accuracy and potential confounding variables. As an example, the high

temporal and spatial resolution obtained in dynamic contrast-enhanced breast images, reflecting tissue physiological parameters related to microvasculature density, permeability, and surface area, is a prognostic and predictive factor for tumor management [23]. Successful study designs comparing well-characterized groups of subjects must be also prospectively validated in individual single-subject analyses.

The next step is the *proof of effectiveness*. Only after successful results at the proof of principle, researchers must conduct a larger multicenter series to validate in routine clinical practice the biomarker. This clinical efficacy groundwork must be performed to qualify the quantitative imaging biomarker for its intended use in the appropriate "real-world" imaging conditions. The study must be extended to large series and should be able to depict any statistically significant conclusion and to assess the ability of the biomarker to appropriately measure the target hallmark in a reproducible, reliable, and accurate way. This study must be multicenter in order to demonstrate the behavior of the developed biomarkers, analyzing the biomarker relationship to the biological or clinical end points in normal practical conditions in different centers. Evaluating the sources of multicenter variation due to the imaging process, and then reducing them where possible, will allow progress to be made. Differences between centers can be characterized, understood, and minimized by the use of phantoms or test objects and normal control subjects [31]. Good multicenter agreement on imaging measurement demonstrates that the imaging technique is appropriate and the results can be applied to other centers. This intercenter agreement also demonstrates that the physical factors involved in the measurement process are well understood and controlled, and the process is reliable and robust. Intercenter agreement must control that serial measurements remain reliable despite imager upgrades [31].

Only multicenter-validated imaging biomarkers should be used as surrogate end points. Decisions to utilize imaging biomarkers depend on the specific proposed topic in addition to the strength of the available evidence. Strong evidence and a compelling context are needed for the use of a biomarker as a surrogate end point, as surrogate end points are meant to replace clinical end points [26]. The developed biomarker must not only be standardized and accurate but also clinically useful, producing a clinical improvement in the patient healthcare cycle (Fig. 2.1) and an increased patient satisfaction [7]. The imaging biomarkers must have sensitivity to the effect they are measuring, high percentage of true positives, and high specificity. Even more, the biomarker should be lowly priced, quickly obtained, and easily applied.

2.6 Innovating with Biomarkers

The adequate development, implementation, and clinical incorporation of new imaging biomarkers into structured reports would be one of the main roles of biomedical engineers in the radiology departments. Radiologists and engineers may lead this personalized and precision innovation task in medicine.

Qualified biomarkers should have biological relevance to the disease process under study, sensitivity to the disease process, and good reproducibility. Biomedical imaging provides parametric and multiparametric morphologic and dynamic structural and biological information that enable focused, minimally invasive treatments and imaging-guided interventions (Fig. 2.7). Imaging biomarkers are key elements in the infrastructure needed for personalized medicine [11]. Biomarkers, as imaging extracted quantifiable features, are useful only if they provide additional accuracy in predicting clinical outcome and disease response beyond that which is attained without them. Their ultimate objective evidence regarding their relationships with health status must be therefore established in clinical and research practice.

The optimal application of quantitative imaging to oncology clinical trials requires systemic evaluation of the ability to standardize modalities and algorithms across different platforms and among multiple institutions, awareness of incremental cost of the imaging techniques, and

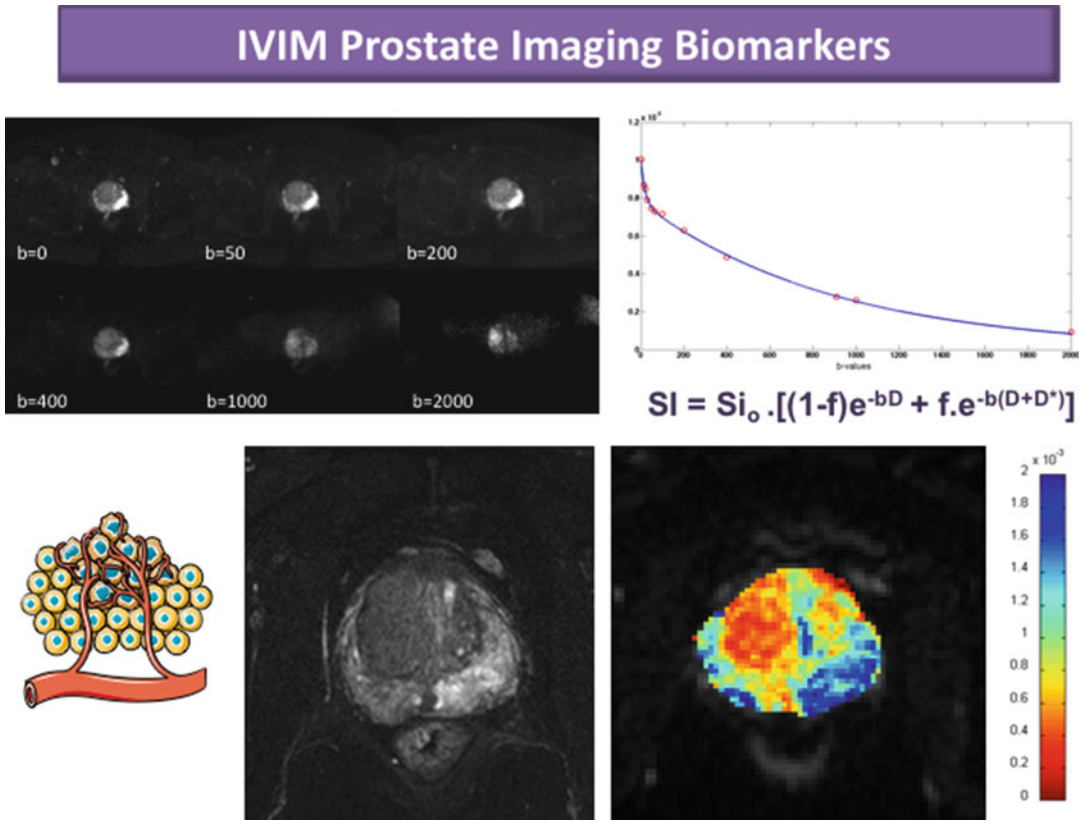


Fig. 2.7 Intravoxel incoherent motion (IVIM) diffusion-weighted MR experiment with 6 b-values (DW images) in a patient with prostate carcinoma (Gleason 7). Signal decay was fitting with a biexponential model after denoising and coregistration. The relationship of signal intensity and D, D* and f is shown in the model. The D parametric map shows the abnormal red values, correlating with higher restriction and higher cell density, having a heterogeneous distribution in the transitional gland. D parametric maps allow guided biopsy and targeted interventions

estimates of potential cost savings by enrichment of responders [14].

New imaging biomarkers will only have an impact in clinical practice if the information they provide answers the clinical problem. To add value, *radiology reports* must be structured, use standardized terminology, convey actionable information, and be limited in their variability, especially with regard to recommendations. To be used, the derived quantitative data must be organized and exposed in simple and intuitive ways. To be useful, radiology reports must show clinical skills, through clarity and pertinence of the results; communication skills, by reporting all the relevant information with standard ontologies; and innovative skills, by incorporating quantitative biomarkers when appropriate. These items will reduce the report variability and uncertainty and improve its utility.

To bring this innovation into clinical practice, biomarker results need to be displayed in an intuitive way. Post-processing platforms and integrated structured imaging biomarker reports should be implemented and their communication skills tested. This would be the best way to assure a paradigm change in the radiological workflow within radiological departments. An appropriate system to convey these results is the *structured reporting* driven by disease. This architecture is designed as a means of encoding documents and exchanging pertinent information through a hierarchical structure. Structured reports define the systematized data, including images, biomarkers, and template report [24]. The presentation of

diagnostic reports requires ensuring that all the pertinent information is checked, including the technique used as reference, the statistical methods and method reproducibility, and the clinical applicability to help in tackling the relevant clinical end points. They also allow for data storing, search, retrieval, statistical analysis, and transfer. Furthermore, digital structured reports including quantitative information extracted from images have the ability to associate the clinical document of the biomarkers with the patient's episode within the electronic health records and PACS. Practical workflow limitations to incorporate quantitative methods into routine radiology reporting have to be solved before the incorporation of more advance imaging metrics into radiology reporting as clinical decision-making paradigms become a reality.

The structured report must comprise complete, comprehensive, and accurate information including the assessment of potential bias and a generalization of the results. Structured reports must be useful for follow-up studies and fully integrated into eHealth solutions. Radiologists should avoid the disconnection between the enthusiasm for quantitative imaging in the academic literature and the skepticism that clinical radiology has traditionally held toward numerical description. Increased incorporation of quantitative imaging into clinical radiology has to be made through the radiology reports [1]. To speed the process of quantitative reporting, both the radiologist and the referring physician may consider that the increased precision resulting from validated quantitative interpretation has a real effect on patient management and would therefore provide an incremental value over a qualitative description under real-life clinical conditions.

Radiologists might interact with patients and referring physicians to confirm that their information is used properly. Radiology reports combined with other available patient information impact on the decision-making process. The interpretative nature of imaging tests, unclear importance of some findings in parametric imaging, variable imaging acquisition and quality, and limitations of available evidence are examples of sources of uncertainty that radiologists have to share with referring physicians and patients [12].

The technological developments for structured reporting implementation are further explained in Chap. 11.

2.7 Biobanks and Biomarkers: Big Data

The new radiological scenario emerging after the introduction of biomarkers provides considerable benefits to the diagnostic imaging professionals as they provide in vivo biological information resolved in time and space. As with other emerging technologies, biomarkers may initially build to a peak of expectations very different from reality; but over time, the quantitative information provided by the biomarkers will finally be incorporated into daily clinical practice.

The ability to accommodate large collections of clinical evidence data for these new biological metrics into stakeholder business models would provide incentives for companies to produce and market them for improved patient care. These large collections of clinical evidence and imaging data are therefore mandatory in the imaging biomarker's development pathway. *Imaging biobanks* allow to study advanced imaging techniques on imaging pools with sufficient sample size, since demonstration of significant associations between an imaging biomarker and a phenotype or a genetic variant requires very large numbers of tested individuals [11]. Once large high-quality and well-curated data sets are available within the biobank, they can be used for data mining, referring to the process of discovering patterns in large data sets.

Medical imaging biobanks have recently emerged for advancing on the study of rare diseases, the identification of early biomarkers and surrogates, and the development of population studies. These big data collections must be efficiently organized. *Structured digital data*, such as images, text, and measurements, are logically integrated and organized within these large repositories. From a single-hospital database to

multicenter electronic health records and PACS repositories, biobanks have emerged as a fundamental tool for clinical research in personalized medicine through quality control issues for sample collections, standardized pathways for quantitative information extraction, and sophisticated protocols for data protection. Imaging biobanks are therefore organized databases of medical images and associated imaging biomarkers, shared among multiple researchers, and linked to a bio-repository. These biobanks will help to evaluate the impact of new imaging biomarkers on early disease diagnosis, disease phenotyping, disease grading, targeting therapies, and evaluation of disease response to treatment. They will also aim to develop, throughout their growing volume and complexity of imaging data, decision-support algorithms required to help physicians apply the most essential patient data for optimal management [11].

Parametric- and multiparametric-derived and source image data and metadata sharing across multiple sites are needed to establish imaging biomarker's models to be used as decision-support tools. The appropriate development of new and reliable imaging biomarkers has encountered bottlenecks in some cases. Such limitations might be addressed through access to structured imaging biobanks, as they will provide standardization (large samples, quality control, access), validation (multicenter, large population studies), and benchmarking (biomarker discovery, image contribution). Biobanks must have structured and fully anonymized information, including medical images and relevant clinical and associated biological data and/or samples. Data dissociation must allow for traceability of cases in unexpected findings, while high-performance computing resources will facilitate image processing comparison, standardization, and validation. Imaging biobanks will surely promote the development of new multivariate or nosologic imaging biomarkers through genetic and circulating sample biomarkers integration. Curation and collection of high-quality imaging data require content expertise to identify and circumscribe, with computer assistance, and annotate, with a standardized and mineable lexicon, the volumes of interest [10].

Imaging biobanks can also generate a space of research, development, and innovation in healthcare management, by promoting new ideas and technological innovation projects based on digital medical imaging, data mining, and clustering. Biobanks must also provide access to quality-defined human health/disease-relevant imaging and biological resources including associated data in an efficient, ethical, and legally compliant manner.

A federation of imaging biobanks would establish an integrated platform providing services of imaging data and processing, easing the long-term access of science researchers to computing infrastructures. This international collaboration will require the advance on both technological and organizational matters (incorporation, procedures, protocols, data sharing, boards, and access criteria). This environment can be considered as a framework on top of a set of computing and data-intensive infrastructures that will provide researchers with tools, protocols, data, and expertise to improve medical imaging research and patient's healthcare.

Large biobanks will improve healthcare through studies of quality control, such as image quality and radiation dose, technical and protocol comparisons, and follow-up assessment of clinical guidelines. They will also improve the therapeutic process through the translation of research findings into clinical trials with guided imaging, assessment of image biomarkers prognostic factors, and early assessment of treatment response. Improvements in epidemiological knowledge of the population through the introduction of quantitative diagnostic tests' markers are also expected. Biobanks will create knowledge databases from representative cases, using semantic and semiotic advanced searches. They will help in the development of new imaging markers of disease progression and therapeutic interventions, capable of validation and subsequent standardization, establishing an attractor for pharmaceutical and medical imaging devices to include value-added companies in clinical trials.

Conclusion

Medicine and computing are dynamic fields. Radiologists must have the skills to think critically. Biomarkers are any validated disease characteristic that can be reliably measured in a cost-effective, repeatable, and generalizable manner. These properties act as a meaningful surrogate for disease presence, activity, or outcome [6]. Biomarkers are important in that they can enable faster clinical trials for interventions, improve understanding of early disease processes, and help healthcare practitioners and patients make decisions.

Biomarkers can be either in vitro specimens, such as those from biologic fluids or pathologic samples, and in vivo imaging measurements, such as tumor pharmacokinetics. Imaging biomarkers monitor and record the spatiotemporal distribution of different tissue processes for diagnostic or therapeutic applications, allowing the advanced visual representation, characterization, and quantification of these biological processes within intact living organisms.

Digital medical imaging and computational processing allow for the extraction of quantitative parameters that might be considered as virtual biopsies based on imaging biomarkers. Quantitative imaging biomarkers offer considerable capabilities in both patient care and drug development but need appropriate clinical evidence to be employed. To be implemented in clinical practice, biomarkers may be useful to the diagnostic, therapeutic, and follow-up processes. To be useful within the clinical care pathways, a quantitative statement of its uncertainty must accompany imaging measurements.

Radiologists and biomedical engineers must check the integrity of the whole development cycle, from the conception to the execution, through the creation and implementation of a process of standardization and harmonization of methods, hardware, and/or software that is sufficient for the development, validation, qualification, and use of accurate, reliable quantitative imaging biomarkers across different imaging instruments and settings. The combination of biological knowledge, digital imaging, and computing analysis will lead to professional radiological success, where imaging will lead an important part of clinical and experimental medicine.

Integrating imaging biomarkers and radiomics data into clinical practice presents a unique opportunity to contribute to the radiology value chain [13]. A multidisciplinary approach will provide better care to patients and better understanding of the disease in vivo by in silico image modeling. Imaging industry collaboration and the effectiveness of working relationships are to the advantage of all involved stakeholders. One of the major challenges limiting biomarker's development in radiology is a lack of communication between those in academics and industry.

Modern medicine depends on biomarkers. Imaging biomarkers open wide new fields of research and bring new potential clinical applications of medical imaging.

References

1. Abramson RG, Su PF, Shyr Y. Quantitative metrics in clinical radiology reporting: a snapshot perspective from a single mixed academic-community practice. Magn Reson Imaging. 2012;30:1357–66.
2. Alberich-Bayarri A, Martí-Bonmatí L, Sanz-Requena R, Belloch E, Moratal D. In vivo trabecular bone morphologic and mechanical relationship using high-resolution 3-T MRI. AJR Am J Roentgenol. 2008;191:721–6.
3. Alberich-Bayarri A, Martí-Bonmatí L, Angeles Pérez M, Sanz-Requena R, Lerma-Garrido JJ, García-Martí G, Moratal D. Assessment of 2D and 3D fractal dimension measurements of trabecular bone from high-spatial resolution magnetic resonance images at 3 T. Med Phys. 2010;37:4930–7.
4. Ashton E. Quantitative MR, in multi-center clinical trials. J Magn Reson Imaging. 2010;31:279–88.
5. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints. Clin Pharmacol Therap. 2001;69:89–95.
6. Buckler AJ, Bresolin L, Dunnick NR, et al. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology. 2011;259:875–84.
7. Carlos RC, Buist DSM, Wernli KJ, et al. JACR 2012. Patient-centered outcomes in imaging: quantifying value. J Am Coll Radiol. 2012;9:725–8.

8. Durá E, Domingo J, Ayala G, Martí-Bonmatí L. Evaluation of the registration of temporal series of contrast-enhanced perfusion magnetic resonance 3D images of the liver. Comput Methods Programs Biomed. 2012;108:932–45.
9. European Society of Radiology Working Group on Imaging Biomarkers, van Beers B, Cuenod CA, Martí-Bonmatí L, Matos C, Niessen W, Padhani A. White paper on imaging biomarkers. Insights Imaging. 2010;1:42–5.
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2015;18:151169.
11. Herold CJ, Lewin JS, Wibmer AG, et al. Imaging in the age of precision medicine: summary of the proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology. Radiology. 2016;279(1):226–38.
12. Kang SK, Fagerlin A, Braithwaite RS. A roadmap for personalized care in radiology. Radiology. 2015;277:638–43.
13. Krishnaraj A, Weinreb JC, Ellenbogen PH, Allen Jr B, Norbash A, Kazerooni EA. The future of imaging biomarkers in radiologic practice: proceedings of the thirteenth annual ACR Forum. J Am Coll Radiol. 2014;11:20–3.
14. Kurlanda BF, Gerstnerb ER, Mountzc JM. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging. 2012;30: 1301–12.
15. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–48.
16. Lee CI, Forman HP. What we can and cannot see coming. Radiology. 2010;257:313–4.
17. Manjón JV, Lull JJ, Carbonell-Caballero J, García-Martí G, Martí-Bonmatí L, Robles M. A nonparametric MRI inhomogeneity correction method. Med Image Anal. 2007;11:336–45.
18. Manjón JV, Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging. 2010;31:192–203.
19. Martí-Bonmatí L, José Lull JJ, García-Martí G, et al. MR analysis of the coincidence between functional and morphological abnormalities in schizophrenic patients with chronic auditory hallucinations. Radiology. 2007;244:549–56.
20. Martí-Bonmatí L, Sanz-Requena R, Alberich-Bayarri A. Pharmacokinetic MR analysis of the cartilage is influenced by field strength. Eur J Radiol. 2008;67:448–52.
21. Martí-Bonmatí L, Alberich-Bayarri A, Sánchez-González J. Overload hepatitides: quanti-qualitative analysis. Abdom Imaging. 2012;37:180–7.
22. Martí-Bonmatí L, Alberich-Bayarri A, García-Martí G, et al. Imaging biomarkers, quantitative imaging, and bioengineering. Radiologia. 2012;54:269–78.
23. Nilsen LB, Fangberget A, Geier OM, Engebraaten O, Borgen E, Olsen DR, Seierstad T. Associations between tumor vascularization assessed by in vivo DCE-MRI and the presence of disseminated tumor cells in bone marrow in breast cancer patients at the time of diagnosis. J Magn Reson Imaging. 2014;40(6):1382–91.
24. Pomar-Nadal A, Pérez-Castillo C, Alberich-Bayarri A, García-Martí G, Sanz-Requena R, Martí-Bonmatí L. Integrating information about imaging biomarkers into structured radiology reports. Radiologia. 2013;55:188–94.
25. Sanz-Requena R, Revert-Ventura A, Martí-Bonmatí L, Alberich-Bayarri A, García-Martí G. Quantitative MR perfusion parameters related to survival time in high-grade gliomas. Eur Radiol. 2013;23:3456–65.
26. Schuster DP. The opportunities and challenges of developing imaging biomarkers to study lung function and disease. Am J Respir Crit Care Med. 2007;176:224–30.
27. Smith JJ, Sorensen AG, Thrall JH. Biomarkers in imaging: realizing radiology's future. Radiology. 2003;227:633–8.
28. Sullivan DC. Imaging as a quantitative science. Radiology. 2008;248:328–32.
29. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. Radiology. 2015;277:813–25.
30. Thornbury JR. Clinical efficacy of diagnostic imaging: love it or leave it. AJR Am J Roentgenol. 1994;162:1–8.
31. Tofts PS, Collins DJ. Multicentre imaging measurements for oncology and in the brain. Br J Radiol. 2011;84 Spec No 2:S213–26.

3 Defining the Biological Basis and Clinical Question (Proof of Concept); Looking for the Interrelationship (Proof of Mechanism)

Fabian Bamberg, Mike Notohamiprodjo, Ulrich Kramer, Marius Horger, and Konstantin Nikolaou

3.1 Introduction

The process of implementation, validation, and standardization of imaging biomarkers poses a major challenge for the radiological community, for a number of reasons. For any newly developed (imaging) biomarker, the underlying pathophysiological mechanism has to be understood, and the corresponding biomarker has to be a valid and reproducible surrogate for this pathophysiological process (proof of concept and proof of mechanism). Hence, the role of the radiological subspecialty societies (e.g., for cardiac, musculoskeletal, or oncologic imaging) is significant, as only by expert knowledge on pathology, diseases, and corresponding imaging techniques and applications, a sustainable development of imaging biomarkers can be propagated. Subspecialty imaging will have to solve the main and relevant issues in the process of developing reliable and reproducible imaging biomarkers, from standardization of image acquisition, standardization of image analysis, quality control, and correct clinical use and implementation.

In the following chapter, examples will be given for established and emerging imaging biomarkers, explaining the concept of evolving and implementing a biomarker, starting from the clinical need and clinical question to be solved, the underlying pathophysiological process, and the necessary steps of clinical implementation and standardization.

3.2 Imaging Biomarker of Atherosclerosis: Coronary Artery Calcification

Cardiovascular disease is a leading cause of morbidity and mortality worldwide. Many novel imaging methods have been developed to study atherosclerosis in patients suffering from coronary artery disease (CAD). Imaging techniques that will allow identifying progressive cardiac disease and prediction of future clinical risk are becoming of more and more importance. Among many potential cardiac imaging biomarkers, coronary artery calcification (CAC) is certainly among the most established biomarkers in clinical routine yet. Coronary calcification can be

F. Bamberg • M. Notohamiprodjo • U. Kramer
M. Horger • K. Nikolaou (✉)
University Hospital of Tuebingen, Department of Diagnostic and Interventional Radiology, Tuebingen, Germany
e-mail: konstantin.nikolaou@med.uni-tuebingen.de

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_3

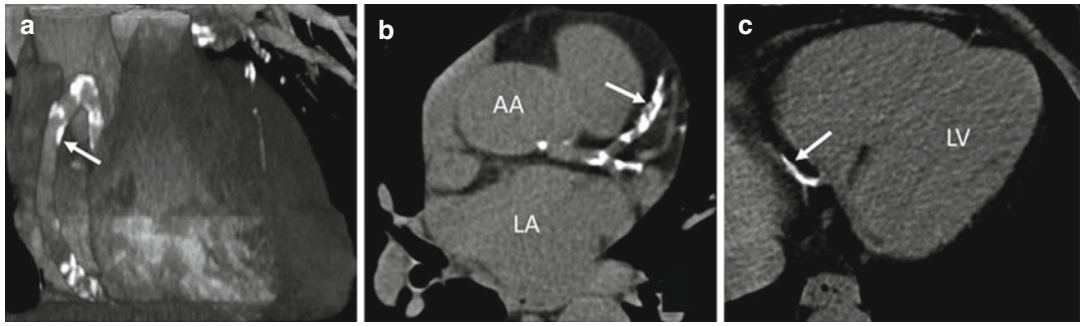


Fig. 3.1 Example of a native coronary artery (CAC) scan acquired by multidetector CT in a 62-year-old male asymptomatic subject. (**a**) Three-dimensional volume-rendered display demonstrating the distribution of the calcified atherosclerotic plaque in the right coronary artery (*arrow*). (**b**, **c**) Axial reconstructions used for quantification of CAC demonstrating extensive CAC in the left anterior descending coronary artery and the distal segment of the right coronary artery (*arrow*, **b** and **c**, respectively). *AA* ascending aorta, *LA* left atrium, and *LV* left ventricle. The Agatston score was 452 in this subject, which puts him on the 57th percentile of his matched age and gender group

assessed noninvasively by computed tomography (CT) and represents a valuable estimate of the presence and extent of coronary atherosclerosis as the underlying cause of stable and acute coronary syndromes. The broad acceptance, both in a clinical and scientific community, is in part attributable to the relatively long period of time (~25 years) this marker of coronary atherosclerosis has been studied but also to the very high level of scientific evidence demonstrating that the degree of coronary calcification is an independent risk factor for cardiovascular events [1]. Also, the widespread availability of CT scanners feasible for CAC scoring without the need for highly advanced CT technology and early methodological, widely accepted definitions of parameters, such as the conventional use of the "Agatston" score, has helped to disseminate this imaging biomarker on a broader basis (Fig. 3.1).

3.2.1 Pathophysiological Correlate and Clinical Problem

Subclinical coronary calcification commonly occurs early in the development of atherosclerotic plaque, preceding the onset of clinical coronary heart disease by years or even decades, therefore serving as an ideal biomarker depicting the degree of atherosclerosis. Calcium is not present in normal arterial vessel walls but is present in atherosclerotic vessels. Previous studies indicate that the amount of calcified atherosclerotic plaque is highly related to the overall plaque burden and is thought to be present in the advanced stages of atherosclerosis [2]. Conversely, noncalcified plaque is considered to be a feature of early atherosclerosis [3] and may be associated with acute coronary syndromes [4]. The pathophysiologic process leading to coronary calcification is a complex one and centrally mediated by inflammation. Current evidence suggests that atherosclerosis is a systemic chronic inflammatory condition involving multiple vascular distributions, associated with a variety of clinical sequels [5]. Through a process of plaque vulnerability, rupture, and subsequent thrombosis, subclinical end-organ damage occurs, as well as clinical evident morbidity and mortality.

3.2.2 Methodological Approaches and Distribution in Normal Cohorts

With the advent of CT scanners with ECG-triggered acquisitions in the late 1990s, electron beam CT (EBCT) followed by ≥4 slice multidetector CT enabled measurement of CAC within a single breath hold [6]. The high-density contrast between calcification and adjacent tissue allowed for quantification of CAC, despite that the detection and

quantification of noncalcified plaque and coronary stenosis on CT angiography (CTA) remained challenging using these first scanner generations.

Given the superb contrast induced by coronary calcification, scans are acquired without the administration of iodinated contrast agents. Traditionally, these scans are associated with a radiation dose between 1 and 2 mSv but lowered to even below 1 mSv when applying a tube voltage of 100 kVp rather than 120 kVp. Recently, additional radiation dose-reduction strategies based on iterative reconstruction have been introduced, resulting in radiation dose reductions of up to 80 % with an effective radiation dose between 0.15 and 0.18 mSv [7].

Initially developed for EBCT, the "Agatston" score, remained the traditional biomarker of CAC as a semiquantitative algorithm based on measurements of plaque area and a weighted plaque density score [8]. This results in a range of continuous values ranging from no calcification (consistent with a zero Agatston score) to extensive calcification that may add up to an Agatston score of more than 3,000. Due to the early implementation in large asymptomatic cohort studies, a normal distribution of the Agatston score with respect to age and gender became available [9]. This allowed for the determination of the percentile distribution of an individual Agatston score with respect to an age- and gender-matched reference cohort and the definition of risk categories [10]. One prominent example is the Multi-ethnic Study of Atherosclerosis (MESA), a large US study that allows to obtain the estimated probability of nonzero calcium, as well as the 25th, 50th, 75th, and 90th percentiles of the CAC distribution for a particular age, gender, and race (http://www.mesa-nhlbi.org/calcium/input.aspx) [11].

3.2.3 Scientific Evidence of Coronary Calcium Score as a Diagnostic and Prognostic Biomarker

Despite its early establishment, CAC scoring was only accepted as a screening tool for asymptomatic subjects within the cardiovascular community after strong data on its prognostic value had been accumulated. In a comprehensive meta-analysis pooling four major studies on the prognostic value of coronary calcification, Pletcher et al. found an adjusted relative risk of death or myocardial infarction of 2.1 for a CAC score of 1–100 compared to a score of 0 [12]. In addition to its generally strong prognostic value, there is similarly high level of evidence that this prognostic value is independent of other markers of risk. For instance, in the MESA study, CAC was compared to high-sensitive CRP, carotid intima-media thickness, ankle brachial index, brachial flow-mediated dilation, and family history of CHD in 1330 individuals without diabetes mellitus [13]. While each of the individual markers was predictive as indicated by an improvement of the area under the receiver-operator characteristic curve (AUC) over a clinical risk score (Framingham Risk), coronary calcification was the strongest predictor of risk associated with the highest increment of the AUC. As a consequence, the 2010 American College of Cardiology/American Heart Association (ACC/AHA) guidelines on screening for coronary artery disease indicated that measurement of CAC is reasonable (level of evidence B) for cardiovascular risk assessment in asymptomatic adults at Framingham intermediate risk (10–20 % 10-year risk) [14].

While the prognostic value in asymptomatic patients is certainly dominating the clinical role of CAC as an established imaging biomarker, there are research findings from the early years documenting a moderately valuable diagnostic role for the detection of significant coronary stenosis. For instance, in a review of 16 studies, the sensitivity and specificity of EBCT for the detection of coronary stenosis were 91 and 49 %, respectively [15]. However, despite these very nonspecific findings, the absence of CAC is highly predictive of the absence of significant stenosis (<1 %) [16]. With the advent of coronary CT angiography, the diagnostic role of coronary artery calcification scanning has diminished, given that even in subjects without coronary calcification, a portion has a plaque and a significant coronary stenosis (13 % and 3.5 %, respectively) [17].

3.2.4 Future Outlook and Developments

Today, assessment of coronary calcification by CT is a highly established biomarker of cardiovascular risk in asymptomatic subjects. However, one of its limitations is that the effectiveness of implementing CAC scanning in therapeutic algorithms has not been established yet in randomized trials. So far, very early evidence stems from the St. Francis Heart Study, in which healthy men and women with CAC scores ≥80th percentile for age and gender were treated with atorvastatin, vitamins C and E, or matching placebos [18]. However, there was only a nonsignificant trend toward a lower rate of cardiovascular events. Due to these limited and nonsignificant findings, pharmacologic therapy to prevent CHD based solely upon the presence of CAC is currently not recommended.

3.3 Imaging Biomarker of Myocardial Fibrosis: Late Gadolinium Enhancement MRI

Up to now, cardiac biomarkers in clinical routine have been dominated by protein molecules that are widely used in the (early) detection of heart failure. For example, serum levels of troponin T (TnT), C-reactive protein (CRP), and brain natriuretic peptide (BNP) have been shown to be sensitive markers of left ventricular (LV) dysfunction and powerful markers of morbidity and mortality in the heart failure setting. One of the current methods of monitoring LV function in patients suffering from heart failure is determination of left ventricular ejection fraction (LVEF) with echocardiography. However, assessment of LVEF is dependent on hemodynamic conditions and fails to detect early subtle alterations in LV systolic function that occurs in the later stages of chronic disease. In addition to echocardiography, cardiac magnetic resonance imaging (MRI) might be used for the noninvasive assessment of LV volumes and LVEF in a heart failure setting [19]. Improvements in both spatial and temporal resolution have made cardiac MRI the gold standard for the noninvasive assessment of LV systolic function, also in patients after myocardial infarction [20].

3.3.1 Clinical Problem

One of the main advantages of cardiac MRI over other imaging modalities is its ability to qualitatively and quantitatively assess changes in myocardial tissue. For example, by using contrast-enhanced pulse sequences, cardiac MRI can be used to assess myocardial viability. Moreover, in myocarditis, this has been used to assess edema, capillary leakage, hyperemia, and, in severe cases, cellular necrosis and fibrosis. This technique has displayed high diagnostic accuracy for acute inflammatory or ischemic injury, which is significant as edema is an important hallmark of inflammatory injury. Nowadays, late gadolinium enhancement (LGE) imaging has become the primary diagnostic tool for assessment of myocardial viability in patients suffering from CAD as well as myocardial inflammation in patients with suspected myocarditis of cardiomyopathies [21]. Furthermore, cardiac stress MRI using adenosine might add a functional component to late gadolinium enhancement imaging, which detects irreversibly damaged myocardium (e.g., scar tissue) with very high accuracy. With LGE, hypo- or akinetic but still viable myocardium can be identified as dysfunctional myocardium without scar or significant remaining viable tissue (<50 % transmurality of scarring). In clinical practice, this additional information will help to decide whether or not the patient will recover after revascularization. However, different noninvasive tests provide different surrogate definitions of hibernating myocardium (e.g., metabolism, scar, or contractile reserve), and consequently the result of a given investigation may depend on the chosen imaging test.

3.3.2 Pathophysiological Correlate

The mechanism of LGE in acute and chronic infarction is the increase of the extracellular

space caused by necrosis (loss of cell membrane integrity), resulting in a significant increase of the extracellular space of scar tissue as compared to normal myocardium. Magnetic resonance contrast agent that diffuse to the interstitial space will be resorbed into the capillary bed and undergo renal excretion. However, when the tissue is damaged, for example, due to infarction, diffuse fibrosis, or even inflammation, the resorption rate of contrast agent will be diminished. At 10–20 min after contrast injection, washout of the contrast agent will be complete in normal myocardium, in contrast to infarcted or edematous tissue. This phenomenon is the basis of "late gadolinium enhancement" imaging. Various studies have shown the relation between myocardial viability and the size of the area displaying late gadolinium enhancement. In MR images, the presence of contrast agent can be detected as a bright area on images acquired with T1-weighted MR images. Currently the principle "bright is dead," indicating that bright areas on a late gadolinium MR image after contrast injection correspond with nonviable myocardium, is subject to a lively debate. Nevertheless, a strong correlation between the transmural extent of hyperenhancement and regional function recovery has been demonstrated in several studies, revealing LGE imaging as a powerful predictor of myocardial damage after myocardial infarction.

The mechanism in nonischemic myocardial diseases is also predominantly, if not exclusively, due to increase in the distribution volume of gadolinium, caused by an increase in the extracellular space of the myocardium. The increase in the extracellular space may be due to necrosis of myocardial cells (myocarditis), replacement of myocardial cells by fibrosis (hypertrophic cardiomyopathy), or replacement of myocardium by various infiltrates (e.g., amyloidosis, sarcoidosis), or a combination of these.

While the distribution of LGE is invariably subendocardial for ischemic disease (either acute or chronic myocardial infarction), the pattern of enhancement has a variable transmural distribution in nonischemic myocardial disease. LGE conforms to the distribution of one or more coronary arteries in ischemic heart disease, while in nonischemic myocardial disease, it does not. With nonischemic myocardial disease, the mural pattern usually can be seen midwall or subepicardial and may be spotty in multiple regions of the ventricle. The pattern of distribution in some may be almost diagnostic of a specific myocardial disease.

3.3.3 Scientific Evidence and Prognostic Impact

Within the bright LGE area of an acute myocardial infarction, microvascular obstruction (MO) or the "no-reflow" phenomenon is known as an established complication of coronary reperfusion therapy, adding another complementary "imaging biomarker" to the clinical value of LGE (Fig. 3.2) [22–24]. The phenomenon of microvascular obstruction is increasingly recognized as a poor prognostic indicator and can serve as an imaging marker of subsequent adverse LV remodeling [25]. Although MO can be assessed using various imaging modalities, evaluation by cardiac MRI is particularly useful in enhancing its detection, diagnosis, and quantification, as well as following its subsequent effects on infarct evolution and healing. MO assessment has become a routine component of the MR evaluation of acute myocardial infarction and will play an increasingly important role in therapeutic decision pathways. MO is characterized by a number of ultrastructural and functional changes at the microvascular level. Understanding these histopathophysiologic changes can enhance the approach by MRI to detecting MO, to understand the results and their subsequent clinical implications. It can also help to potentially improve how MO is assessed by CMR, which has implications with regard to the understanding of infarct evolution as well as the evaluation of the efficacy and mechanism of reperfusion treatment.

LGE and MO have been found to be predictive of clinical outcome, independently of or when adjusted for other indices such as infarct size and LV ejection fraction (EF) [26]. Many of these outcome studies showed a severe relationship between the size of LGE and/or the presence of

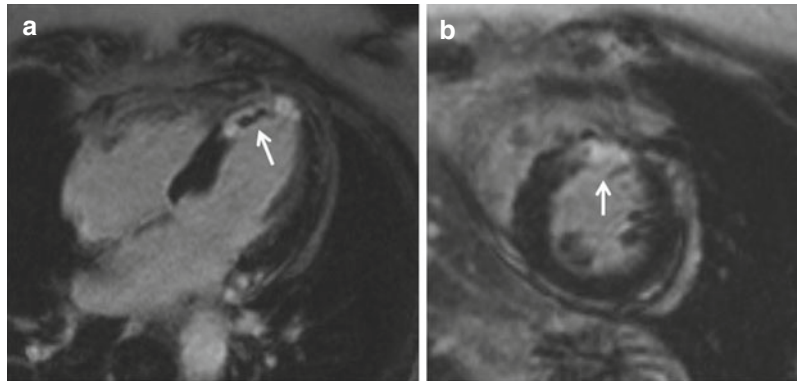


Fig. 3.2 A 77-year-old female patient suffering from an acute myocardial infarction. Invasive catheter coronary angiography showed complete occlusion of the left anterior descending artery (LAD), with following revascularization of the LAD territory after successful PTCA and stent placement. In the four-chamber view of the late gadolinium enhancement (LGE) image (**a**), a myocardial infarction can be seen by pathological uptake of the contrast agent (*arrow*) and with an area of microvascular obstruction in the center of the infarcted myocardium (black central area within the LGE area). Short-axis late gadolinium enhancement image through the mid-ventricle at the level of the papillary muscles (**b**) demonstrates the corresponding myocardial delayed enhancement in the anterior-septal wall segments (*arrow*)

MO and adverse LV remodeling, with reduced global systolic function and pathologically enlarged LV volumes at follow-up exams, suggesting a possible mechanism for the poor prognosis. Moreover, MO is increasingly incorporated into clinical trial methodology as a surrogate clinical outcome in studies investigating reperfusion strategies, particularly when the treatment has the potential to directly impact microvascular function.

3.4 Imaging Biomarker of Tumor Angiogenesis: Assessment of Perfusion

Angiogenesis is a complex process and a constant companion of normal tissue development and tissue formation but shows certain alterations in pathologic processes like tumor growth and/or inflammatory processes. Perfusion-based imaging enables objectification and quantification of angiogenesis and thus opens insights into a functional process generating surrogate biomarkers that are complementary to, e.g., morphologic tumor imaging. The knowledge that angiogenesis is a prerequisite for tumor growth and tumor spread makes this biomarker a plausible target not only for tumor and tissue characterization purposes but also for prognosis evaluation and treatment monitoring. In case of inflammatory processes, perfusion-based imaging allows indirect assessment of inflammatory activity which is usually going along with an increased tissue perfusion and offers potential tool for monitoring anti-inflammatory therapy regimens.

The most frequently used parameters, or quantitative imaging biomarkers of tissue perfusion, are the blood flow (the rate of blood passing through the vasculature in a tissue region), the blood volume (the volume of blood that is actually following within the vasculature), and the mean transit time (the average time of blood traversing from the arterial input to the venous outlet and k-trans (a surrogate measure of vascular leakiness reflecting the flux of solutes from blood plasma to the interstitial space)). Each of these parameters represents a facet of the complex process of angiogenesis and can be used for tumor characterization, for understanding of the mechanisms of action of molecular (antiangiogenic) therapeutic agents, and thus for a potentially more sensitive response monitoring (early detection of response and/or early recognition of tumor breakthrough) (Fig. 3.3).

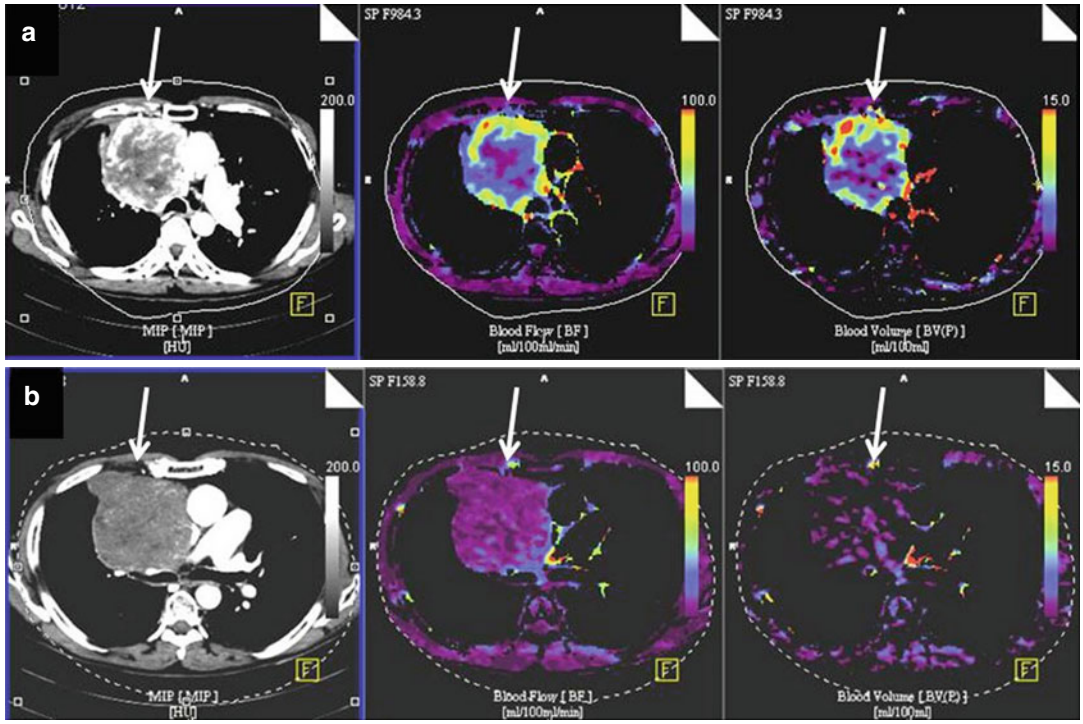


Fig. 3.3 Solitary fibrous pleural tumor (Panel **a** and **b**, *arrows*), receiving experimental therapy with bevacizumab (VEGFR inhibitor) and temozolomide as sixth-line treatment attempt. Perfusion CT performed at baseline (*upper panel*, **a**) shows high levels of tumor perfusion (from left to the right morphologic image/blood flow [BF] color-coded map/blood volume [BV] color-coded map) at the tumor edges with large core necrosis. Six weeks later, during ongoing therapy, a typical on/off angiogenic switch could be observed (*lower panel*, **b**)

3.4.1 The Clinical Problem and Pathophysiological Correlate

Recent developments in tissue characterization in terms of histological differentiation, grading, but in particular immunohistochemical profiling (i.e., receptor expression) as well as genome typing, needing invasive procedures such as biopsies, have somewhat interfered with the constant ambition of establishing and optimizing imaging biomarkers for noninvasive tumor characterization. Nonetheless, relevant clinical applications focusing on perfusion-based tumor identification and characterization are being developed and constantly enhanced. One of the most important clinical applications for perfusion imaging is the characterization of hepatic nodules occurring in patients with known liver cirrhosis, fatty liver, hepatitis C, or hemochromatosis. In these cases, current diagnostic guidelines recommend a noninvasive HCC diagnosis based on the presence of typical enhancement patterns like washin and washout phenomena, in contrast-enhanced CT or MR imaging. Unfortunately, these typical qualitative enhancement patterns are not always present in HCC nodules, and potentially being missed, in particular in smaller lesions. In this clinical setting with unclear imaging results, an invasive approach, i.e., biopsy, is usually recommended. The rationale for implementing a quantitative, perfusion-based imaging biomarker of HCC and of HCC precursors is the knowledge that along the so-called multistep process of carcinogenesis, liver nodules create an arteriolar network, gradually replacing the normal sinusoidal architecture and thus becoming predominantly or exclusively supplied by arterial blood [27]. Hence, with ongoing tumor differentiation, the number of so-called non-triadal arteries is

increasing which has been intensively analyzed by comparison with histology [28]. At this point, the use of perfusion imaging (e.g., perfusion CT) with the option for a separate calculation of the dual liver blood supply (arterial and portal venous blood supply) offers an ideal tool for a noninvasive classification of liver lesions [29]. Additionally, differences in the perfusion between parenchymatous organs like the liver and metastases are known to improve tumor detection both for hypervascularized and hypovascularized lesions, based on the fact that all these lesions are primarily supplied by arterial blood [30]. In several tumors, a direct correlation between them and other invasive histological biomarkers of angiogenesis like microvessel density and VEGF values has been reported [31, 32]. Despite the need for a histologic proof in the primary diagnosis, detection of tumor relapse could benefit from noninvasive imaging-based characterization, avoiding unnecessary risks related to invasive diagnosis. In particular in anatomical region like the CNS, avoidance of open biopsy has emerged to an issue of great debate. A number of studies have used perfusion CT for tumor grading showing that BF, BV, and k-trans values were higher in high-grade glioma vs. low-grade glioma [33]. Moreover, differentiation of high-grade gliomas from other brain lesions including lymphomas has been reported based on differences in BF and BV [34].

3.4.2 Perfusion Imaging as a Biomarker for Prognosis and Response

Lately, perfusion parameters have been increasingly used to predict overall survival after therapy or to predict response to different treatment regimens [35]. Some authors reported about the potential association between the degree of tumor vascularization and tumor aggressiveness [36]. An increased local failure rate was reported in squamous cell carcinoma of the head and neck presenting with low perfusion values before treatment [37]. Concordantly, a better response to both chemotherapy and radiation therapy was described in tumors exhibiting higher BF values in the baseline [38]. Early efficacy prediction of sorafenib in the treatment of hepatocellular carcinoma (HCC) could be demonstrated in rats [39]. Pretreatment perfusion values were found to be a good predictor in many other tumors undergoing antiangiogenic therapy [40].

A response biomarker may be defined as a characteristic that can be objectively measured as an indicator of pharmaceutical response. The way and magnitude by which tumor vascularization is affected by a certain drug depends both on the characteristics of the tumor's own vessel network, that of its microenvironment as well as on the type of antitumoral agent applied. Knowledge about all these variables helps for adequate use of perfusion measurements considering any potential clinical setting. In the long term, most effective therapy regimens lead to a reduction in perfusion CT parameters. Even during standard chemotherapy, cell death is expected to be followed by a loss in angiogenic support leading to measurable hypovascularization. In particular, of interest are the angiogenesis inhibitors which in many cases are administered as monotherapy. Their antiangiogenic effects depend on the mechanism of action of the drug applied as well as on the specific tumor entity, whereas morphologic changes are generally not expected at least in the early phases of treatment. Despite concurrent imaging modalities including positron emission therapy using different radiopharmaceuticals, targeting the course of tumor vascularization seems to be the more accurate way of response assessment during antiangiogenic treatment. Moreover, perfusion parameters are also suited for confident response monitoring for both local antivascular therapies like chemoembolization and systemic treatment. Even during radiation therapy, changes in perfusion characteristics can be predictive for progressive-free survival and outcome [41]. Perfusion CT also enhances our understanding of the pharmacodynamics of novel drugs and how they should be combined with each other in order to achieve best therapeutic effects. In this respect, one good example is the class of vascular disrupting agents that target also the mature vasculature potentially impacting the arrival time and dose of

other complementary-given chemotherapeutics. Using perfusion as a complementary biomarker helps for developing new more proper response criteria, avoiding confusion caused by newly described response categories like pseudo-response, pseudostabilization, or even pseudo-progression, which severely affect patients' management [42]. Many of the so-called targeted drugs directly or indirectly affecting the signaling pathways of angiogenesis, like EGFR, VEGFR, and PDGF inhibitors, interferon, proteasome inhibitors, and thalidomide and successor drugs, have already been extensively analyzed by means of perfusion imaging techniques [43, 44].

In conclusion, perfusion-based imaging is a potent imaging tool which requires profound understanding of tumor behavior and mechanisms of action of the drugs to be monitored. Once these aspects have been elucidated, its broad clinical use may be advocated.

3.5 Imaging Biomarker of Osteoarthritis: T2 Relaxation Time of the Articular Cartilage

3.5.1 Clinical Background

Osteoarthritis is a common degenerative joint disease, affecting more than 20 million people in the USA alone. The incidence of osteoarthritis is rising, and in 2030, at least 70 million people (USA) will be affected [45]. Osteoarthritis-related costs are a relevant socioeconomic factor. The established therapy of manifest osteoarthritis is cartilage repair, cartilage transplantation, or arthroplasty, while lifestyle modification is currently the only method of potential disease prevention [46]. Disease-modifying therapies which may prevent or prolong the course of osteoarthritis have been developed and evaluated but are still not broadly approved for clinical routine.

The only endpoint for clinical trials concerning "imaging biomarkers" accepted by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) is the quantitative measurement of joint space width [47]. However, this parameter does not depict the cartilage itself and is obviously limited in assessing discrete changes of cartilage composition and cartilage quality during disease development and under disease-modifying treatments. However, early detection of potentially reversible biochemical degradation is of high interest, as consecutive morphological changes are typically irreversible. Similarly, short-term variation in symptoms is not sufficiently reflected in radiographs [48].

Magnetic resonance imaging (MRI), using moderately T2-weighted fast spin-echo or T1-weighted gradient-echo sequences, is superior to any other cross-sectional imaging method for direct visualization of articular cartilage. Several semiquantitative MRI scoring systems for osteoarthritis have been developed, such as the Knee Osteoarthritis Scoring System [49] and the Whole-Organ Magnetic Resonance Imaging Score [50]. Furthermore, assessment of cartilage thickness and volume has been evaluated and proposed as potential biomarkers in the development of osteoarthritis [51].

3.5.2 Underlying Pathophysiology

Measurement of cartilage thickness and cartilage volume only provides information on the tissue morphology and not on the biochemical composition of the joint cartilage. To establish relevant surrogate imaging biomarkers for osteoarthritis, one has to understand the histological composition of joint cartilage and the pathophysiology of osteoarthritis.

The main components of hyaline cartilage are water and the collagenous extracellular matrix. Hyaline articular cartilage is largely acellular, and only 4% of its wet weight are formed by chondrocytes, which produce the extracellular matrix. Histologically, cartilage is classified in three types, i.e., as *elastic cartilage*, which can be found, e.g., in the outer ear or epiglottis, the glossy *hyaline cartilage* of the joint surfaces, and the rigid and dense *fibrocartilage*, which is found in menisci and the annulus fibrosus of the intervertebral disks.

The main function of hyaline cartilage is to provide adequate cushioning and gliding of the articular joint surfaces. Its function is dependent on the molecular composition of the extracellular matrix, which mainly consists of proteoglycans and collagen type 2. Fibrous cartilage is the only histologic subtype to additionally contain collagen type 1. The main proteoglycan is aggrecan, forming large aggregates with hyaluronan. These aggregates are negatively charged and hold water in the tissue and are constrained by the collagen. The collagen has a specific arrangement in the tissue, beginning at the surface as a zone of tangentially aligned fibril sheets. Below this is a transitional zone of randomly aligned fibrils, followed by a deep radial zone with fibrils aligned perpendicular to the articular surface. The presence of a high proteoglycan concentration immobilized in the collagen network creates a high osmotic pressure drawing water molecules into the tissue. This complex molecular structure provides the unique properties of articular cartilage. Cartilage is resilient and viscoelastic, making it resistant to frictional, compressive, shear, and tensile loading [52]. However, the reparative potential of cartilage as compared to other connective tissues is low. Cartilage has a very slow turnover and does not contain blood vessels, so that the chondrocytes are supplied by diffusion. If damage to hyaline cartilage does affect the blood vessels of the affected bone, the microvascular blood supply will lead to formation of new cartilage. However, the new cartilage filling the defect is fibrous cartilage. Thus, this scar of fibrous cartilage does not provide the ideal physiologic properties for articular surfaces but is still preferable to a full-thickness defect of the cartilage [53].

3.5.3 Establishing Imaging Biomarkers of Cartilage Degeneration and Trauma

Several imaging biomarkers have been proposed to reflect the different components of articular cartilage [53]. The amount of cartilage proteoglycans can be assessed with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC), Na23 imaging, spin-lattice relaxation time in the rotating frame (T1rho), and glycosaminoglycan chemical exchange saturation transfer (GAG-CEST) [54]. Diffusion-weighted imaging allows estimating the apparent diffusion coefficient, which is related to the water content of articular cartilage. The advanced diffusion-weighted imaging technique called “diffusion tensor imaging” (DTI) additionally allows quantifying the diffusion anisotropy of articular cartilage, which is considered to be related to the amount of collagen [55]. Probably, together with dGEMRIC, the most broadly evaluated imaging biomarker of cartilage biochemistry is assessment of the T2 relaxation time, which is related to the amount of water and the collagen structure [56, 57]. In contrast to dGEMRIC, no exogenous contrast agent is required for application of this technique.

For physiological interpretation of T2 relaxation, the pathogenesis of osteoarthritis has to be taken into account. In early stages, tissue damage stimulates a synthetic and proliferative response of the embedded chondrocytes. These may maintain or even restore the articular cartilage. However, in ongoing osteoarthritis, the chondrocyte response declines, leading to cartilage thinning and volume loss. Early degeneration of articular cartilage begins with disruption or alteration of the molecular structure and composition of the extracellular matrix. A loss of proteoglycans and an increase of water concentration occur. Some of the early matrix changes in articular cartilage degeneration include loss of proteoglycans. The size of proteoglycan aggregates decreases significantly with age and joint degeneration due to degradation of the proteoglycans and alteration of proteoglycan synthesis. Thus, the abovementioned aggrecan molecules become shorter, and the mean number of aggrecans in each aggregate decreases. Furthermore, osteoarthritis causes upregulation of collagen cross-linking enzymes. Thus, the collagen network is capable to hold more water molecules, increasing the water content of articular cartilage. Studies have shown that the water content of cartilage affected by osteoarthritis increases

by 10 % [58]. These changes can be monitored using T2-weighted sequences with multiple increasing echoes, allowing to quantify the T2 relaxation time.

Typically, up to 10 echoes are acquired using spin-echo sequences [59]. The signal decay is then fitted mono-exponentially on a voxelwise or ROI basis. Regional distribution is commonly visualized using color-coded parameter maps (Fig. 3.4). Analysis is usually performed taking into account the zonal architecture of articular cartilage. Segmentation of articular cartilage is either performed on the T2-weighted images or on overlying gradient-echo sequences. T2 values of normal cartilage follow the physiological architecture with a significant T2 increase from deep to superficial layers. Early studies have shown that T2 values of articular cartilage increase with aging. Furthermore, osteoarthritis causes a T2 increase correlating with the degree of degeneration. Increased T2 values compared to the uninjured side can be observed after knee surgery, e.g., ACL reconstruction despite normal morphology and volume [60], suggesting that T2 relaxation time may serve as an early predictive marker of joint degeneration [61].

Physiological exercises such as loading and unloading of the respective joint cause alteration of the water content of the cartilage and thus change in T2 relaxation. Cartilage repair tissue shows an even more pronounced response to these maneuvers compared to normal articular cartilage, suggesting that repair cartilage exhibits a different ultrastructure and the quantitative T2 measurements allow for differentiation of normal and abnormal cartilage. In contrast to normal cartilage, repair tissue after autologous chondrocyte transplantation shows no zonal alteration of T2 in early follow-up examinations. However, with progressive healing, a zonal stratification can be observed in later exams [62]. Compared to morphological scores, e.g., magnetic resonance observation of cartilage repair tissue (MOCART), T2 relaxation assessment provides

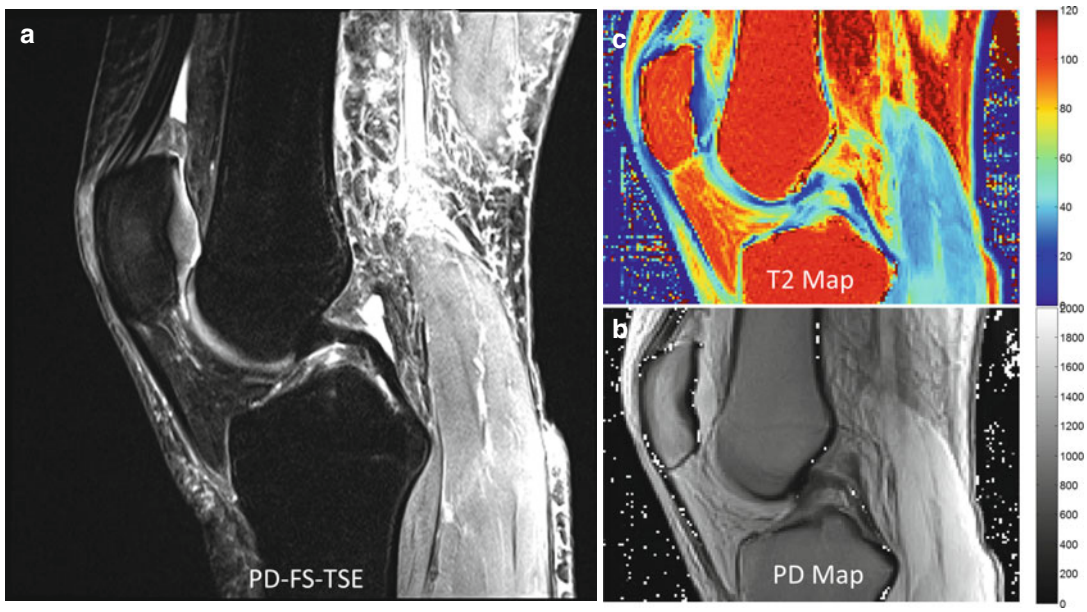


Fig. 3.4 Healthy cartilage shows several zones between the articular surface and bone interface, marked by collagen orientation of collagen, which is highly organized with a high proteoglycan content. Here, morphological proton-density fat-saturated turbo spin-echo (PD_FS_TSE) imaging is shown (**a**), with exemplary PD map (**b**) and T2 map (**c**) of a healthy knee. The zonal differentiation can be readily visualized, particularly for the retropatellar cartilage

additional information regarding the biochemical composition of the repair tissue [63]. Major constraints for implementation of quantitative assessment of T2 relaxation time are the dependency of diagnostic accuracy on the applied fitting method and variation of repeated measurements [64]. The mean coefficient of variant for quantitative T2 relaxation is about 3–10 % [64], so that changes related to disease should exceed these thresholds. As mentioned above, the changes of water content range around these values, so that careful interpretation of the data is warranted.

Concluding, quantitative T2 relaxation time is a noninvasive surrogate biomarker for the water content of articular cartilage, which is dependent on underlying the collagen structure. It is one of the best evaluated parameters for evaluation of cartilage biochemistry and may serve as a valuable endpoint of clinical studies examining potential therapies for osteoarthritis.

Conclusions

In the process of introducing new, noninvasive imaging biomarkers for any kind of clinical entity, treatment, or medical need, the first and most important steps are a thorough understanding of the underlying pathology, the proof of mechanism for a given drug or disease and the corresponding imaging process, and the proof of concept that the underlying pathophysiology is adequately reflected by the newly developed imaging biomarker. Only if this first logical step is successfully taken, the important tasks of biomarker standardization, validation, and reproducibility can be solved. With the examples given in this chapter, a variety of the most important established or emerging imaging biomarkers have been discussed, including cardiovascular, oncologic, and musculoskeletal medical entities. For each of the given biomarkers, the basic pathophysiology and the corresponding proof of concept and proof of mechanism have been explained, and the methodology of imaging biomarker correlation and implementation has been elucidated.

References

1. Taylor AJ, Cerqueira M, Hodgson JM, et al. ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 Appropriate Use Criteria for Cardiac Computed Tomography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. Circulation. 2010;122:e525–55.
2. Rumberger JA, Simons DB, Fitzpatrick LA, Sheedy PF, Schwartz RS. Coronary artery calcium area by electron-beam computed tomography and coronary atherosclerotic plaque area. A histopathologic correlative study. Circulation. 1995;92:2157–62.
3. Bamberg F, Dannemann N, Shapiro MD, et al. Association between cardiovascular risk profiles and the presence and extent of different types of coronary atherosclerotic plaque as detected by multidetector computed tomography. Arterioscler Thromb Vasc Biol. 2008;28:568–74.
4. Hoffmann U, Moselewski F, Nieman K, et al. Noninvasive assessment of plaque morphology and composition in culprit and stable lesions in acute coronary syndrome and stable lesions in stable angina by multidetector computed tomography. J Am Coll Cardiol. 2006;47:1655–62.
5. Naghavi M, Libby P, Falk E, et al. From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: part I. Circulation. 2003;108:1664–72.
6. Arad Y, Spadaro LA, Roth M, et al. Correlations between vascular calcification and atherosclerosis: a comparative electron beam CT study of the coronary and carotid arteries. J Comput Assist Tomogr. 1998;22:207–11.
7. Willemink MJ, den Harder AM, Foppen W, et al. Finding the optimal dose reduction and iterative reconstruction level for coronary calcium scoring. J Cardiovasc Comput Tomogr. 2015;10(1):69–75.
8. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte Jr M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol. 1990;15:827–32.
9. Hoff JA, Chomka EV, Krainik AJ, Daviglus M, Rich S, Kondos GT. Age and gender distributions of coronary artery calcium detected by electron beam tomography in 35,246 adults. Am J Cardiol. 2001;87:1335–9.
10. Raggi P, Callister TQ, Cooil B. Calcium scoring of the coronary artery by electron beam CT: how to apply an individual attenuation threshold. AJR Am J Roentgenol. 2002;178:497–502.

11. McClelland RL, Chung H, Detrano R, Post W, Kronmal RA. Distribution of coronary artery calcium by race, gender, and age: results from the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation. 2006;113:30–7.
12. Pletcher MJ, Tice JA, Pignone M, Browner WS. Using the coronary artery calcium score to predict coronary heart disease events: a systematic review and meta-analysis. Arch Intern Med. 2004;164:1285–92.
13. Yeboah J, McClelland RL, Polonsky TS, et al. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. JAMA. 2012;308:788–95.
14. Greenland P, Alpert JS, Beller GA, et al. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2010;56:e50–103.
15. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. Circulation. 2000;102:126–40.
16. Haberl R, Becker A, Leber A, et al. Correlation of coronary calcification and angiographically documented stenoses in patients with suspected coronary artery disease: results of 1,764 patients. J Am Coll Cardiol. 2001;37:451–7.
17. Villines TC, Hulten EA, Shaw LJ, et al. Prevalence and severity of coronary artery disease and adverse events among symptomatic patients with coronary artery calcification scores of zero undergoing coronary computed tomography angiography: results from the CONFIRM (Coronary CT Angiography Evaluation for Clinical Outcomes: An International Multicenter) registry. J Am Coll Cardiol. 2011;58:2533–40.
18. Arad Y, Spadaro LA, Roth M, Newstein D, Guerci AD. Treatment of asymptomatic adults with elevated coronary calcium scores with atorvastatin, vitamin C, and vitamin E: the St. Francis Heart Study randomized clinical trial. J Am Coll Cardiol. 2005;46:166–72.
19. Won E, Donnino R, Srichai MB, Sedlis SP, Feit F, Rolnitzky L, Miller LH, Iqbal SN, Axel L, Nguyen B, Slater J, Shah B. Diagnostic accuracy of cardiac magnetic resonance imaging in the evaluation of newly diagnosed heart failure with reduced left ventricular ejection fraction. Am J Cardiol. 2015;116(7):1082–7.
20. Palazzuoli A, Beltrami M, Gennari L, Dastidar AG, Nuti R, McAlindon E, Angelini GD, Bucciarelli-Ducci C. The impact of infarct size on regional and global left ventricular systolic function: a cardiac magnetic resonance imaging study. Int J Cardiovasc Imaging. 2015;31(5):1037–44.
21. Klopotowski M, Kukula K, Malek LA, Spiewak M, Polanska-Skrzypczyk M, Jamiolkowski J, Dabrowski M, Baranowski R, Klisiewicz A, Kusmierczyk M, Jasinska A, Jarmus E, Kruk M, Ruzyllo W, Witkowski A, Chojnowska L. The value of cardiac magnetic resonance and distribution of late gadolinium enhancement for risk stratification of sudden cardiac death in patients with hypertrophic cardiomyopathy. J Cardiol. 2016;68:49–56. pii: S0914-5087(15)00253-1.
22. Orn S, Manhenke C, Greve OJ, Larsen AI, Bonarjee VV, Edvardsen T, Dickstein K. Microvascular obstruction is a major determinant of infarct healing and subsequent left ventricular remodelling following primary percutaneous coronary intervention. Eur Heart J. 2009;30:1978–85.
23. Mather AN, Lockie T, Nagel E, Marber M, Perera D, Redwood S, Radjenovic A, Saha A, Greenwood JP, Plein S. Appearance of microvascular obstruction on high resolution first-pass perfusion, early and late gadolinium enhancement CMR in patients with acute myocardial infarction. J Cardiovasc Magn Reson. 2009;11:11–33.
24. Bekkers SC, Backes WH, Kim RJ, Snoep G, Gorgels AP, Passos VL, Waltenberger J, Crijns HJ, Schalla S. Detection and characteristics of microvascular obstruction in reperfused acute myocardial infarction using an optimized protocol for contrast-enhanced cardiovascular magnetic resonance imaging. Eur Radiol. 2009;19:2904–12.
25. de Waha S, Desch S, Eitel I, Fuernau G, Zachrau J, Leuschner A, Gutberlet M, Schuler G, Thiele H. Impact of early vs. late microvascular obstruction assessed by magnetic resonance imaging on long-term outcome after ST-elevation myocardial infarction: a comparison with traditional prognostic markers. Eur Heart J. 2010;31:2660–8.
26. Larsen TH, Stugaard M, Rotevatn S, Nygård O, Nordrehaug JE. Clinical significance of late enhancement and regional wall remodeling assessed by 3T magnetic resonance imaging. Clin Med Insights Cardiol. 2015;9:17–24.
27. Park YN, Yang CP, Fernandez GJ, Cubukcu O, Thung SN, Theise ND. Neoangiogenesis and sinusoidal "capillarization" in dysplastic nodules of the liver. Am J Surg Pathol. 1998;22(6):656–62.
28. Tarhan NC, Hatipoğlu T, Ercan E, Bener M, Keleş G, Başaran C, Bilezikçi B. Correlation of dynamic multidetector CT findings with pathological grades of hepatocellular carcinoma. Diagn Interv Radiol. 2011;17(4):328–33.
29. Kaufmann S, Horger T, Oelker A, Beck S, Schulze M, Nikolaou K, Ketelsen D, Horger M. Volume perfusion computed tomography (VPCT)-based evaluation of response to TACE using two different sized drug eluting beads in patients with nonresectable hepatocellular carcinoma: Impact on tumor and liver parenchymal vascularisation. Eur J Radiol. 2015. pii: S0720-048X(15)30101-7. doi:10.1016/j.ejrad.2015.09.009.
30. Reiner CS, Goetti R, Burger IA, Fischer MA, Frauenfelder T, Knuth A, Pfammatter T, Schaefer N, Alkadhi H. Liver perfusion imaging in patients with primary and metastatic liver malignancy: prospective

comparison between 99mTc-MAA spect and dynamic CT perfusion. Acad Radiol. 2012;19(5):613–21.
31. d'Assignies G, Couvelard A, Bahrami S, Vullierme MP, Hammel P, Hentic O, Sauvanet A, Bedossa P, Ruszniewski P, Vilgrain V. Pancreatic endocrine tumors: tumor blood flow assessed with perfusion CT reflects angiogenesis and correlates with prognostic factors. Radiology. 2009;250(2):407–16.
32. Ash L, Teknos TN, Gandhi D, Patel S, Mukherji SK. Head and neck squamous cell carcinoma: CT perfusion can help noninvasively predict intratumoral microvessel density. Radiology. 2009;251(2):422–8.
33. Jain R, Gutierrez J, Narang J, Scarpace L, Schultz LR, Lemke N, Patel SC, Mikkelsen T, Rock JP. In vivo correlation of tumor blood volume and permeability with histologic and molecular angiogenic markers in gliomas. AJNR Am J Neuroradiol. 2011;32(2):388–94.
34. Jain R, Ellika SK, Scarpace L, Schultz LR, Rock JP, Gutierrez J, Patel SC, Ewing J, Mikkelsen T. Quantitative estimation of permeability surface-area product in astroglial brain tumors using perfusion CT and correlation with histopathologic grade. AJNR Am J Neuroradiol. 2008;29(4):694–700.
35. Jain R, Narang J, Griffith B, Bagher-Ebadian H, Scarpace L, Mikkelsen T, Littenberg B, Schultz LR. Prognostic vascular imaging biomarkers in high-grade gliomas: tumor permeability as an adjunct to blood volume estimates. Acad Radiol. 2013;20(4):478–85.
36. García-Figueiras R, Goh VJ, Padhani AR, Baleato-González S, Garrido M, León L, Gómez-Caamaño A. CT perfusion in oncologic imaging: a useful tool? AJR Am J Roentgenol. 2013;200(1):8–19.
37. Wang J, Wu N, Cham MD, Song Y. Tumor response in patients with advanced non-small cell lung cancer: perfusion CT evaluation of chemotherapy and radiation therapy. AJR Am J Roentgenol. 2009;193(4):1090–6.
38. Hermans R, Meijerink M, Van den Bogaert W, Rijnders A, Weltens C, Lambin P. Tumor perfusion rate determined noninvasively by dynamic computed tomography predicts outcome in head-and-neck cancer after radiotherapy. Int J Radiat Oncol Biol Phys. 2003;57(5):1351–6.
39. Wang Q, Shi G, Wang L, Liu X, Wu R. Early prediction of response of sorafenib on hepatocellular carcinoma by CT perfusion imaging: an animal study. Br J Radiol. 2014;87(1035):20130695. doi:10.1259/bjr.20130695. Epub 2014 Jan 27.
40. Han KS, Jung DC, Choi HJ, Jeong MS, Cho KS, Joung JY, Seo HK, Lee KH, Chung J. Pretreatment assessment of tumor enhancement on contrast-enhanced computed tomography as a potential predictor of treatment outcome in metastatic renal cell carcinoma patients receiving antiangiogenic therapy. Cancer. 2010;116(10):2332–42.
41. Scharpfenecker M, Floot B, Russell NS, Coppes RP, Stewart FA. Thalidomide ameliorates inflammation and vascular injury but aggravates tubular damage in the irradiated mouse kidney. Int J Radiat Oncol Biol Phys. 2014;89(3):599–606. doi:10.1016/j.ijrobp.2014.03.039.
42. Betz M, Kopp HG, Spira D, Claussen CD, Horger M. The benefit of using CT-perfusion imaging for reliable response monitoring in patients with gastrointestinal stromal tumor (GIST) undergoing treatment with novel targeted agents. Acta Radiol. 2013;54(7):711–21.
43. Hygino da Cruz Jr LC, Rodriguez I, Domingues RC, Gasparetto EL, Sorensen AG. Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. AJNR Am J Neuroradiol. 2011;32(11):1978–85.
44. Zhu AX, Holalkere NS, Muzikansky A, Horgan K, Sahani DV. Early antiangiogenic activity of bevacizumab evaluated by computed tomography perfusion scan in patients with advanced hepatocellular carcinoma. Oncologist. 2008;13(2):120–5.
45. Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. Best Pract Res Clin Rheumatol. 2014;28(1):5–15.
46. Cooper C, et al. How to define responders in osteoarthritis. Curr Med Res Opin. 2013;29(6):719–29.
47. Conaghan PG, et al. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthritis Cartilage. 2011;19(5):606–10.
48. Liess C, et al. Detection of changes in cartilage water content using MRI T2-mapping in vivo. Osteoarthritis Cartilage. 2002;10(12):907–13.K.
49. Kornaat PR, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)–inter-observer and intra-observer reproducibility of a compartment-based scoring system. Skeletal Radiol. 2005;34(2):95–102.
50. Peterfy CG, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage. 2004;12(3):177–90.
51. Horng A, et al. Topographic deformation patterns of knee cartilage after exercises with high knee flexion: an in vivo 3D MRI study using voxel-based analysis at 3T. Eur Radiol. 2015;25(6):1731–41.
52. Horng A, et al. Locoregional deformation pattern of the patellar cartilage after different loading types: high-resolution 3D-MRI volumetry at 3 T in-vivo. Rofo. 2011;183(5):432–40.
53. Hunter DJ, et al. Imaging biomarker validation and qualification report: sixth OARSI Workshop on Imaging in Osteoarthritis combined with third OA Biomarkers Workshop. Osteoarthritis Cartilage. 2013;21(7):939–42.
54. Matzat SJ, et al. Quantitative MRI techniques of cartilage composition. Quant Imaging Med Surg. 2013;3(3):162–74.
55. Raya JG, et al. Articular cartilage: in vivo diffusion-tensor imaging. Radiology. 2012;262(2):550–9.
56. Mamisch TC, et al. Quantitative T2 mapping of knee cartilage: differentiation of healthy control cartilage and cartilage repair tissue in the knee with unloading–initial results. Radiology. 2010;254(3):818–26.
57. Mayerhoefer ME, et al. The in vivo effects of unloading and compression on T1-Gd (f) relaxation times in healthy articular knee cartilage at 3.0 Tesla. Eur Radiol. 2010;20(2):443–9.

58. Guilak F, et al. Mechanical and biochemical changes in the superficial zone of articular cartilage in canine experimental osteoarthritis. J Orthop Res. 1994;12(4):474–84.
59. Raya JG, et al. T2 measurement in articular cartilage: impact of the fitting method on accuracy and precision at low SNR. Magn Reson Med. 2010;63(1): 181–93.
60. Arnoldi AP, et al. MRI based volumetric assessment of knee cartilage after ACL-reconstruction, correlated with qualitative morphologic changes in the joint and with clinical outcome. Is there evidence for early posttraumatic degeneration? Rofo. 2011;183(12):1138–44.
61. Bae JH, et al. Articular cartilage of the knee 3 years after ACL reconstruction. Acta Orthop. 2015;86(5):605–10.
62. Welsch GH, et al. Cartilage T2 assessment at 3-T MR imaging: in vivo differentiation of normal hyaline cartilage from reparative tissue after two cartilage repair procedures–initial experience. Radiology. 2008;247(1):154–61.
63. Welsch GH, et al. Quantitative T2 mapping during follow-up after matrix-associated autologous chondrocyte transplantation (MACT): full-thickness and zonal evaluation to visualize the maturation of cartilage repair tissue. J Orthop Res. 2009;27(7):957–63.
64. Raya JG, et al. Voxel-based reproducibility of T2 relaxation time in patellar cartilage at 1.5 T with a new validated 3D rigid registration algorithm. MAGMA. 2009;22(4):229–39.

Image Acquisition: Modality and Protocol Definition

4

Javier Sánchez-González and Paula Montesinos

Medical imaging has a key role in improving diagnosis and treatment of different diseases and has been used as a cost-effective solution to reduce costs in clinical trials [1]. Medical imaging overcomes the three main difficulties that traditional clinical endpoints have, such as (1) their difficulty to be standardized or quantified, (2) the long time required to be manifested, and (3) their high costs, particularly when long-term endpoints, such as mortality, are used. All these difficulties can be overcome by the so-called surrogate endpoints that are intended to substitute clinical endpoints and are expected to predict clinical benefits based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidences. The utility of surrogate endpoints is defined based on their ability to be measured earlier and more frequently than traditional clinical endpoints, facilitating the access to the final results.

On the other hand, a biomarker is defined as a medical indication of a certain medical state that can be measured in an accurate and reproducible way [2]. By definition, all measurable surrogate endpoints are biomarkers, but not all biomarkers are surrogate endpoints. For a biomarker to become a surrogate endpoint, it is necessary to establish a strong scientific connection between the illness pathogenesis and the reported biomarker [2, 3]. This scientific connection needs to ensure that the defined biomarker is closely related to the presence of the target disease or condition, to a therapeutic effect, and to the true endpoint of the evaluated therapy. Morevover, the quantitative measurement must be accurate, specific, feasible and reproducible over time [4]. Apart from this connection, the quality of a biomarker is assessed by its accuracy to express the agreement between the obtained measurements and the accepted reference and also by the reproducibility of the provided quantitative measurements.

The concepts of accuracy and reproducibility are also applied to the particular case of imaging biomarkers, imposing strong requirements to reduce subjectivity and variability [5, 6] in the process of generating imaging biomarkers, which is generally a very complex procedure. In order to obtain reliable information from images, it is necessary to standardize the way the information is generated by imposing different procedure requirements. This must cover from image acquisition to image processing, archiving, and interpretation [7].

To fulfill the above requirements, special attention needs to be paid in the specificity and accuracy of the proposed imaging biomarker. The specificity of a certain imaging biomarker can be defined as the capability to unequivocally

J. Sánchez-González (✉) • P. Montesinos
Philips Healthcare, Iberia,
C\María de Portugal 1, Madrid 28050, Spain
e-mail: javier.sanchez.gonzalez@philips.com

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_4

assess the imaging endpoint under the presence of undesirable components that may be expected to be present. These components are normally referred as image artifacts and are potential confounding factors that may affect the final quantitative values. As an example, if lesion volume is used as a biomarker for hepatic lesion, it is important to choose an imaging procedure that ensures few motion artifacts, reducing the impact of respiratory movements, and provides adequate lesion-to-background contrast to be able to depict lesion borders.

In the definition of imaging biomarkers, specificity and accuracy are closely related. Specificity could be linked to image modality selection, which is related with the biological process being assessed. Accuracy could be related to the adjustment of the imaging acquisition protocol that ensures precision and linearity of the measurements.

The Radiological Society of North America (RSNA) started the initiative called Quantitative Imaging Biomarkers Alliance (QIBA) (detailed in Chap. 1) [8]. This initiative is intended to describe step by step the process that the development of a medical biomarker has to follow in order to be used in clinical trials. This initiative covers all the aspects in the development of an imaging biomarker, paying special attention to the image acquisition process. During the generation of an imaging biomarker, it is necessary to take into account all the steps regarding image acquisition, from image modality selection to protocol imaging definition.

The QIBA initiative describes biomarker profiles and protocols. A QIBA Profile addresses different aspects of the development of an imaging biomarker; the "Clinical Context" section describes the clinical context where the biomarker can be used, while "Profile Claims" describes the information to be obtained from the biomarker and the expected within-subject variability. Finally, "Groundwork" step includes the definition of quality control metrics; patient preparation, intended to reduce variability due to external uncontrolled sources; and other details as software version and determination of performance claims. Taken into consideration all the information described in the QIBA Profile, a so-called QIBA protocol is generated.

The rest of this chapter explains the rationale behind the selection of a specific imaging modality and a particular image acquisition protocol. For a better understanding of the selection process, this chapter will be guided with illustrative examples.

4.1 Medical Image Modality Selection

The basic step in defining an imaging biomarker is to find the imaging modality that best serves to measure the underlying biological process being monitored. For example, hepatic lesion size is a well-established biomarker used to evaluate metastatic diseases; reduction in tumor size has been shown to correlate with prolonged patient survival [9]. In order to provide accurate volume measurements, imaging modalities with 3D representation of the anatomy are preferred over projective modalities like diagnostic radiography. Among 3D medical imaging technologies, those with higher spatial resolution and adequate coverage, like computed tomography (CT) or magnetic resonance (MR), are preferred to those with less spatial resolution or higher subjectivity, as ultrasound (US) or nuclear medicine technology. During the image modality selection process, all procedural aspects that may make biomarkers more reproducible and less sensible to potential artifacts must be taken into account. Continuing with the tumor volume example, CT is preferred to MR, due to shorter acquisition times, which requires shorter patient breath-holding and therefore makes CT less prone to respiratory artifacts.

As another example, perfusion is a biomarker very well established in the clinical field [10]. In the clinical area, perfusion can be assessed by US [11], CT [12], positron emission tomography (PET) [13], and MR. To define a quantitative

biomarker like perfusion from an imaging modality, there are some requirements that must be fulfilled, as coverage, spatiotemporal resolution, and a well-established relation between signal changes and contrast concentration. In the case of US, the mechanism to generate a signal change due to the presence of contrast is highly nonlinear [14], making it difficult to assess the contrast concentration from signal intensity changes in the images. Nuclear medicine techniques have been proposed to assess absolute cardiac perfusion due to their high sensitivity and the good linearity between signal and tracer concentration [15]. To avoid absorption problems, short half-life tracers, as ^{13}N-ammonia and ^{15}O-labeled water, are generally used, which requires to have a cyclotron close to the nuclear medicine departments. This lack of accessibility makes PET a difficult modality to be used in clinical trials.

On the other side, CT and MR are the most suitable modalities to perform quantitative perfusion measurements. CT modality properties such as its linear behaviour between signal changes and contrast concentration, its high aquisition speed and its high spatial resolution make this technique the preferred choice for quantitative perfusion assessment [16]. However, the major drawback of CT perfusion is the radiation exposure due to the great amount of scans acquired over the same region. These limitations can be overcome with advanced reconstruction techniques [17] and multi-detector technologies with enough spatial coverage. Although both technologic advances are becoming a standard, up to now, they are not widely available in hospitals.

On the other side, MR is an option widely available in many hospitals. This modality allows high temporal and spatial resolution, with a good spatial coverage and without radiation exposure. The main drawback of MR is the nonlinear relation between signal intensity changes and contrast concentration, even so this relation is very well established in the literature. These characteristics make MR a good modality to develop a biomarker for quantitative perfusion assessment.

4.2 Imaging Protocol Definition

The accuracy of a given measurement is evaluated from multiple samples acquired under the prescribed conditions, in a subject or group of subjects; those measurements have to be compared with an accepted reference. Accuracy is composed by both linearity and precision. The linearity is understood as the average deviation between the accepted true value (or reference) and the quantitative value provided by the proposed biomarker. On the other hand, precision is expressed as the closeness of two or more measurements obtained from multiple samples acquired under the prescribed conditions. Once the imaging modality is selected, to provide an accurate and reproducible biomarker, both linearity and precision must be taken into account and combined in the definition of the image acquisition protocol and the image processing.

It is known that not all imaging systems perform with equivalent capabilities. Hardware characteristics of the scanners influence the definition of the acquisition protocols and the quality of the biomarkers. For example, in MR the characteristics of the magnetic field gradients will define the resolution of the images obtained, the acquisition speed, and also other important acquisition parameters as the echo time (TE). Differences among different imaging systems are taken into account in the QIBA protocols that define three different levels of compliance (acceptable, target, and ideal):

- *Acceptable* defines the minimum quality required in the acquired data. If the imaging protocol does not reach these requirements, it is difficult to ensure reliable quantitative data.
- *Target* represents an acquisition protocol that is considered to be achievable with reasonable effort and that provides better quantitative results than that of the previous case.
- *Ideal* represents a protocol that can reach better quantitative results of the biomarker, but requires a strong effort from the imaging parameters point of view.

The first step to develop a biomarker is to define the information that images must contain and that is required by the post-processing step. Also, the minimum acquisition requirements (coverage, spatial and temporal resolution) to obtain the mentioned information must be defined. All the mentioned requirements need to be identified to ensure biomarker accuracy. Biomarker reproducibility can be improved by taking into account all external factors that might influence the final quantitative values. These factors need to be carefully described in the complete imaging protocol, covering from patient preparation to image acquisition parameters. This work has been done by the QIBA initiative for several biomarkers, including MR perfusion.

Regarding the perfusion example, quantitative perfusion basically depends on the amount of blood that reaches the desired parenchyma through the corresponding artery, which is called vascular input function (VIF), and on the amount of blood that is released to the organ of interest. To include VIF in the post-processing helps to eliminate external physiological effects, as, for example, different cardiac outputs between scans and patients, or a variable contrast distribution in the body; this way more reproducible measurements will be obtained.

Regarding the patient preparation process for quantitative perfusion measurements, it is critical to standardize any aspect related with contrast concentration, including dose and injection rate. Both dose and injection rate have a relevant impact on the estimation of the VIF and on the tissue contrast uptake. The first factor that needs to be defined is the contrast dose; safety specifications of the different contrast agents must be followed.

Once the safety limits are fixed, there are other factors that need to be taken into account. For example, if we assess quantitative perfusion by PET, it is important to take into account some factors as the scanner dead time factor under high activity conditions. This parameter is essential to avoid scanner saturation due to high activity, especially during the first pass of the tracer, as it requires to process too many events at the same time. In the case of MR, the use of high concentrated contrast bolus, obtained from fast injection rates, will produce a signal drop in the VIF estimation during first pass of the contrast, due to T2* relaxation effects. This nonlinearity will affect the final estimation and must be handled with an appropriate injection rate. On the contrary, to evaluate contrast organ uptake, a higher injection rate is preferred, since contrast concentration per pixel is smaller in the tissue than in the VIF. In conclusion, to obtain an accurate VIF and a proper evaluation of tissue contrast uptake, a balance between contrast volume and injection rate needs to be found. A good overview of all the potential factors that can affect absolute cardiac perfusion assessment in MR can be found in the work presented by Sánchez-González J et al. [18]. In this article the authors used 3 ml/s injection rate to avoid T2* signal decay in the VIF under a wider range TEs.

Besides, there is some physiological information that must be recorded during patient preparation; this information will be used in the final parameter estimation part. Continuing with the perfusion case, patient hematocrit is a good example of it. The contrast is distributed in the blood plasma; nonetheless for the estimation of the blood flow and volume, the whole blood volume needs to be taken into account. Hematocrit can change among subjects; thus it is desirable to correct these variations to improve measurement reproducibility.

Also, during the patient preparation part, it is extremely important to check safety issues related to any potential risk of the imaging procedure. A clear example in the case of dynamic contrast-enhanced MRI is to check for creatinine levels before contrast administration; creatinine will serve to evaluate renal function to assure a proper contrast elimination.

Once the patient preparation criteria are fixed, it is important to define the proper image acquisition parameters that will provide the most accurate measurements. In the example of quantitative perfusion assessment by MR, to gather the information required to evaluate the VIF and the contrast uptake in the tissue, it is important to ensure a temporal sampling rate high enough to capture fast signal changes. In addition, to assess reliable measurements of perfusion, it is important to accurately estimate the contrast concentration in

both the blood and the tissue. These two physiological factors, VIF and contrast uptake in the tissue, define the key information that needs to be acquired for a proper perfusion assessment.

As mentioned before, sampling rate must be high enough to be able to measure the fast signal intensity changes in the VIF and in the tissue contrast uptake. To better understand this, Fig. 4.1 shows the VIF measured in the descending aorta between 10 and 60 s after the injection, with four different sampling rate values, 4, 5, 6, and 7 s. In this case the contrast bolus was injected at an injection rate of 5 ml/s. Sampling rates between 4 and 6 s provide almost equivalent VIF results, while lower sampling rates do not provide sufficient temporal resolution to ensure an accurate VIF acquisition, this is translated in a signal intensity drop. In the QIBA protocol for MR perfusion, it is established that a dynamic interval between 3 and 5 s must be used to ensure correct sampling. Minimum spatial coverage and adequate image resolution must be used to allow for a proper organ study, maintaning at the same time the mentioned temporal resolution. Another aspect to take into account is to acquire samples with enough total duration to accurately assess contrast washout.

Another key factor is the transformation of signal intensity changes into contrast concentration changes. In imaging modalities like CT or PET, there is a linear relationship between signal changes and contrast concentration changes. However, in other modalities like MR, this relation is not linear and needs to be carefully taken into account during image acquisition and image processing. On top of that, these signal changes cannot be directly obtained from the signal, and some basal calibrations are required for a proper quantification. As a result, from the nonlinearity between the MR signal and the contrast concentration, the effect of contrast over singal changes depends on the baseline T1 values of the tissues. For those tissues with short baseline T1 values, the MR signal change due to contrast uptake will be smaller than for tissues with longer baseline T1 values. Figure 4.2 shows a simulation of the effect of contrast on the MR signal in two different tissues, equally perfused but with different baseline T1 values (260 and 1500 ms). Figure 4.2b shows the normalized signal with a relative signal difference of 80 % between both tissues. As both tissues have the same perfusion value and contrast concentration, the difference is just derived from different baseline T1 values. To compensate

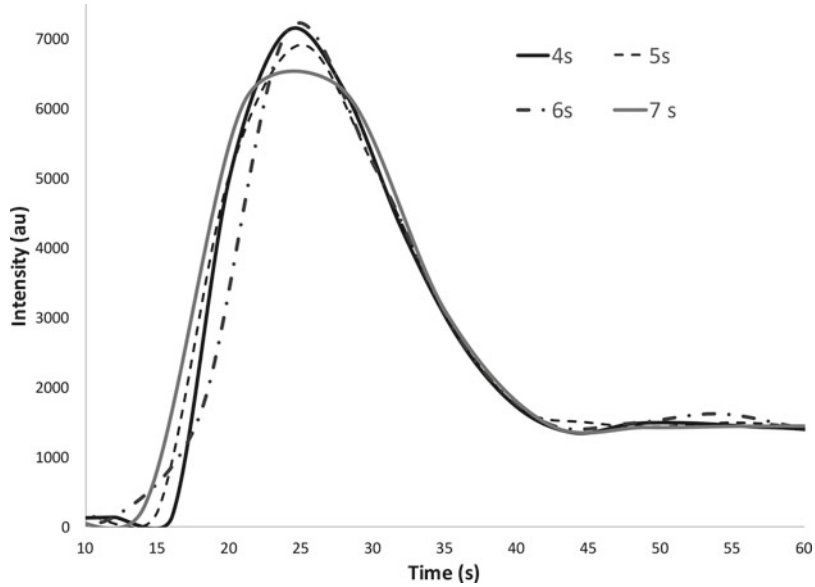


Fig. 4.1 Zoom of the signal intensity curves showing signal change during first pass of contrast bolus acquired with four different sampling rates (4, 5, 6, and 7 s)

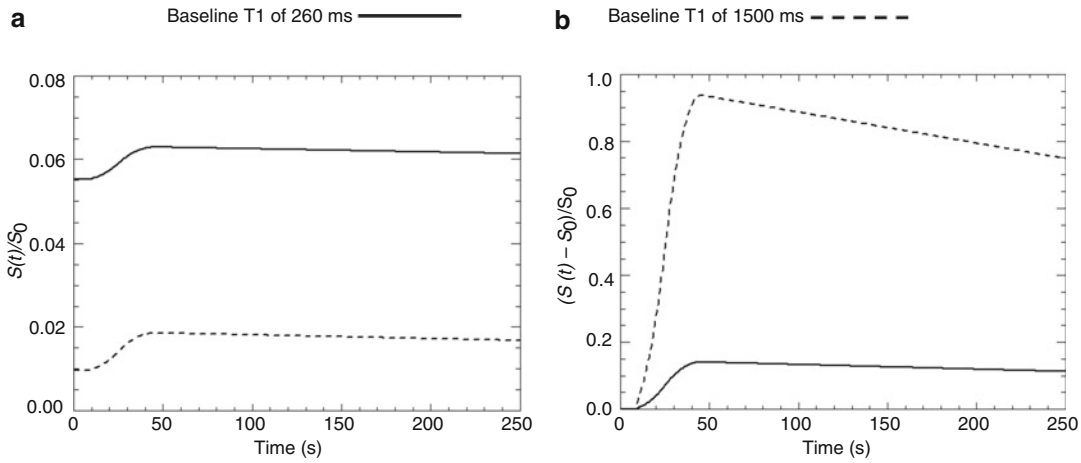


Fig. 4.2 Simulation of contrast uptake in two tissues with different baseline T1 values. Both tissues have equal perfusion and contrast arrival. (**a**) Signal intensity during the first pass of the contrast normalized by the signal before contrast arrival (S_0). (**b**) Signal intensity change compared with baseline signal (S_0), normalized by the baseline signal

the impact of baseline T1 differences on quantitative MR perfusion, some baseline measurements must be acquired during the imaging acquisition protocol; this way reproducibility between different scans and patients will be improved. These baseline T1 measurements are generally addressed using special image acquisition MR sequences [19, 20].

During the contrast administration, a dynamic protocol must be acquired. This dynamic protocol must be defined to enhance the image contrast to provide a more precise analysis. In certain medical imaging modalities, such as CT, besides improving contrast sensitivity, special attention has to be paid to reduce the radiation exposure. Iterative reconstruction techniques have a big impact on both aspects [17, 21], allowing to reduce the total radiation dose with minor impact on image quality while maintaining excellent contrast sensitivity [17].

MR image contrast depends on many different acquisition parameters that need to be properly tuned and combined. For dynamic contrast-enhanced experiments, it is required to enhance the T1 properties of the contrast, avoiding at the same time potential signal saturations. A good review of the technical details and the process to assess perfusion parameters by MR can be found in [22]. For cardiac perfusion, different approaches have been proposed to avoid signal saturation due to high contrast concentration, especially in the assessment of VIF while maintaining adequate uptake in the cardiac muscle [18, 23, 24]. In other organs, such as the prostate, the standard approach is to use 3D acquisitions covering the whole organ. On those sequences, spoiled gradient echo sequences are generally applied; signal modeling is perfectly described in these sequences, which ensures a proper conversion from signal variation to contrast concentration changes. In the case of 3D acquisitions, in order to avoid T1 saturation of the signal caused by high contrast concentrations, relative high excitation pulses are usually applied with flip angles between 20 and 30°. In addition to the T1 effect, it is important to reduce the impact of T2* in VIF estimation. To avoid this T2* effects, it is recommended to acquire the images with minimum TE. In the QIBA protocol for MR perfusion biomarker, a TE lower than 1.5 ms is proposed as ideal and a range between 2 and 2.5 ms as acceptable. In the case of the repetition time (TR), the ideal values would be lower than 3 ms, while values between 5 and 7 ms are acceptable.

Conclusions

In this chapter we have introduced the rationale to define an imaging protocol to generate imaging biomarkers from medical images;

this rationale has been illustrated using different practical examples.

Accuracy and reproducibility of any biomarker are the key properties that make it suitable to be used in clinical decisions. To fulfill these requirements, it is essential to choose the right medical image modality that must ensure high sensitivity to the physiological property that wants to be measured, minimizing at the same time potential artifacts in the final measurements. Once the image modality is selected, the full biomarker protocol needs to be defined; this protocol must cover from patient preparation to imaging protocol definition. Patient preparation needs to fulfill all safety aspects related to the imaging test and to collect all the physiological information required in the post-processing step. Finally, the image acquisition protocol needs to be adjusted to gather all the information that is required by the post-processing to provide accurate outcomes, avoiding any potential confounding factor.

References

1. O'Neill RT. FDA's critical path initiative: a perspective on contributions of biostatistics. Biom J. 2006;48(4):559–64.
2. Strimbu K, Tavel JA. NIH public access. Curr Opin HIV AIDS. 2011;5(6):463–6.
3. Aronson JK. Biomarkers and surrogate endpoints. Br J Clin Pharmacol. 2005;59(5):491–4.
4. Smith JJ, Sorensen AG, Thrall JH. Biomarkers in imaging: realizing radiology's future. Radiology. 2003;227(3):633–8.
5. Brealey S. Measuring the effects of image interpretation: an evaluative framework. Clin Radiol. 2001;56(5):341–7.
6. Sica GT. Bias in research studies. Radiology. 2006;238(3):780–9.
7. Endpoints CT. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. Biotechnol Law Rep. 2007;26(4):375–86.
8. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M, Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27–67.
9. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. Lancet (London, England). 2000;356(9227):373–8.
10. Functional imaging in oncology: clinical applications – Volume | Antonio Luna | Springer. [Online]. Available: http://www.springer.com/us/book/9783642405815. Accessed 01 Nov 2015.
11. Westwood M, Joore M, Grutters J, Redekop K, Armstrong N, Lee K, Gloy V, Raatz H, Misso K, Severens J, Kleijnen J. Contrast-enhanced ultrasound using SonoVue® (sulphur hexafluoride microbubbles) compared with contrast-enhanced computed tomography and contrast-enhanced magnetic resonance imaging for the characterisation of focal liver lesions and detection of liver met. Health Technol Assess. 2013;17(16):1–243.
12. Ogul H, Kantarci M, Genc B, Pirimoglu B, Cullu N, Kizrak Y, Yilmaz O, Karabulut N. Perfusion CT imaging of the liver: review of clinical applications. Diagnostic Interv Radiol. 2014;20(5):379–89.
13. Schaefer WM, Nowak B, Kaiser H-J, Koch K-C, Block S, vom Dahl J, Buell U. Comparison of microsphere-equivalent blood flow (15O-water PET) and relative perfusion (99mTc-tetrofosmin SPECT) in myocardium showing metabolism-perfusion mismatch. J Nucl Med. 2003;44(1):33–9.
14. Sboros V, Tang M-X. The assessment of microvascular flow and tissue perfusion using ultrasound imaging. Proc Inst Mech Eng H. 2010;224(2):273–90.
15. Wijns W, Camici PG. The value of quantitative myocardial perfusion imaging with Positron Emission Tomography in coronary artery disease. Herz. 1997;22(2):87–95.
16. Oğul H, Kantarcı M, Genç B, Pirimoğlu B, Cullu N, Kızrak Y, Yılmaz O, Karabulut N. Perfusion CT imaging of the liver: review of clinical applications. Diagn Interv Radiol. 2014;20(5):379–89.
17. Mehta D, Thompson R, Morton T, Dhanantwari A, Shefer E, Healthcare P. Iterative model reconstruction: simultaneously lowered computed tomography radiation dose and improved image quality. Med Phys Int. 2013;1:147–55.
18. Sánchez-González J, Fernandez-Jiménez R, Nothnagel ND, López-Martín G, Fuster V, Ibañez B. Optimization of dual-saturation single bolus acquisition for quantitative cardiac perfusion and myocardial blood flow maps. J Cardiovasc Magn Reson. 2015;17(1):21.
19. Deoni SCL, Rutt BK, Peters TM. Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. Magn Reson Med. 2003;49(3):515–26.
20. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified look-locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. Magn Reson Med. 2004;52(1):141–6.

21. Löve A, Olsson ML, Siemund R, Stålhammar F, Björkman-Burtscher IM, Söderberg M. Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. Br J Radiol. 2013;86(1031):20130388.
22. Sánchez-González J, Luna A, Hygino da Cruz LC. Perfusion imaging by magnetic resonance. In: Luna A, Vilanova JC, Hygino da Cruz LC, Rossi SE, editors. Functional imaging in oncology: biophysical basis and technical approaches. Springer-Verlag Berlin Heidelberg; 2014. p. 341–76.
23. Gatehouse PD, Elkington AG, Ablitt NA, Yang G-Z, Pennell DJ, Firmin DN. Accurate assessment of the arterial input function during high-dose myocardial perfusion cardiovascular magnetic resonance. J Magn Reson Imaging. 2004;20(1):39–45.
24. Christian TF, Rettmann DW, Aletras AH, Liao SL, Taylor JL, Balaban RS, Arai AE. Absolute myocardial perfusion in canines measured by using dual-bolus first-pass MR imaging. Radiology. 2004;232:677–84.

MRI Preprocessing

5

José V. Manjón

5.1 Introduction

Medical imaging plays a major role on modern health care since it allows to unveil the interior of the human body in a nonintrusive manner. There are several medical image modalities that are used to show anatomical and/or functional information. For example, X-ray shows anatomical information using the density difference among organs and/or body parts, PET images are used to obtain metabolic information through the use of positron-emitting isotopes introduced on specific biologically active molecules, and magnetic resonance (MR) imaging can be used to analyze either anatomical or functional data from the scanned object.

All of these image modalities have their specific acquisition protocols in order to produce clinically usable data. In the past, medical image analysis consisted of fundamental steps: (1) image acquisition and (2) image observation. As a result, problems on image acquisition used to lead to problems in diagnostics. With the introduction of computers in the clinical settings, this process have gained flexibility (and complexity) allowing to perform complex anatomical or functional image analysis tasks leveraging the amount of useful information that we are able to extract from the images.

This image-specific automatic analysis pipelines are based on the assumption that the acquired images fulfill with a set of requirements such as a minimum image quality and the presence of specific properties to be analyzed. Image preprocessing is a fundamental step in those pipelines that tries to improve the image quality (degraded during the acquisition process) and/or to normalize the images to set them on a specific geometric or intensity space. In this chapter, we will review some of the most common preprocessing steps in MR imaging. However, most of these steps can be shared among different image modalities (although with some specific modality dependent adaptation).

Magnetic resonance imaging has become one of the most used and versatile medical image modalities nowadays. While in the beginning of this technique the analysis was fundamentally qualitative and based merely in the observation of the images, currently several technical developments have made possible to derive quantitative measurements from these data.

However, MR images are normally affected by different types of artifacts that need to be minimized before applying any quantitative biomarker estimation pipeline. In this chapter, we will describe some of the most common MR

J.V. Manjón
IBIME Research Group, Polytechnic University of Valencia, Valencia, Spain
e-mail: jmanjon@fis.upv.es

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_5

imaging preprocessing steps normally applied to the raw MR images in order to improve their quality or to set them on a specific geometric or intensity space making easier the subsequent processing and analysis.

Since these preprocessing steps can be sequence specific, we will focus our attention on anatomical MR image sequences explicitly excluding functional and dynamic sequences such as perfusion-weighted or BOLD MR sequences. We have to note though that most of the methods we will revise here can be applied to different sequences.

Preprocessing steps described in this chapter will be:

1. Denoising
2. Inhomogeneity correction
3. Superresolution
4. Registration
5. Intensity standardization

Bellow, in the next pages, these steps will be described in detail, and the relevant references will be presented.

5.2 Denoising

MR images are inherently corrupted by random noise from the image acquisition process. Such noise introduces uncertainties in the measurement of any quantitative biomarker.

MR image noise can be effectively reduced by simply averaging multiple acquisitions directly in the scanner. However, this is not a common practice in clinical settings since this technique significantly increases the acquisition time. Instead, filtering methods are normally applied in the preprocessing stage of many analysis pipelines.

There is a large amount of denoising methods in literature, being this field one of the most prolific in medical image processing [1]. First denoising methods had the drawback that while removing noise, they also removed high-frequency signal components, thereby blurring the edges in the images. This is, as an example, the case of classical low-pass filters like the Gaussian filter. However, adaptive image denoising methods can mitigate these drawbacks. One of the first edge preserving filters was the well-known anisotropic diffusion filter (ADF) [2, 3] which was able to effectively remove the noise, being more respectful with the image edges. Also, wavelet-based filters have been applied successfully to MR denoising [4–7]. Current *state-of-the-art* denoising methods are based on patch-wise image processing approaches exploiting sparseness or self-similarity properties of the medical images or both.

Sparseness-based methods reduce the noise by assuming that the noisy data can be represented in a lower dimensionality space. This means that most of the signals can be sparsely represented using few atoms/bases enabling to discard noise-related components or simply approximate noisy patterns by their corresponding noise-free patterns. A classic example of these techniques are the fast Fourier transform (FFT)- or discrete cosine transform (DCT)-based methods where standard atoms/bases such as sin or cosine functions are used to represent the images [8, 9]. In these methods, noise reduction is achieved by simply removing noise-related coefficients in the transform domain using either soft or hard thresholding techniques. More recently, newer techniques that learn image-specific bases have been proposed [10–12]. These techniques learn a set of bases from the images to create a dictionary to sparsely represent image patches as a linear combination of dictionary entries [13]. The advantage of these dictionaries over standard ones such those used on DCT or FFT is the fact they are better adapted to the images to be processed, which enables to obtain a sparser representation and therefore a better signal/noise separation. In MR imaging, sparse theory has been used in many recent methods [14–16].

On the other hand, self-similarity methods reduce noise by taking benefit of the natural pattern redundancy of the images. A good example of a self-similarity-based denoising method is the well-known nonlocal means (NLM) filter, first introduced by Buades et al. [17]. This method effectively reduces the noise while respecting the underlying anatomy. First application of this filter

in medical image was simultaneously but independently published in 2008 [18–20]; in both cases, the proposed methods explicitly treated the Rician nature of MR image noise. The publications related to this method are extensive [21–31].

Concretely, the NLM filter restores every pixel x_i in the image by computing a weighted average of surrounding pixels using a robust similarity measure that takes into account the neighboring pixels surrounding the pixel being compared:

$$NLM\left(x_i\right) = \sum_{\forall x_j \in \Omega} w\left(x_i, x_j\right) x_j \qquad (5.1)$$

where $w(x_i, x_j)$ is a weight assigned to value x_j representing the similarity between the local patches N_i and N_j of radius r centered on voxels x_i and x_j, and Ω represents a local search of x_i:

$$w\left(x_i, x_j\right) = \frac{1}{Z_i} e^{\frac{N_i - N_j^2}{h^2}} \qquad (5.2)$$

where Z_i is a normalization constant ensuring that $\sum_{\forall x_j \in \Omega} w\left(x_i, x_j\right) = 1$ and h acts as a filtering parameter controlling the decay of the exponential function. In Fig. 5.1, an example of the typical NLM filter output is shown demonstrating the excellent noise reduction result.

Many adaptations and incremental improvements of the NLM filter have been proposed over the last decade. Manjón et al. [21] proposed a multicomponent version of the NLM filter that benefits from the intrinsic multicomponent nature of MR images in a similar manner as is done for RGB images in photography. Modern MR image sequences use parallel imaging to accelerate the image acquisition. As a result, the noise variance is spatially modulated resulting in a spatially varying noise patterns across the images. In 2010, Manjón et al. proposed a spatially adaptive NLM filter that automatically estimated the local noise level. One key advantage of this filter is that it can be applied as a black box to many different types of MR images since it can deal with both stationary and nonstationary noise. In fact, this filter has been extensively used as a part of the well-known VBM toolbox (http://www.neuro.uni-jena.de/).

More recently, new versions of NLM method combine it with prefiltering strategies; this has boosted the accuracy of the method. In Manjón et al. [23], a very efficient DCT-based prefilter was used before applying a rotational invariant version of the NLM method. Finally, in 2015, it was proposed to change the DCT-based prefilter with a PCA-based prefilter which significantly improved the denoising performance representing the current state of the art on MR imaging denoising [32]. This filter, named PRI-NLPCA, is also able to deal with spatially varying noise patterns.

Some denoising methods are devoted to other types of MR images, such as those used on diffusion-weighted imaging (DWI). For example, Tristán-Vega and Aja-Fernández [24] proposed a method that used orientation information to filter each DW image using the correlations with images of similar orientations. Sparse reconstruction-based methods have been also pro-

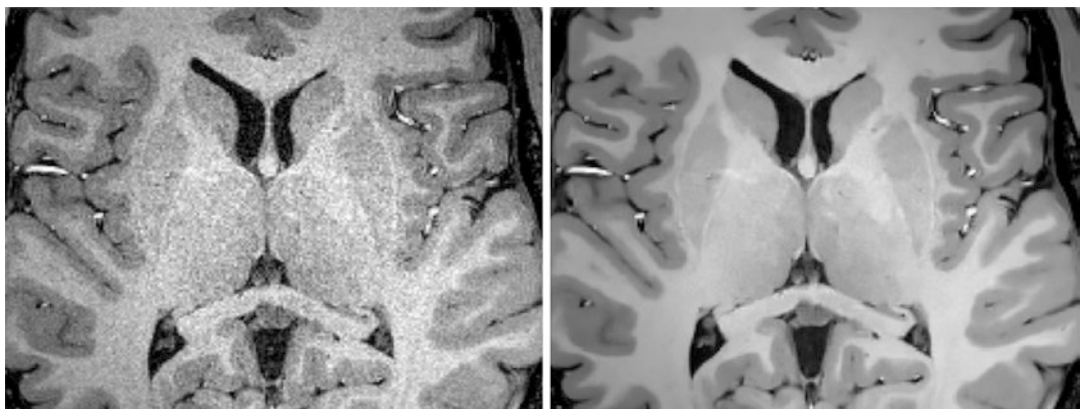


Fig. 5.1 *Left*: original noisy T1-weighted image. *Right*: denoised image using the NLM filter

posed to reduce the noise in DW images using dictionary-based approaches [16] or PCA-based decomposition [15, 33].

5.3 Inhomogeneity Correction

MR images are normally affected by signal intensity inhomogeneity which is mainly produced by imperfections in the radio-frequency coils and object-dependent interactions (Sled et al. 1998). Such artifact is perceived as a low-frequency variation of the signal intensity across the image.

Many quantitative methods, such as image registration and segmentation, rely on the assumption that a given tissue is represented by similar voxel intensities throughout the data. Therefore, correction of inhomogeneous data must be performed prior to any quantitative MR analysis. There is a large amount of bibliography dealing with the so-called bias field correction [34].

The common MR imaging signal intensity model including the inhomogeneity effect is a multiplicative model with additive noise:

$$Y = x\beta + n \tag{5.3}$$

where Y is the observed voxel intensity, β is the corresponding value of the bias field supposed to be smooth, x is the true emitted intensity, and n is a Rician distributed additive noise [35–38].

There are two main approaches for the inhomogeneity correction: prospective and retrospective strategies. The prospective methods try to avoid this type of artifact during the acquisition process by using special hardware or specific sequences such as the 3D magnetization-prepared rapid acquisition gradient echo (MP2RAGE) sequences. However, these techniques require additional hardware or extension of the acquisition time, which limits their usefulness. Retrospective methods have been more intensively used since they do not require any special acquisition protocol and can be applied *off-line* as a preprocessing step on any analysis pipeline.

There are two main classes of retrospective bias correction methods, those that model the bias field during the segmentation process and those that work directly with image features. Segmentation-based methods estimate the bias field as an image model parameter during the segmentation [38–42]. This parameter estimation is usually performed using an expectation-maximization (EM) algorithm [43].

In the SPM2 (statistical parametric mapping, Wellcome Institute, London, United Kingdom) software, Ashburner [44] modeled the bias field by using a combination of discrete cosine transform basis functions whose parameters were adjusted by the minimization of the negative log-likelihood of the log-transformed data (which is equivalent to the image entropy). This technique was further improved on subsequent versions of the software (the current one is SPM12). Although it obtains remarkable results, it is limited to brain imaging.

On the other hand, there are methods which operate directly with image properties and make minimal assumptions about the image characteristics, such as the number of tissues or location, which make them more general. Some of these methods estimate the bias field using a set of low-frequency basis functions and the minimization of some cost function related with the homogeneity of the corrected image [45–48].

However, probably the most used and referenced method is the well-known N3 method (Sled et al. 1998), which despite of its simplicity has gained a huge popularity probably thanks to its robustness. This method estimates the bias field by sharpening the image histogram using a Gaussian deconvolution and smoothing the resulting bias field estimation using B-spline fitting. More recently, an incremental improvement of N3 method called N4 has been proposed which automatically sets some of the method parameters by using a hierarchical optimization scheme [49] (Fig. 5.2).

5.4 Superresolution

In MR imaging, the images are acquired with a specific resolution which is limited by several factors including the signal-to-noise ratio (SNR), dynamic considerations, hardware, time limitations, and patient's comfort resulting in an insufficient sampling density for certain applications.

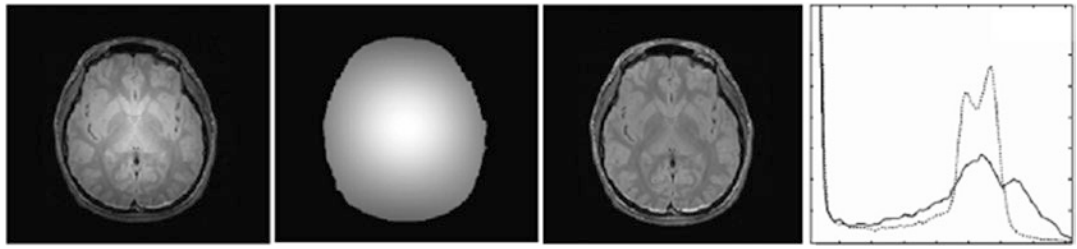


Fig. 5.2 Example of inhomogeneity correction process. From *left* to *right*: original bias field corrupted image, estimated bias field, inhomogeneity-corrected image, and histogram of the original and corrected images. Note that corrected image has a sharper histogram with well-clustered intensities

In typical clinical settings, several types of images are obtained with different voxel resolutions. Traditionally, in-plane resolution is normally higher than resolution in the slice direction, yielding non-isotropic voxel sizes. In multimodal image applications, such as image segmentation or registration, low-resolution (LR) data has to be upsampled to match a specific voxel size to make it compatible with a higher-resolution (HR) image data [50, 51]. In such cases, interpolation techniques [52, 53] have been traditionally applied. Techniques such as linear interpolation or spline-based methods have been extensively used to increase the apparent data resolution. However, such techniques estimate new points assuming that the existing ones (in the LR image) have the same value in the HR images which is only valid within homogeneous regions. As a result, interpolated images are typically blurred versions of the underlying HR images.

A better approach to effectively increase the resolution of an LR dataset is to use superresolution techniques [54]. Superresolution is a term used to refer to the process to infer a HR image from one or several LR images.

Specifically, in MR, image voxels in LR data y can be related to the corresponding underlying HR voxels x through a simple degradation model:

$$Y = DHx + n \tag{5.4}$$

where D is a decimation operator (defined as taking each Lth value starting from zero in each dimension), H is the convolution matrix, x is the underlying HR data, and n is a Rician distributed random noise [55]. In MRI, H can be roughly approximated by a 3D boxcar or Gaussian functions representing the point spread function (PSF) of the acquisition system.

Therefore, the value y_j of any voxel in the LR image can be expressed as follows:

$$y_j = \frac{1}{N}\sum_{i=1}^{N} x_i + n \tag{5.5}$$

where the value of the LR voxel y_j is the average of the corresponding N x_i voxels in the subjacent HR image (assuming equal weights for all HR voxels) plus some noise from the measurement process.

Given this MR image formation model, the aim of any superresolution method is to find the HR x_i values from the LR y_j values. This is a very ill-posed problem since there are infinite x_i values that meet such condition. A common approach to solve this problem is to minimize a merit function such as:

$$\hat{x} = \operatorname{argmin} y - DHx^2 \tag{5.6}$$

Due to the nonuniqueness of the solution for this problem, extra information is needed to constrain the possible solutions to obtain plausible results. One commonly used approach is to apply smoothness constraints in the reconstruction process that are based on the assumption of smoothness of the reconstructed data:

$$\hat{x} = \operatorname{argmin}\left(y - DHx^2 + 1R(x)\right) \tag{5.7}$$

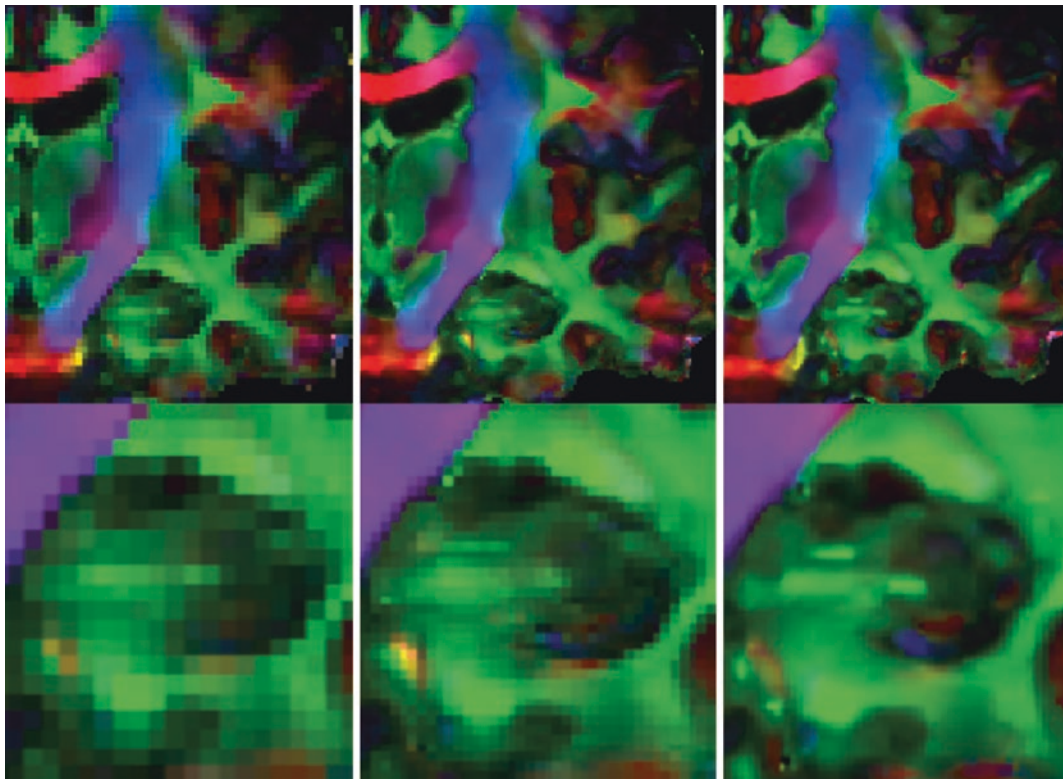
where $R(x)$ is a regularization term and λ is a weight that balances the contribution of

smoothness and data fidelity terms. However, such smoothness assumption penalizes high-frequency content of the reconstructed image that is precisely what we want to obtain. Current superresolution methods use $R(x)$ terms enforcing regularity rather than smoothness.

Superresolution techniques have been previously applied to increase image resolution in functional MR (fMR) imaging [56] and diffusion tensor imaging (DTI) studies [57]. However, most of such techniques are based on the acquisition of multiple LR images (typically orthogonal) with small shifts, a process which is time consuming and therefore not adequate for typical clinical settings. In contrast, single-image superresolution techniques do not increase acquisition time and can be applied to any dataset at the preprocessing stage of any image analysis pipeline.

An example of the single-image superresolution technique was proposed by Manjón et al. [22] where nonlocal pattern redundancy and inter-scale constraints were used to restrict the solution space for this otherwise very ill-posed problem. In this method, a nonlocal means filter is used to enforce the regularity of the image, while a mean constraint assures the inter-scale image fidelity term. Another interesting approach was also proposed by Manjón et al. [28] which benefits from the fact that in clinical settings, both LR and HR images of the patient are acquired in the same session. As a result, acquired HR images can be used to reconstruct the LR images from the same patient.

A similar approach was used in diffusion-weighted imaging to increase the resolution of the DW images [58]. In this case, the B0 image was upsampled using the nonlocal upsampling method [22], and this upsampled version (which has a higher SNR) was used to increase the resolution of the different gradient images. In Fig. 5.3, an example of the results obtained with this technique is shown.



Gold standard at 1.2x1.2x1.2 mm^3 | *Gold standard* reconstructed at 0.6x0.6x0.6mm^3 using CLASR | *Gold standard* reconstructed at 0.4x0.4x0.4mm^3 using CLASR

Fig. 5.3 *Left*: gold standard 1.2 mm^3 resolution color-coded map. *Center*: result using CLASR to reconstruct at 0.6 mm^3 (factor 2). *Right*: result using CLASR to reconstruct at 0.4 mm^3 (upsampling factor 3)

5.5 Registration

Image registration is the process of mapping different images into the same coordinate system so equivalent points of the different images share the same location in a common geometric space. Registration is necessary in order to be able to compare or integrate data from multiple sources (multimodal imaging) or to apply some analysis pipeline (e.g., segmentation).

Registration process first estimates the transformation parameters needed to map the different images. After this, the estimated transformation is applied to the moving image/s to locate them in the reference image space. Depending on the complexity of the transformation, we can have linear or nonlinear registrations. In linear registration, the same transformation is applied to every voxel in the moving image/s (specified in an affine transformation matrix encoding translations, rotations, scaling, and shears), while in nonlinear registration, different voxels may have a different transformation (specified in the so-called deformation fields).

The geometric transformation mapping from the source space to the target space is usually estimated through an optimization process. Such a process requires the use of a similarity measure to evaluate the goodness of the current transformation. Some similarity measures (Fig. 5.4) normally used in MR imaging registration are mean-squared differences (MSD), correlation coefficient (CC), or normalized mutual information (NMI). In order to reduce the computational burden of this complex estimation problem, gradient descent techniques are normally used. Recently, the use of graphical processing units (GPU) has also reduced considerably the registration time by using massively parallel approaches [59].

In MR preprocessing, registration is a typical step needed to integrate different image modalities/sequences (e.g., T1- and T2-weighted images of the same subject) or to locate the images in a

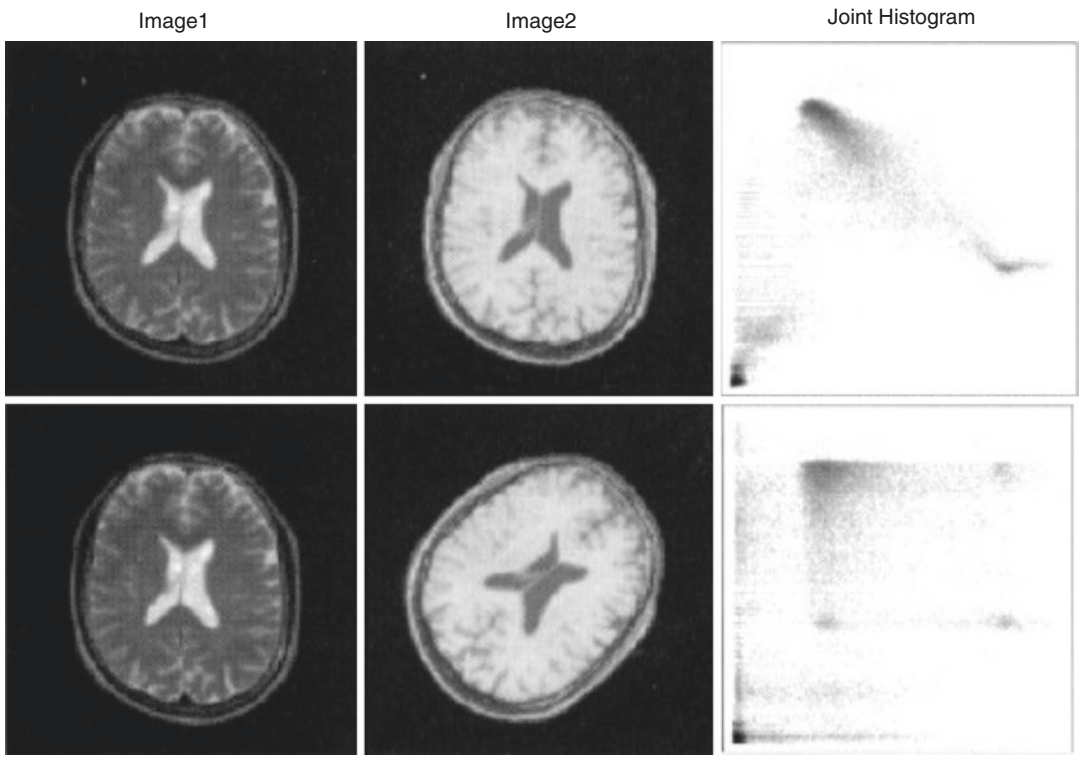


Fig. 5.4 Example of a joint histogram of a registered image pair (*upper row*) and a non-registered image pair (*bottom row*). Note the intensity dispersion patterns on both histograms (registered images show a well-grouped clusters, while non-registered images show a wider intensity distributions)

specific standard space such as the Talairach or the Montreal Neurological Institute (MNI) space where a population analysis can be performed.

There is a large number of registration methods publically available. In Klein et al.'s [60] paper, the largest evaluation of nonlinear deformation algorithms applied to brain image registration was conducted. Fourteen algorithms were evaluated over more than 45,000 registrations. The compared algorithms were AIR [61], ANIMAL [62], ART [63], Diffeomorphic Demons (Vercauteren et al. 2007), FNIRT [64], IRTK [65], JRD-fluid [66], ROMEO (Hellier et al. 2001), SICLE [67], SyN (Avants et al. 2008), and four different SPM5 algorithms ("SPM2-type" and regular normalization, unified segmentation, and the DARTEL Toolbox) [68–70]. It was concluded that ART, SyN, IRTK, and SPM's DARTEL Toolbox gave the best results (especially ART and SyN).

5.6 Intensity Standardization

Typically, MR images acquired with a similar protocol, such as T1-weighted images, do not share similar intensities across scanners. Even within the same scanner and setting, there is a variability on the intensity patterns of the acquired images on different sessions. This intensity variability does not correspond to bias field or noise, and it is significantly difficult to obtain quantitative measures directly from the data (in contrast with other medical image modalities like CT and the Hounsfield units, where same tissues have same intensities across scanners). Intensity standardization techniques in MR imaging try to correct this scanner-dependent intensity variations.

Most simple approaches to standardize MR intensities rely on the use of histogram matching techniques. Histogram matching is the transformation of an image so that its histogram matches a specified histogram [71].

In addition, scaling intensities with a simple linear transformation have been probed not sufficient since the influence of the MRI acquisition in the image intensities is nonlinear [72]. As a result, piecewise linear transformation has been widely used since it is able to model histogram intensities in a more flexible way and it also allows to incorporate anatomical information that can help on the standardization process.

For example, the technique developed by Nyul et al. [73] matches the input image histogram landmarks onto a standard histogram landmarks, obtained during an optimization process, linearly interpolating intensities between the landmarks using a piecewise linear transformation. This approach can be used to jointly standardize multimodal data (e.g., T1 and T2) using a bidimensional histogram [74].

Using a priori knowledge, the technique proposed by Hellier [72] approximates the input image histogram with a mixture of Gaussians and aligns their means with those of the standard image (reference) through a polynomial function.

Using the same philosophy, Lötjönen et al. [75] used a multiple landmark approach with a piecewise linear function. These landmarks corresponded to the means of the three main brain components (i.e., CSF, GM, and WM) which can be easily estimated if a prior segmentation is available (otherwise, Lötjönen proposed to use the classifier of Leemput et al. (1999)).

Also, Manjón et al. [76] recently proposed a similar method where the mean value of each brain tissue was estimated using the trimmed mean segmentation (TMS) method [19, 20] which robustly estimates the mean values of the different tissues by excluding partial volume voxels from the estimation jointly with the use of an unbiased robust mean estimator. Such estimation was performed using only voxels within the standard brain mask area of MNI152 template to minimize the inclusion of external tissues (Fig. 5.5).

5.7 Preprocessing Pipeline

All the described steps are aimed to either improve image data quality or to locate it at a specific geometric and/or intensity space so the following analysis steps are properly and optimally

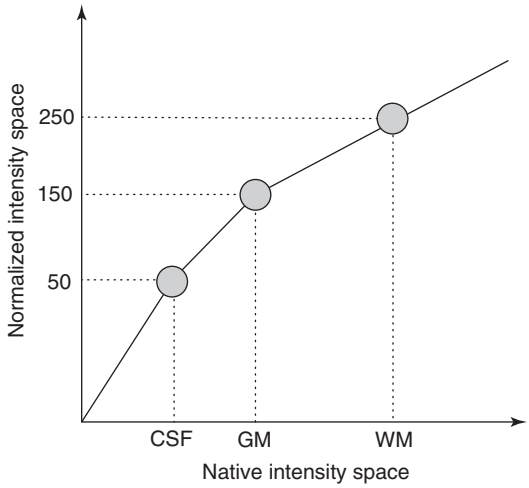


Fig. 5.5 Example of intensity normalization via a piecewise linear mapping. CSF, GM, and WM mean values are automatically estimated using TMS method and mapped to their corresponding normalized values (50, 150, and 250)

performed. The order or omission of each of these steps is highly problem dependent and the combination has to be always empirically estimated.

However, there are some considerations that can help to choose the optimal order in most of our preprocessing pipelines. For example, denoising is typically the first step because the dispersion of the signal may impact the following steps such as the homogeneity correction or registration (prefiltering the images normally helps to obtain optimal results). Besides, if inhomogeneity correction is done before filtering, this will result in a noise modulation making more difficult to restore the original signal.

After denoising, inhomogeneity correction can be helpful during the registration process. It can be also useful if a superresolution step is applied to increase the resolution of the images (especially if using self-similarity-based approaches). Nevertheless, there are methods for inhomogeneity correction combining both registration and IH correction (such as SPM) taking benefit from the expected intensity distributions on a given tissue.

Superresolution can be applied either in native (prior to the registration) or in the transformed space (e.g., MNI). This will depend on the limitations of the used data and the features of the SR method used.

Finally, intensity standardization is used to be the last step in the preprocessing as its aim is to set the images in a standardized space making possible meaningful comparisons between different acquisitions or subjects.

References

1. Mohan J, Krishnaveni V, Guo Y. A survey on the magnetic resonance image denoising methods. Biomed Signal Process Control. 2014;9:56–69.
2. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Machine Intell. 1990;12:629–39.
3. Gerig G, Kikinis R, Kubler O, Jolesz FA. Nonlinear anisotropic filtering of MRI data. IEEE Trans Med Imaging. 1992;11:221–32.
4. Donoho DL, Johnstone IM. Ideal spatial adaptation via wavelet shrinkage. Biometrika. 1994;81:425–55.
5. Kuwamura S. Wavelet denoising for tomographically reconstructed image. Opt Rev. 2006;13:129–37.
6. Nowak R. Wavelet-based Rician noise removal for magnetic resonance imaging. IEEE Trans Image Process. 1999;8:1408–19.
7. Pizurica A, Philips W, Lemahieu I, Acheroy M. A versatile wavelet domain noise filtration technique for medical imaging. IEEE Trans Med Imaging. 2003;22:323–31.
8. Guleryuz OG. Weighted overcomplete denoising. In: Proceedings of the Asilomar Conference on Signals and Systems. 2003.
9. Yaroslavsky LP, Egiazarian K, Astola J. Transform domain image restoration methods: review, comparison and interpretation. TICSP Series #9, TUT, Tampere; 2000. ISBN: 952-15-0471-4.
10. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. 2006;15(12):3736–45.
11. Mairal J, Elad M, Sapiro G. Sparse learned representations for image restoration. In: IASC2008. Yokohama; 2008.
12. Protter M, Elad M. Image sequence denoising via sparse and redundant representations. IEEE Trans Image Process. 2009;18(1):27–36.
13. Aharon M, Elad M, Bruckstein AM. K-SVD: an algorithm for designing over complete dictionaries for sparse representation. IEEE Trans Sig Process. 2006;54:4311–22.
14. Bao L, Liu W, Zhu Y, Pu Z, Magnin I. Sparse representation based MRI denoising with total variation. In: ICSP2008 Proceedings. 2008.
15. Bao L, Robini M, Liu W, Zhu Y. Structure-adaptive sparse denoising for diffusion-tensor MRI. Med Image Anal. 2013;17(4):442–57.
16. Patel V, Shi, Y, Thompson, PM, Toga, AW. K-SVD for Hardi denoising. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2011.

17. Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. In: IEEE International Conference on Computer Vision and Pattern Recognition (CPVR), 2005;2: p. 60–5.
18. Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging. 2008;27:425–41.
19. Manjón JV, Carbonell-Caballero J, Lull JJ, Garcia-Martí G, Martí-Bonmatí L, Robles M. MRI denoising using non-local means. Med Image Anal. 2008;4:514–23.
20. Manjón JV, Tohka J, García-Martí G, et al. Robust MRI brain tissue parameter estimation by multistage outlier rejection. Magn Reson Med. 2008;59(4):866–73.
21. Manjón JV, Thacker NA, Lull JJ, Garcia-Martí G, Martí-Bonmatí L, Robles M. Multicomponent MR image denoising. Int J Biomed Imaging. Article ID 756897. 2009.
22. Manjón JV, Coupé P, Martí-Bonmatí L, Robles M, Collins DL. Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging. 2010;31:192–203.
23. Manjón JV, Coupé P, Buades A, Collins DL, Robles M. New methods for MRI denoising based on sparseness and self-similarity. Med Image Anal. 2012;16(1):18–27.
24. Tristán-Vega A, Aja-Fernández S. DWI filtering using joint information for DTI and HARDI. Med Image Anal. 2010;14(2):205–18.
25. He L, Greenshields IR. A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images. IEEE Trans Med Imaging. 2009;28:165–72.
26. Katkovnik V, Foi A, Egiazarian K, Astola J. From local kernel to nonlocal multiple-model image denoising. Int J Comput Vis. 2010;86(1):1–32.
27. Rajan J, Den Dekker A, Sijbers J. A new non local maximum likelihood estimation method for Rician noise reduction in magnetic resonance images using the Kolmogorov-Smirnov test. Sig Process. 2014;103:16–23.
28. Manjón JV, Coupé P, Buades A, Fonov V, Louis Collins D, Robles M. Non-local MRI upsampling. Med Image Anal. 2010a;14(6):784–92.
29. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. MRI superresolution using self similarity and image priors. Int J Biomed Imaging. Article ID 425891. 2010b.
30. J. Rajan, J. Veraat, J.V. Audekerke, M. Verhoye, J. Sijbers, Nonlocal maximumlikelihood estimation method for denoising multiple-coil magnetic resonanceimages, Magn. Reson. Imaging 30 (2012) 1512–1518.
31. Nicolas Wiest-Daesslé, Sylvain Prima, Pierrick Coupé,Sean Patrick Morrissey, and Christian Barillot. Rician noise removal by non-Local Means filtering for low signal-to-noise ratio MRI: applications to DT-MRI. MICCAI 2008; 11(Pt 2): 171–179.
32. Manjón JV, Coupe P, Buades A. MRI noise estimation and denoising using non-local PCA. Med Image Anal. 2015;22:35–47.
33. Manjón JV, Coupé P, Concha L, Buades A, Collins DL, Robles M. Diffusion weighted image denoising using overcomplete local PCA. PLoS One. 2013;8(9), e73021. doi:10.1371/journal.pone.0073021.
34. Belaroussi B, Milles J, Carme S, Min Zhu Y, Benoit H. Intensity non-uniformity correction in MRI: existing methods and their validation. Med Image Anal. 2005;10:121–32.
35. Axel L, Constantini J, Listerud J. Intensity correction in surfacecoil MR imaging. Am J Roentgenol. 1987;148:418–20.
36. Dawant B, Zijdenbos A, Margolin R. Correction of intensity variations in MR images for computer-aided tissue classification. IEEE Trans Med Imaging. 1993;12(4):770–81.
37. Meyer C, Bland P, Pipe J. Retrospective correction of intensity inhomogeneities in MRI. IEEE Trans Med Imaging. 1995;14(1):36–41.
38. Wells III W, Grimson W, Kikinis R, Jolesz F. Adaptative segmentation of MRI data. IEEE Trans Med Imaging. 1996;15(4):429–42.
39. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging. 1999;18(10):885–96.
40. Gispert J, Reig S, Pascau J, Vaquero J, Garca-Barreno P, Desco M. Method for bias field correction of brain T1-weighted magnetic resonance images minimizing segmentation error. Hum Brain Mapp. 2004;22(2):133–44.
41. Guillemaud R, Brady M. Estimating the bias field of MR images. IEEE Trans Med Imaging. 1997;16(3):238–51.
42. van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imag. 1999;18:897–908.
43. Dempster A, Laird N, Rubing D. Maximum likelihood from incomplete data via EM algorithm. J Royal Soc. 1977;39:1–38.
44. Ashburner J. Another MRI bias correction approach. In: Eighth International Conference on Functional Mapping of the Human Brain. Sendai; 2002.
45. Mangin JF. Entropy minimization for automatic correction of intensity nonuniformity. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. 2000. p. 162–9.
46. Likar B, Viergever M, Pernus F. Retrospective correction of MR intensity inhomogeneity by information minimization. IEEE Trans Med Imaging. 2001;20(12):1398–410.
47. Vovk U, Pernus F, Likar B. MRI intensity inhomogeneity correction by combining intensity and spatial information. Phys Med Biol. 2004;49(17):4119–33.
48. Manjón JV, Lull JJ, Carbonell-Caballero J, García-Martí G, Martí-Bonmatí L, Robles M. A nonpara-

metric MRI inhomogeneity correction method. Med Image Anal. 2007;11(4):336–45.
49. Tustison NJ, Avants BB, Cook P, Zheng Y, Egan A, Yushkevich P, Gee JC. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310–20.
50. Atkins MS, Siu K, Law B, Orchard J, Rosenbaum W. Difficulties of T1 brain MRI segmentation techniques. In: The International Society for Optical Engineering, vol. 4684 of Proceedings of SPIE. 2002. p. 1837–44.
51. Vansteenkiste E, Vandemeulebroucke J, Philips W. 2D/3D registration of neonatal brain images. In: Proceedings of the The Workshop on Biomedical Image Registration (WBIR '06), 2006; p. 272–79.
52. Thevenaz P, Blu T, Unser M. Interpolation revisited. IEEE Trans Med Imaging. 2000;19(7):739–58.
53. Lehmann TM, G¨onner C, Spitzer K. Survey: interpolation methods in medical image processing. IEEE Trans Med Imaging. 1999;18(11):1049–75.
54. Carmi E, Liu S, Alon N, Fiat A, Fiat D. Resolution enhancement in MRI. Magn Reson Imaging. 2006;24(2):133–54.
55. Sijbers J. Signal and noise estimation from magnetic resonance images, Doctoral thesis, Antwepen, 1998.
56. Kornprobst P, Peelers R, Nikolova M, Deriche R, Ng M, Van Hecke P. A superresolution framework for fMRI sequences and its impact on resulting activation maps. In: Proceedings 6th International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI '03). 2003;2879; p. 117–25.
57. Peled S, Yeshurun Y. Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. Magn Reson Med. 2001;45(1):29–35.
58. Coupé P, Manjon JV, Chamberland M, Descoteaux M. Collaborative patch-based super-resolution for diffusion-weighted images. Neuroimage. 2013;83: 245–61.
59. Lin S, Liu W, Zhang H, Xie Y, Wang D. A survey of GPU-based medical image computing techniques. Quant Imaging Med Surg. 2012;2(3):188–206.
60. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009;46:786–802.
61. Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reslicing PET images. J Comput Assist Tomo. 1992;16:620–33.
62. Collins D, Holmes C, Peters T, Evans A. Automatic 3D model-based neuroanatomical segmentation. Hum Brain Mapp. 1995;3(3):190–208.
63. Ardekani BA, Bachman AH. Model-based automatic detection of the anterior and posterior commissures on MRI scans. Neuroimage. 2009;46(3):677–82.
64. Andersson J, Smith S, Jenkinson M. FNIRT—FMRIB' non-linear image registration tool. Hum Brain Map. 2008; Poster #496.
65. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ. Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imaging. 1999;18(8):712–21.
66. Chiang M-C, Dutton RA, Hayashi KM, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM. 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. Neuroimage. 2007;34:44–60.
67. Christensen G, Johnson H. Consistent image registration. IEEE Trans Med Imag. 2001;20:568–82.
68. Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. Hum Brain Map. 1999;7:254–66.
69. Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007;38:95–113.
70. Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005;26:839–51.
71. Gonzalez RC, Woods RE. Digital Image Processing. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2008. p. 128. ISBN 9780131687288.
72. Hellier P. Consistent intensity correction of MR images. In: Proceedings of the International Conference on Image Processing (ICIP '03), 2003; p. 1109–12.
73. Nyul LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging. 2000;19(2):143–50.
74. Jager F, Nyul L, Frericks B, Wacker F, Hornegger J. Whole body MRI intensity standardization. In: Horsch A, Deserno TM, Handels H, Meinzer H-P, Tolxdorff T, editors. Bildverarbeitung fur dieMedizin. Berlin: Springer; 2007. p. 459–63.
75. Lötjönen JM, Wolz R, Koikkalainen JR, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage. 2010;49(3):2352–65.
76. Manjón JV, Eskildsen SF, Coupé P, Romero JE, Louis Collins D, Robles M. Non-local intracranial cavity extraction. IJBI. Article ID 820205. 2014.

Imaging Biomarker Structural Analysis

6

Angel Alberich-Bayarri

6.1 Introduction

Medical imaging has a key role in current workflows for the assessment of clinical decisions in many disease scenarios. Concretely, imaging biomarkers are transforming the way radiology has taken part in the healthcare cycle, from conventional workflows based on qualitative criteria and the experience of the radiologist toward having a powerful measurement tool in each hospital, allowing for the extraction of quantitative indicators of tissue and organ characteristics by the application of image processing methods and algorithms to medical images from modalities like X-ray (XR), magnetic resonance (MR) imaging, computed tomography (CT), ultrasound (US), positron emission tomography (PET), single-photon emission computed tomography (SPECT), among others.

Imaging biomarkers analysis methods can be structured both in those related to structural properties and those focused on analyzing dynamic features. In the present chapter, we will focus on the explanation and description of structural imaging biomarkers, providing a classification according to the nature of the computational algorithm in which the biomarker is based. The dynamic models for the analysis of imaging biomarkers characterizing physiological phenomena (such as pharmacokinetics modeling or cellularity) will be reviewed in the following chapter.

The most basic imaging biomarkers providing structural information are those related to distances and areas of specific regions. As an example, the maximum diameter is considered as an easy and reproducible imaging biomarker for monitoring treatment response in solid tumors, through the well-known Response Evaluation Criteria in Solid Tumors (RECIST) [1]. However, recently more tissue comprehensive biomarkers like the volume of regions are preferred due to the better sensitivity to lesion size changes in the follow-up process of patients [2].

Several morphological quantitative descriptors can also be obtained from specific regions in medical images. As in the case of volumetry techniques, the most important application is in the field of oncology. Morphology analysis algorithms allow to provide biomarkers such as the spicularity, concavity, compactness, or acutance.

There are also descriptors about structural complexity that can be quantified that allow for obtaining measurements of the degree of irregularity. Irregularity of tissues has been proven to be related to several pathological processes (i.e., microarchitecture alteration of bone in osteoporosis; vessel tortuosity in tumors). Two of the

A. Alberich-Bayarri
Biomedical Imaging Research Group, La Fe Health Research Institute, Valencia, Spain

QUIBIM SL, Valencia, Spain
e-mail: alberich_ang@gva.es

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_6

most relevant indicators of irregularity are those related to complexity of tissues when "filling" a region that can be quantified by the fractal dimension parameter using specific image processing algorithms and those related to tortuosity of specific tissues (i.e., vessels) that can be calculated by the ratio between the geodesic distance and the Euclidean distance [3].

Finally, texture descriptors have become one of the most valuable tools for the characterization of heterogeneity and related properties from the quantitative analysis of the voxel intensities within a tissue or region. Typical parameters are the skewness, the kurtosis, and the entropy and can be used for the characterization of the different habitats within lesion and tissue structural alterations.

In the present chapter, a detailed description of the main structural imaging biomarkers and the most common algorithms to extract them is provided, focusing in morphology, volumetry, irregularity, and texture parameters. Combining large numbers of structural features as discussed in this chapter is also considered as the basis of radiomics, which are further explained in Chap. 8. Basically, radiomics are focused in linking the features with the clinical endpoints of the disease in order to define new prognostic imaging biomarkers of the disease.

6.2 Morphology and Volumetry Biomarkers

One of the very basic characterizations of specific regions, tissues, and organs is related to the extraction of geometrical properties such as the maximum diameter, the area, or the volume. In fact, in oncology the diameter has remained to be the main criteria for evaluating the response in solid tumors, as it is extracted from the revised guidelines of RECIST 1.1 [1]: "Must be accurately measured in at least one dimension (longest diameter in the plane of measurement is to be recorded)." One-dimensional measurements like the diameter are fast, reliable, and reproducible, and therefore, its application to clinical practice is straightforward; however, the use of areas or volumes is more representative of the real geometrical characteristics of the region of interest. Specifically, in cancer, volumetry has been shown traditionally to be superior to distances for the evaluation of the treatment response [4–6].

One of the most important parameters evaluated in this section is the tumoral volume, which consists in accounting the number of voxels of the lesion and extract the volume by multiplying it by the voxel volume. For that, the DICOM file parameters of pixel spacing (PixelSpacing, tag 0028–0030), slice thickness (SliceThickness, tag 0018–0050), and spacing between slices (SpacingBetweenSlices, tag 0018–0088) must be taken into consideration.

The spatial resolution of the acquired images has a high relevancy if volumetry has to be extracted for the follow-up of the disease. For this reason, the Quantitative Imaging Biomarkers Alliance (QIBA) already introduced in Chap. 1 has created a working group in the field of volumetry. As an example, QIBA document entitled "Lung Nodule Volume Assessment and Monitoring in Low Dose CT Screening Quantification Profile (v1.0)" which is still under discussion recommends a slice thickness small relative to the size of the smallest nodules detected and followed by CT screening therefore specifying a thickness of 1.25 mm or less for lung nodules.

6.3 Irregularity Biomarkers

Using the fractal theory, it is possible to quantify structures with complex characteristics and irregularity through the fractal dimension (D), a parameter which indicates how an irregular structure tends to fill space after the observation at different scales [7]. In order to calculate fractal dimension values of a given structure, the most extended algorithm is the so-called box counting, which can be applied for either 2D or 3D structures. The algorithm divides the structure in different regular regions (boxes) progressively varying their size and counts the number of boxes containing structural elements for each box size. Finally, a relationship between the number of

boxes containing contour and the different box sizes can be built, as seen in Eq. 6.1.

$$\log(N) = -D \bullet \log(\lambda) + k \tag{6.1}$$

Equation 6.1 gives the relationship between the number of contour boxes (N), the corresponding box size (λ), the box-counting fractal dimension parameter (D), and a proportionality constant (k).

An example of the application to trabecular bone can be observed in Fig. 6.1.

The tortuosity, τ, characterizes the sinuosity of a structure. Geometrically, it is defined as the ratio between the geodesic distance and the Euclidean distance (L_G/L_E). The concept can be observed in Fig. 6.2. This approach allows to quantitatively classify the tortuosity of any non-regular structure, such as trabecular bone or chaotic vessels feeding a malignant tumor. Therefore, a structure is considered as tortuous if $\tau > 1$ [3].

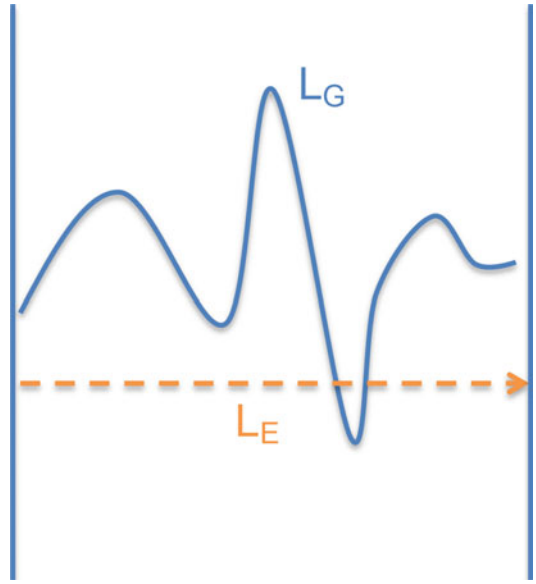


Fig. 6.2 Difference between Euclidean and geodesic distances in a filamentous structure

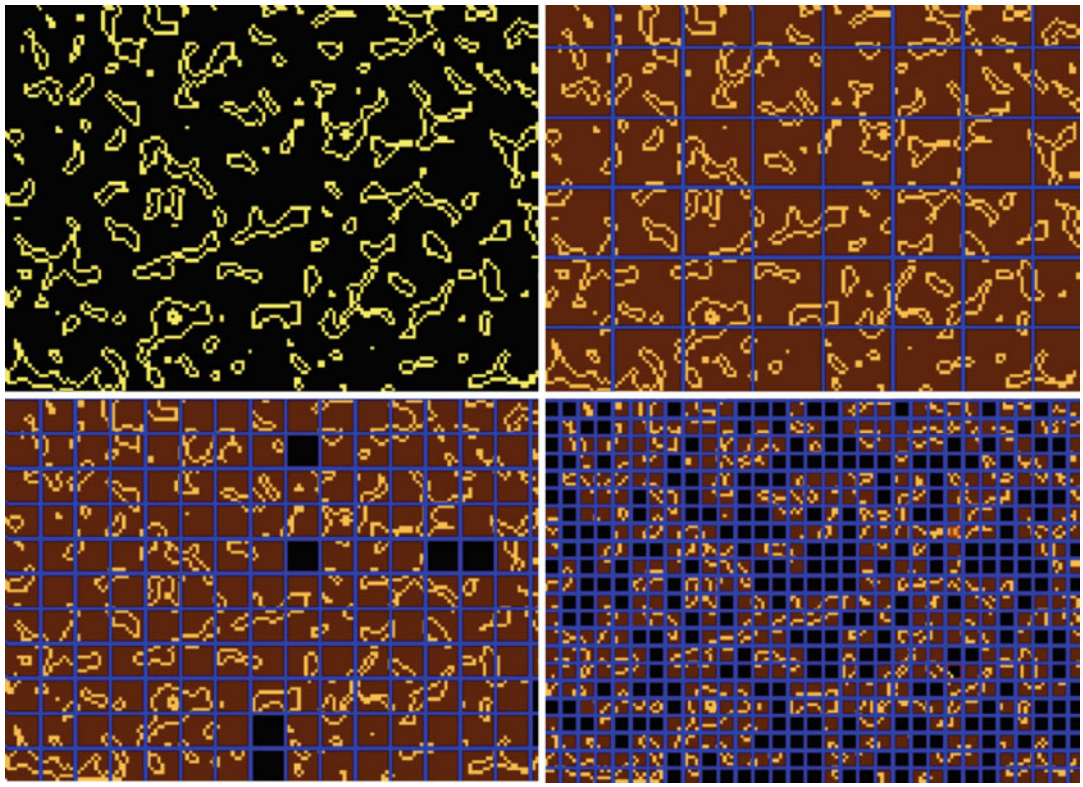


Fig. 6.1 Box-counting algorithm applied to boundaries of trabecular bone

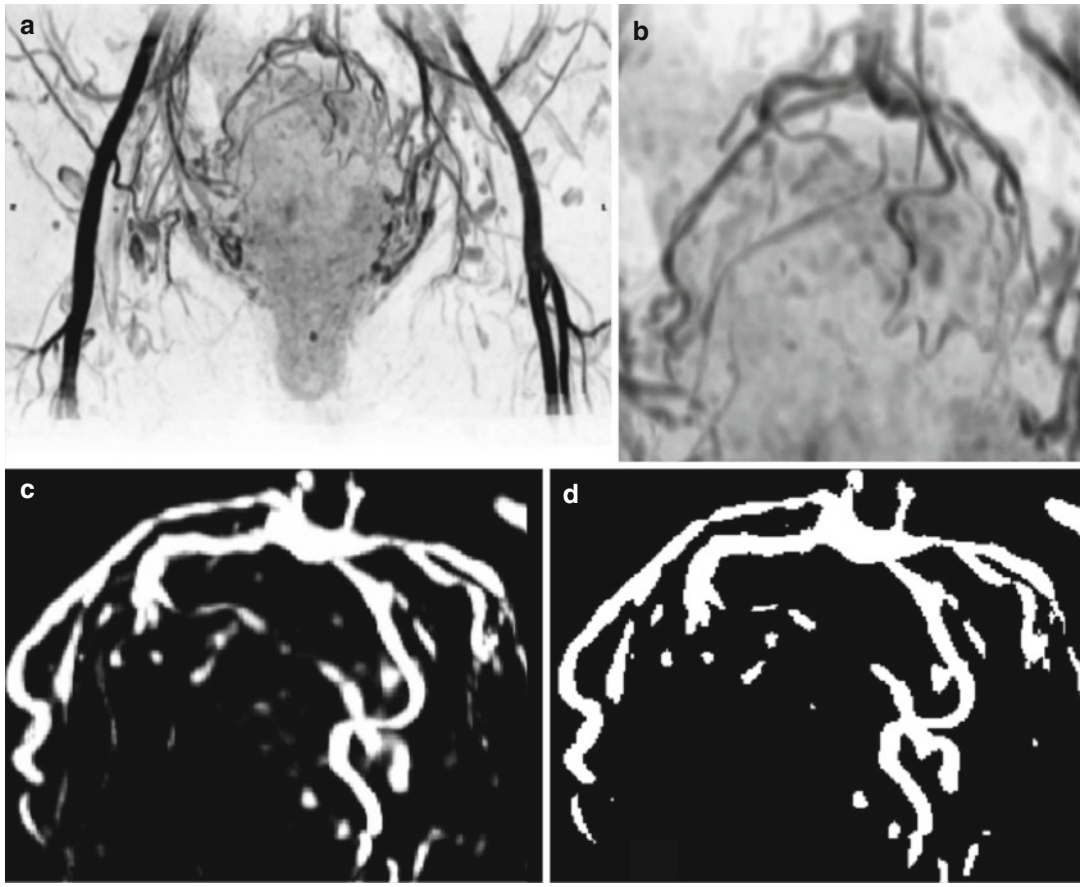


Fig. 6.3 Vascular characterization of rectal cancer. In (**a**), inverted grayscale image of a maximum intensity projection (MIP) reconstructed from a dynamic contrast-enhanced MR (DCE-MR) sequence. In (**b**), segmentation of the main vascular input to the tumor. In (**c**), grayscale. In (**d**), binarized image ready for the application of 2D tortuosity and fractal analysis techniques. For this case, $D = 1.43$ and $\tau = 1.58$

These methods can be applied to quantitatively evaluate tumor irregularity, either the tumor itself or any of its characteristics, like the different enhancement regions [8], or the tortuous and irregular feeding vessels (Fig. 6.3).

6.4 Texture Biomarkers

Biological tissue composition and organization present a high variability depending on the organ and function in which their cells are specialized. Many tissues have an isotropic distribution, not being formed by elements aligned in a specific space direction, such as fat that can be considered as an isotropic tissue. Some others, however, have a clear arrangement in fiber bundles, such as muscle or white matter. Pathologic tissues like tumors are in most cases highly heterogeneous, combining regions with necrosis, edema, infiltration, hypoxia, and many other biological phenomena. In fact, tumor heterogeneity is considered to provide valuable prognosis information [9]. Different tissue regions can be analyzed by image processing techniques looking for mathematical descriptors of the grayscale pixel intensities, also called texture analysis. Also in the touch sense, the texture-based techniques allow for the characterization of the tissues from medical images in terms of uniformity, smoothness, granularity, and roughness.

From the technical perspective, a co-occurrence matrix is calculated to extract pattern information of the grayscale intensities in the image and the changes of a pixel with respect to the neighboring pixels, considering the position, the intensity value difference, and distance between them. Textural properties can also be understood as a function from different signal intensities. Normalization must be applied to the co-occurrence matrix in order to calculate specific parameters which represent some texture characteristics like the correlation, contrast, entropy, energy, and homogeneity [10]).

As can be concluded, texture analysis methods can be applied either to original medical imaging data directly provided by the modalities or also to parametric or multiparametric images.

Conclusions

The authors have presented in this chapter the main families of structural imaging biomarkers that can be applied to extract quantitative information from non-dynamic data. The progress in these topics will help to introduce Imaging Biomarkers in diagnosis and treatments follow-up criteria and new indicators, especially in oncology, where measurements like volumes and textures will be progressively integrated in response criteria guidelines.

References

1. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228–47.
2. Lin M, Pellerin O, Bhagat N, Rao PP, Loffroy R, Ardon R, Mory B, Reyes DK, Geschwind JF. Quantitative and volumetric European Association for the Study of the Liver and Response Evaluation Criteria in Solid Tumors measurements: feasibility of a semiautomated software method to assess tumor response after transcatheter arterial chemoembolization. J Vasc Interv Radiol. 2012;23(12):1629–37.
3. Roque WL, Arcaro K, Alberich-Bayarri A. Mechanical competence of bone: a new parameter to grade trabecular bone fragility from tortuosity and elasticity. IEEE Trans Biomed Eng. 2013;60(5):1363–70.
4. Nougaret S, et al. MR Volumetric Measurement of Low Rectal Cancer Helps Predict Tumor Response and Outcome after Combined Chemotherapy and Radiation Therapy. Radiology. 2012;263(2):409–18.
5. Mayr NA, Magnotta VA, Ehrhardt JC, Wheeler JA, Sorosky JI, Wen BC, Davis CS, Pelsang RE, Anderson B, Doornbos JF, Hussey DH, Yuh WT. Usefulness of tumor volumetry by magnetic resonance imaging in assessing response to radiation therapy in carcinoma of the uterine cervix. Int J Radiat Oncol Biol Phys. 1996;35:915–24.
6. Andreopoulou E, Andreopoulos D, Adamidis K, Fountzila-Kalogera A, Fountzilas G, Dimopoulos MA, Aravantinos G, Zamboglou N, Baltas D, Pavlidis N. Tumor volumetry as predictive and prognostic factor in the management of ovarian cancer. Anticancer Res. 2002;22:1903–8.
7. Mandelbrot BB. The Fractal Geometry of Nature. New York: W.H. Freeman and Company; 1982.
8. Sanghera B, Banerjee D, Khan A, Simcock I, Stirling JJ, Glynne-Jones R, Goh V. Reproducibility of 2D and 3D fractal analysis techniques for the assessment of spatial heterogeneity of regional blood flow in rectal cancer. Radiology. 2012;263(3):865–73.
9. Lee J, Narang S, Martinez JJ, Rao G, Rao A. Associating spatial diversity features of radiologically defined tumor habitats with epidermal growth factor receptor driver status and 12-month survival in glioblastoma: methods and preliminary investigation. J Med Imaging (Bellingham). 2015;2(4):041006.
10. Mahmoud-Ghoneim D, Toussaint G, Constans JM, de Certaines JD. Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. Magn Reson Imaging. 2003 Nov;21(9):983–7. PubMed PMID: 14684200.

Imaging Biomarker Model-Based Analysis

7

George C. Manikis, Eleftherios Kontopodis, Katerina Nikiforaki, Konstantinos Marias, and Nickolas Papanikolaou

7.1 Introduction

Magnetic resonance imaging (MRI) is an imaging technique that is based on the interactions of water with external magnetic fields. Magnetic properties of water molecules are analyzed in order to sketch the profile of tissues, and they may be related to a variety of aspects including internal structure, tissue integrity, molecular environment, and others. In order to elucidate tissue properties, it is often necessary to acquire multiple series of images and quantify the progress of a certain parameter in time or the degree of response to an external perturbation. After careful sequence optimization in order to selectively trigger the process, one needs to appreciate all the possible factors affecting the evolution in order to constitute a robust model. Complex phenomena taking place after excitation are decomposed in one or more mathematical terms of an appropriate form and weighting, to comply with the physical rules behind the sequence of events taking place. When model predicted data converge to the experimental measurements, the model can be considered as reliable, in the frame of a carefully designed imaging protocol. In this chapter, we will focus on the most important models used to extract imaging biomarkers related to diffusion and perfusion studies.

7.2 Diffusion-Weighted Imaging (DWI) MRI

Diffusion is the process of random motion of water molecules in a free medium. For human tissues, water mobility can be assessed in the intracellular, extracellular, and intravascular spaces. All media have a different degree of structure and thus pose a variant level of difficulty in water mobility that is called "diffusivity." A sequence sensitized to microscopic water mobility by means of strong gradient pulses can be utilized to provide insights in the complexity of the environment which in turn can reveal information related to tissue microarchitecture.

A major requirement in diffusion imaging is to select ultrafast pulse sequences that may freeze macroscopic motion in the form of respiration, peristalsis, or patient motion in general. For this reason, echo planar imaging (EPI) sequences modified with the addition of two identical strong diffusion gradients are routinely used to provide diffusion images. The amplitude and duration of the diffusion gradients is represented by the

G.C. Manikis, MSc • E. Kontopodis, MSc
K. Nikiforaki, MSc • K. Marias, PhD
Computational BioMedicine Laboratory (CBML), Institute of Computer Science (ICS) Foundation for Research and Technology – Hellas, Nikolaou Plastira 100, Heraklion, Greece

N. Papanikolaou, PhD (✉)
Karolinska University Hospital, Stockholm, Sweden
e-mail: npapan@npapan.com

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_7

"*b*-value" (measured in s/mm^2), an index used to control the sensitivity of DWI contrast to water mobility.

7.2.1 DWI Modeling

7.2.1.1 Gaussian Mono-exponential

Apparent diffusion coefficient (ADC) [1] was the first and most widely used quantitative biomarker associated to cellular density and the extracellular space fraction [2] with "apparent" giving away a reluctance to use it literally as the distance traveled from the water molecule in a certain time in certain surroundings. The simplest model assumes exponential signal decay where exponential coefficient correlates with the product of *b**ADC for each tissue type. For a given *b*-value, the signal intensity of the diffusion-weighted image (DWI) depends on the ADC of tissue in each individual pixel. ADC can be estimated with acknowledgment of two or more measurements on different *b*-values (one with a *b*-value of zero and at least one with a higher *b*-value) as described in the following mono-exponential equation:

$$S_b = S_0 * \exp\left(-b * \mathrm{ADC}\right)$$

S_b is the measured signal intensity of the DWI with gradient factor attenuation *b* (s/mm^2), and S_0 is the measured signal intensity in the absence of diffusion weighting (Fig. 7.1).

7.2.1.2 Gaussian Bi-exponential

After the introduction of mono-exponential decay, a more complex model was proposed in order to add sensitivity to the arbitrary motion related to micro-capillary perfusion which induced deviation from the initially assumed decay. The previously described mono-exponential decay in many cases failed to estimate fast-decaying signal appearing in the low-*b*-value area, and therefore a more complex model, the intravoxel incoherent motion (IVIM) [4] model, was proposed to take into account flow phenomena that contributed to DWI contrast (Fig. 7.2). The IVIM model assumes that tissue is primarily characterized by two distinct compartments (an intravascular and an interstitial space) with negligible water exchange between them, where the DW signal of each pixel can be expressed from the following bi-exponential equation:

$$S_b = S_0 * \left((1-f) * \exp\left(-b * D\right) + f * \exp\left(-b * D^*\right)\right)$$

Similarly to the mono-exponential fit equation, S_b is the measured signal intensity of the diffusion-weighted image with a gradient factor attenuation *b* (s/mm^2), and S_0 is the measured signal intensity in the absence of diffusion weighting. This new model attempted to measure the diffusion signal contamination with the added term of "microperfusion" (*D**), representing signal loss resulting from other processes, most likely microperfusion of blood nutrients at capillary level. *D** is associated with blood velocity and capillary vessels geometry.

Fitting the bi-exponential model, values for the true diffusion coefficient (*D*), microperfusion coefficient (*D**), and perfusion fraction (*f*) are calculated. If the model is applied in a pixel-based level, then the IVIM coefficients are derived for every pixel and displayed as parametric maps. IVIM is an extended model of the conventional mono-exponential diffusion model which equates to a mono-exponential form in the absence of the perfusion fraction (*f*). The effect of the microcirculation of blood, and thus the amount of vascularity, in the bi-exponential signal decay of the IVIM model varies according to the studied organ. Studies report low perfusion fractions in the brain [5], whereas in the body, the perfusion activity can be much more significant [6]. An indicative comparison between mono- and bi-exponential models when applied to normal liver tissue is shown in Fig. 7.2.

7.2.1.3 Non-Gaussian Mono-exponential

Both the mono-exponential and the bi-exponential models rely on the assumption that water mobility follows a random, unrestricted pattern which can be considered as a Gaussian displacement distribution. However, in biological tissues the

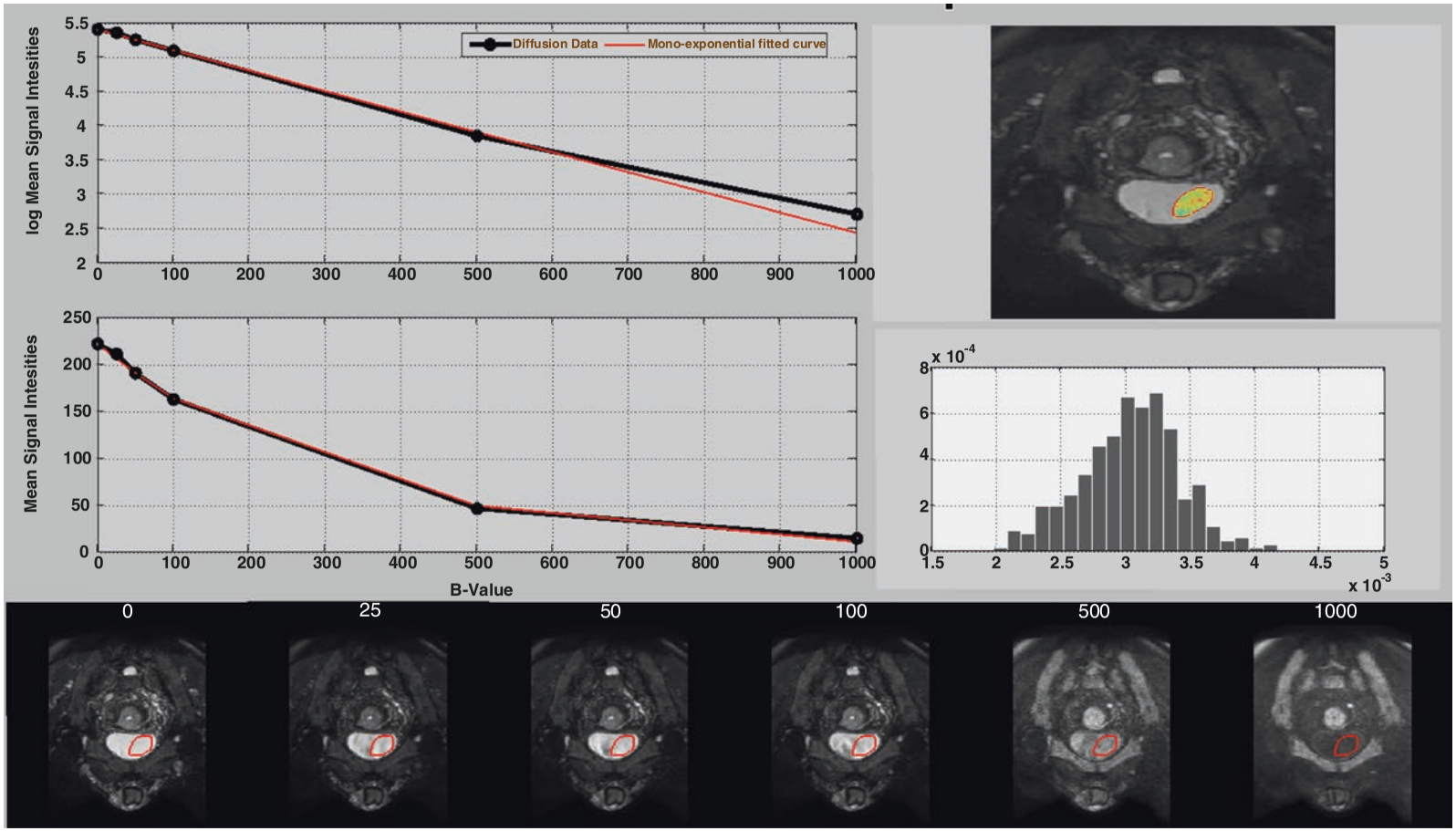


Fig. 7.1 Gaussian mono-exponential analysis in the bladder with a mean value of ADC = 3.04*10^{-3} mm^2/s. *Top left*: Mean value of the DWI signal intensities in an examined region of interest (ROI) and the mean fitted signal attenuation decaying curve from the mono-exponential model (*black* and *red solid line*, *respectively*). In the top sub-figure, the vertical axis is shown in semilogarithmic form. *Top right*: The examined ROI (*red contour*) in a specific slice of the DWI. *Bottom*: The examined slice with signal intensities decaying as *b*-value increases [3]

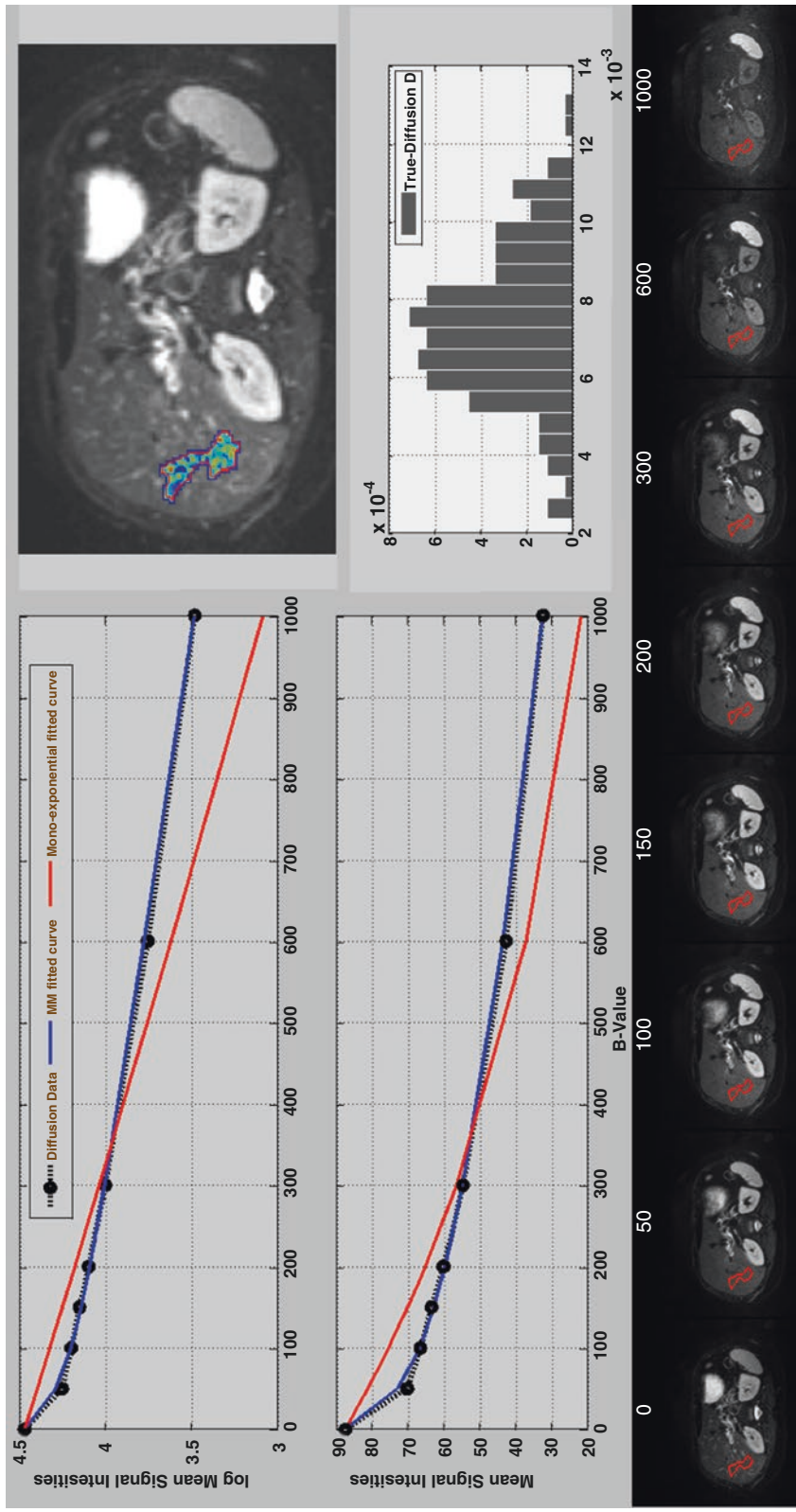


Fig. 7.2 Comparison between mono- and bi-exponential models when applied to normal liver tissue. Top sub-figure displays the mean DWI and the fitted signal intensities of the ROI in semilogarithmic form. The mono-exponential model yields a mean ADC value equal to $1.46*10^{-3}$ mm^2/s, almost doubled to true diffusion (D) from the IVIM model ($0.75*10^{-3}$ mm^2/s). Mean microperfusion (D^*) and perfusion fraction (f) equal to $34.1*10^{-3}$ mm^2/s and 21.77 %, respectively. Notice the poor fitting of mono-exponential model (*red line*) to the experimental data (*back dots*) and the almost perfect fit of the bi-exponential model (*blue line*) that takes into account microperfusion effects [3]

presence of physical barriers like cell membranes or compartments (intracellular and extracellular spaces) restricts the Brownian motion of water [7]. When incorporating the assumption of a restrictive environment, the displacement probability distribution for the water molecules deviates from the Gaussian shape, and the degree of this deviation is quantified by kurtosis. Kurtosis is a dimensionless metric expressing the difference of an arbitrary distribution from a Gaussian with the same variance in terms of more or less weight on the center and tails:

$$K = \frac{M_4}{M_2^2} - 3$$

where M_n is the nth moment of the arbitrary distribution. Similarly, to the ADC, the diffusional kurtosis (K) is not specific for any tissue property, and thus its interpretation in terms of tissue structure is not always well defined. For example, changes in K might be the overall result of more than one complex process in tissue and are unable to identify the precise biological mechanisms behind this change (Fig. 7.3). Several models have been proposed in order to study the value of kurtosis parameters in clinical practice.

The most widespread model for kurtosis stems from the expansion of diffusion tensor imaging (DTI) approach where the exponential decay of the signal is analyzed by Taylor series [8]:

$$\ln\left(S_b\right) = \ln S_0 - b * D(t) + O\left(b^2\right) + \cdots$$

With the introduction of high ($b > 1000$) values in clinical practice, the contribution from the second-order term cannot be considered negligible as when employed for DTI calculations, and the above expression can be rewritten as:

$$\ln\left(S_b\right) = \ln S_0 - b * D + \frac{1}{6} * b^2 * D^2 * K$$

Similarly, when b exceeds a certain upper limit, the latter expression may also suffer from systematic errors in the calculation of D and K from the omission of even higher-order terms.

7.2.1.4 Non-Gaussian Bi-exponential (IVIM-Kurtosis)

The non-Gaussian mono-exponential model may lead to D and K miscalculation in the two-compartment hypothesis unless it is extended to also account for microperfusion [9]. The modified MR signal that takes into account simultaneously the existence of two distinct compartments as well as the presence of structure in tissue into a single framework can be expressed as:

$$S_b = S_0 * \left((1-f) * \exp\left(-b * D + \frac{1}{6} * b^2 * D^2 * K \right) + f * \exp\left(-b * D^* \right) \right)$$

where four parameters (D, D^*, f, and K) (Fig. 7.4) need to be estimated. In order to fit experimental data to this model, the acquisition sequence must be tailored to adequately sample signal decay both at the low-b-value area (0–200 s/mm^2) to capture capillary flow effects and at the high-b-value area (more than 200 s/mm^2) to quantify signal loss related to true diffusion effects. These requirements render the acquisition protocol demanding in terms of succeeding a good compromise between acquisition time, noise level, and spatial resolution.

7.2.1.5 Stretched Exponential Model

Intravoxel heterogeneity in the distribution of diffusion coefficients because of heterogeneity in fluid viscosity or diffusive restrictions has been quantified in the stretched exponential model [10]. The existence of multiple pools rather than only two inside a region of interest (ROI) or pixel, together with proton exchange between pools, has been pinpointed as the reason of mismatch observed between expected volume fractions and fitted results from DWI data [11]. The proposed model assumes continuous distribution

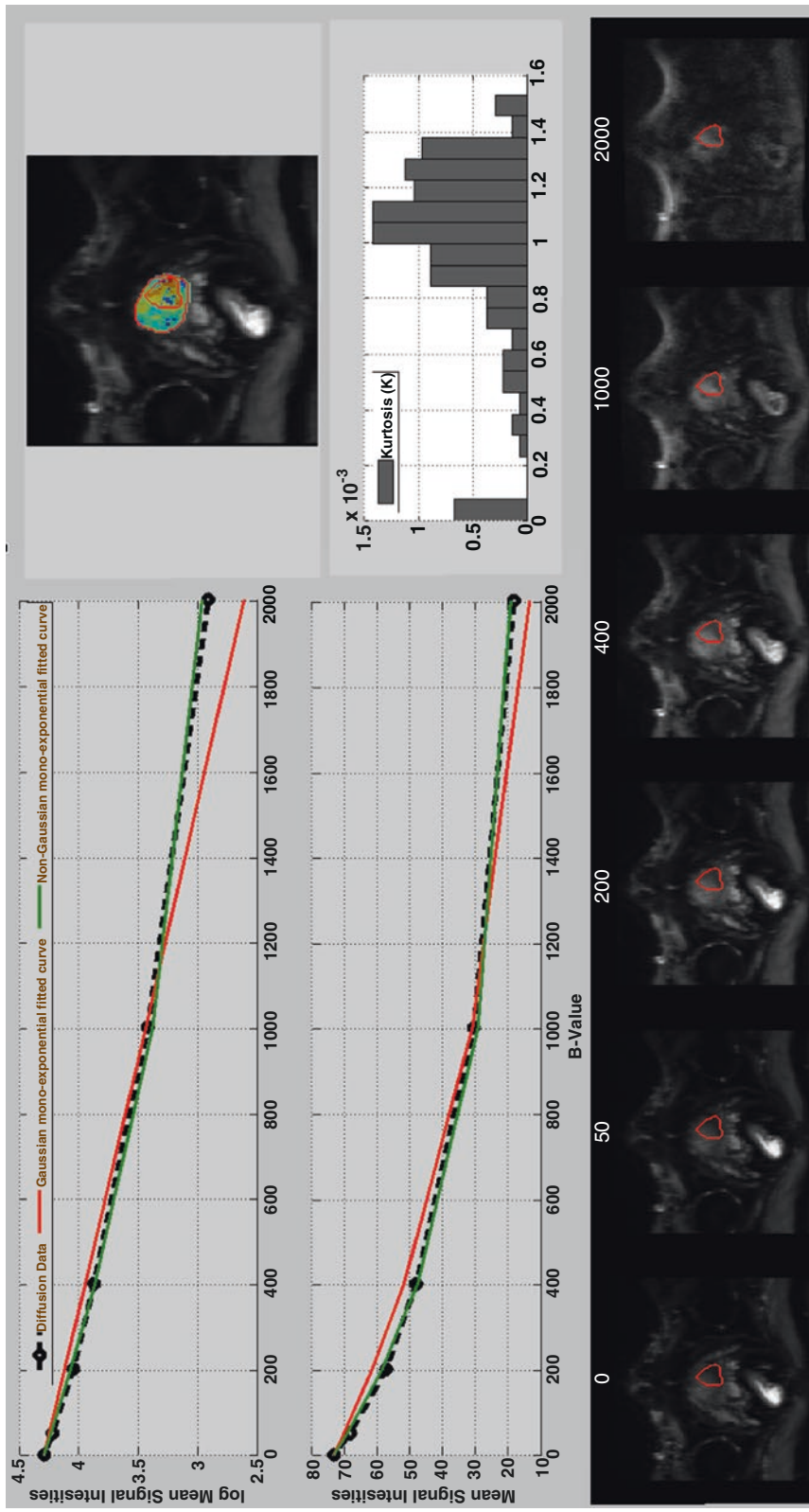


Fig. 7.3 Patient with aggressive prostate cancer (Gleason score 4+4). Simple mono-exponential fitting (*red line*) starts to considerably deviate from experimental data, and at higher *b*-values (>1000), this model is unable to approximate the data without incorporating non-Gaussian term (*green line*), reflecting increased tissue heterogeneity at the site of the tumor (region of interest shown on multiple *b*-values at the lower row) [3]

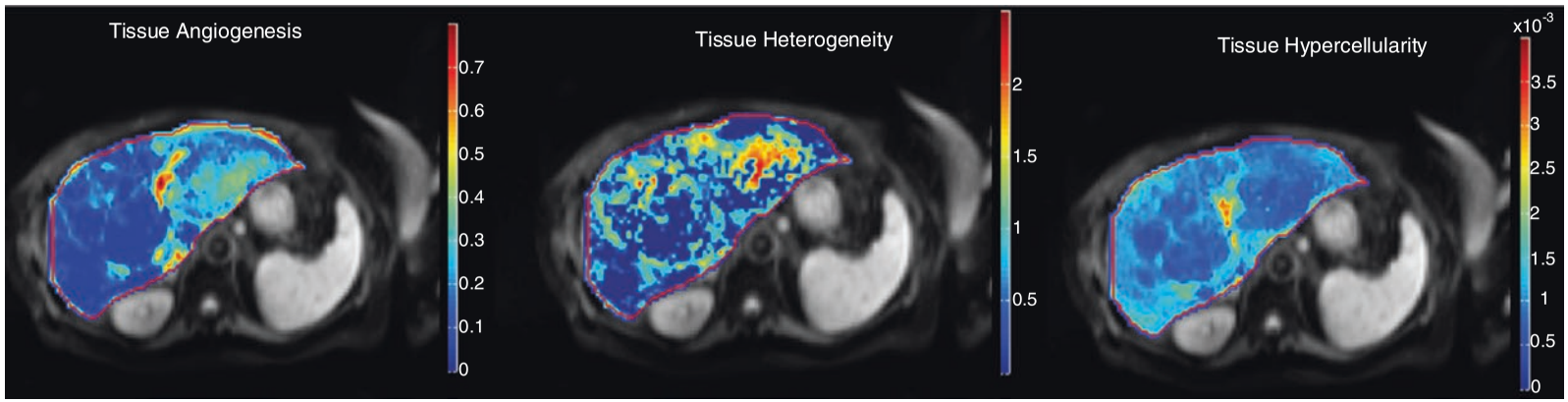


Fig. 7.4 Comprehensive parametric mapping from non-Gaussian bi-exponential model (IVIM-Kurtosis) application in a patient with hepatocellular carcinoma. From *left to right*: a perfusion fraction (*f*) map is shown that reflects tissue angiogenesis information, and a kurtosis (*K*) map reveals tissue heterogeneity, while a true diffusion (*D*) map presents information related to tissue cellularity [3]

of sources decaying at different rates without any restriction in the number of participating sources, and thus the signal attenuation can be attributed to the aggregation of a number of uncoupled decay processes, such that:

$$S_b = S_0 * \exp\left[-(b * \mathrm{ADC})^{\alpha}\right]$$

where α is the stretching parameter and a measure of deviation of the signal decay from the simple mono-exponential behavior. Range of α is from 0 to 1. Lower values of parameter α would imply presence of multiple compartments within the ROI, while at the upper limit ($\alpha = 1$) the model coincides with a simple exponential decay of a homogeneous sample. Incorporation of IVIM theory to the latter model would result in a double-stretched exponential model that has not yet been extensively studied.

7.2.2 DWI Analysis

7.2.2.1 Data Fitting

Several mathematical models have been proposed to quantify the DWI signal decay into diffusion biomarkers. In contradiction to the mono-exponential model and the single ADC biomarker calculated, a more complex mathematical framework is required for fitting the diffusion signal according to the IVIM and the non-Gaussian models. In case of the IVIM and the extended non-Gaussian IVIM model, two main categories are presented in the literature: (a) complete fitting methods for calculating simultaneously all the biomarkers using nonlinear regression models and (b) partial fitting methods that provide biomarkers in a more simplified way based on observations related to the behavior of the true diffusion and microperfusion effects in the *b*-value range used.

Complete Fitting

In order to extract multiple biomarkers from the IVIM and the non-Gaussian models, nonlinear least squares (NLLS) are widely used. The NLLS fitting technique is based on the Levenberg-Marquardt algorithm [12]. NLLS are minimization problems in mathematics that given initial, lower, and upper bounds for the estimated parameters (i.e., *D*, *D**, and f in case of the IVIM model) approximate the diffusion model by a linear one and iteratively refine the values of the parameters to reach their optimal values. The initial point is of crucial significance for the convergence of the algorithm, and a lot of attention has been paid for assessing the precision and uncertainty in the estimation of the diffusion biomarkers as reported in [13].

Partial Fitting

Many studies in the literature use partial fitting for calculating IVIM-related biomarkers. Every partial fitting method relies on the fact that, as stated in the IVIM theory, *D** is roughly one order of magnitude greater than *D* [14] at high *b*-values ($b > 200$ s/mm^2) and therefore the microperfusion term in the IVIM model can be neglected. According to [15], the microperfusion effect in high *b*-values is eliminated, and the IVIM bi-exponential fit equation is simplified to the following mono-exponential where *D* can be obtained linearly using least-squares regression:

$$S_b = S_0 * \exp(-b * \mathrm{ADC})$$

The fitted curve from the mono-exponential model is then extrapolated at $b = 0$, and the ratio between the *x*-intercept and the DW-MRI data at $b = 0$ gives an estimation of the perfusion fraction *f*. Biomarkers *D* and *f* are then substituted into the IVIM bi-exponential equation, and nonlinear least squares are applied to the entire *b*-value range for calculating *D**. Alternatives can be also found in [16].

7.2.2.2 Evaluation of the DWI Models and Comparative Studies

Assessing how well models fit the real diffusion signal in every *b*-value is a crucial step in the analysis. In the absence of ground-truth knowledge of the true values of the estimated biomarkers, an accurate fit of the diffusion data can provide increased confidence in the results. Several statistical measures from the regression analysis can be used for evaluating models'

goodness of fit to the diffusion data. *R-squared* (R^2) is one of the most commonly used statistical measures for assessing the goodness of fit and is given by:

$$R^2 = 1 - \frac{SS_E}{SS_T}$$

where SS_E is the residual sum of squares and SS_T is the total sum of squares. *R-squared* values range from 0 to 1 with 1 indicating a perfect fit and 0 a total dissimilarity. Alternatively, adjusted *R-squared* (Fig. 7.5) can be used since it takes into consideration the number of the *b*-values used in the analysis, as well as the number of the parameters provided by each model. Other statistical measures rely on the root-mean-square error (RMSE) which estimates the differences in signal intensity between the real and the modeled diffusion data in every

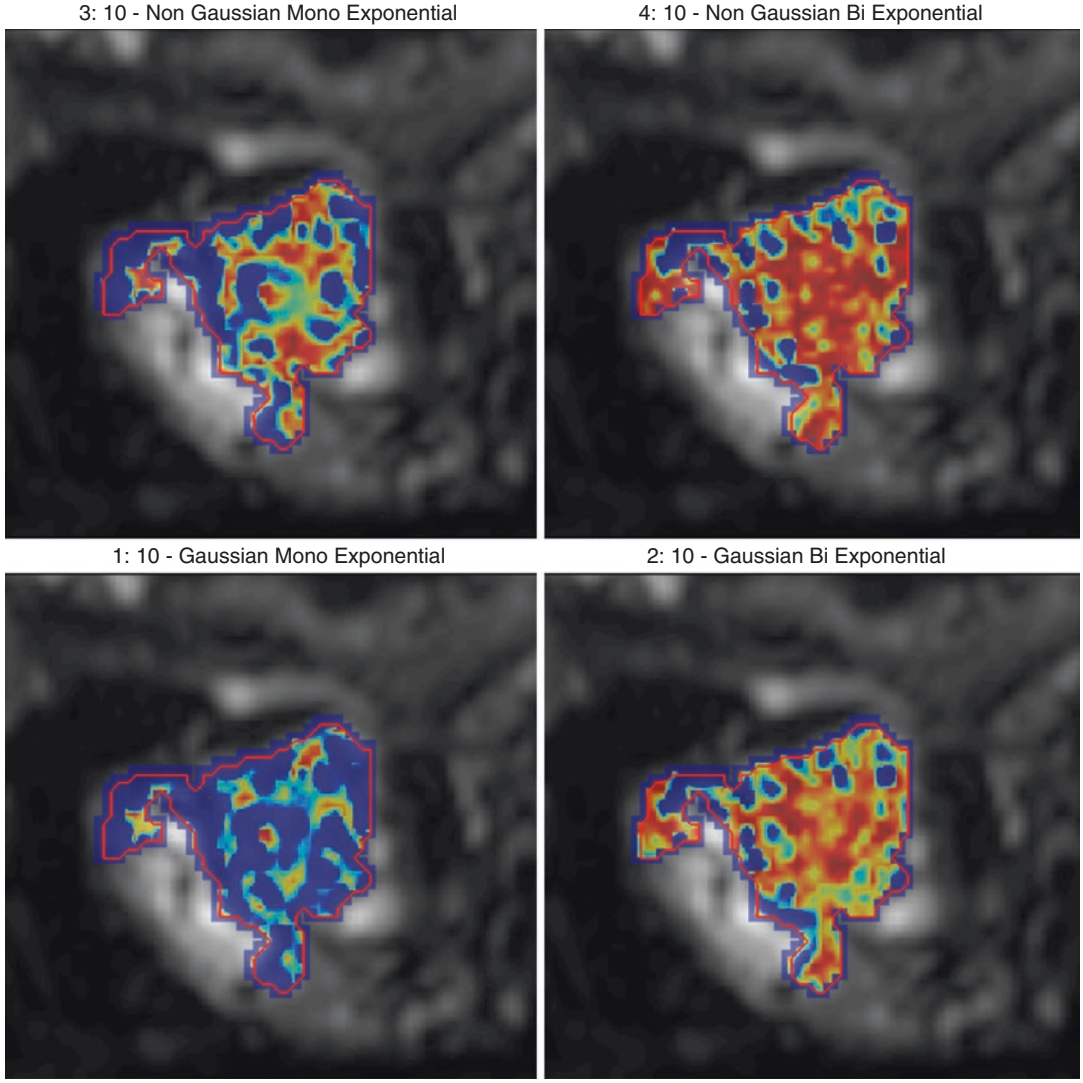


Fig. 7.5 Adjusted *R-squared* in a patient with rectal cancer. The best performing model is the non-Gaussian bi-exponential (*top-right*) where most of the tumoral pixels are rendered with red color (very high adjusted R-squared), while at *top left* the non-Gaussian mono-exponential, *bottom left* the Gaussian mono-exponential, and *bottom right* the Gaussian bi-exponential (IVIM) are shown [3]

b-value, the Akaike information criteria (AIC), and its corrected version (AICc) [17] indicating a balance between the fitness and the simplicity of the model and a Bayesian modification of the AIC known as the Bayesian information criterion (BIC). An extensive analysis is given in [18] in which the fitted quality and repeatability of different diffusion models when applied to normal and cancer prostate data are assessed.

7.2.2.3 Qualitative and Quantitative Data Presentation

Pixel-based calculation of DW-related parameters has the advantage of a detailed view over the area of interest compared to simple ROI techniques [19], but the size of data can be overwhelming and difficult to handle. In order to gain an immediate insight into the complete information contained in the calculated indices, parametric maps and related histograms are produced. The quantification of specific histogram metrics may serve as the ground for statistical interpretation of the results and possibly indicate clinically important biomarkers for certain pathologies (Fig. 7.6). Histogram metrics in some cases are able to discriminate two groups, a task that simple ROI methods fail to achieve [20]. Most histogram analyses use descriptive statistics such as the mean, standard deviation, quartiles, minimum, maximum, skewness, and kurtosis to characterize and compare distributions of the diffusion biomarkers in examined ROIs in a quantitative manner. The most important advantage of histogram comparing to ROI-based analysis is that histogram offers the possibility to evaluate tumor heterogeneity, in other words overcoming the averaging effects that ROI-based analysis is sensitive to.

7.3 Dynamic Contrast-Enhanced (DCE)-MRI

Imaging modalities such as MRI, PET, and CT have been used to noninvasively assess tissue perfusion, by means of imaging biomarkers that can be related to tumor angiogenesis [21]. Sequential acquisition of MRI data image sets and utilization of small molecular weight paramagnetic contrast agents resulted in significant developments in the field of the assessment and monitoring of tumor treatment response [22] (Fig. 7.7).

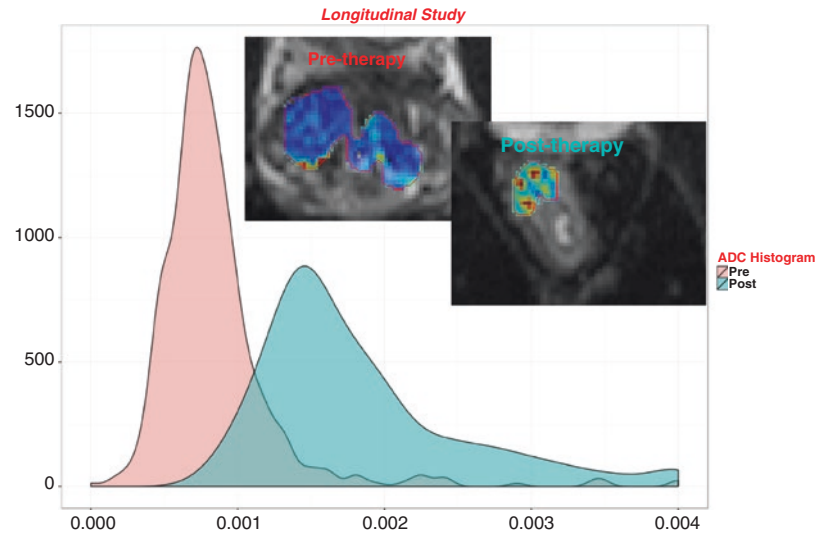


Fig. 7.6 Longitudinal histogram study for a patient with rectal cancer before and after chemoradiation therapy. There is an obvious shift of the ADC values to the right (higher ADC) indicating response to the selected therapy plan. The histogram shift is a comprehensive way to retrieve bulk information from the data that would otherwise demand time-consuming statistical calculations

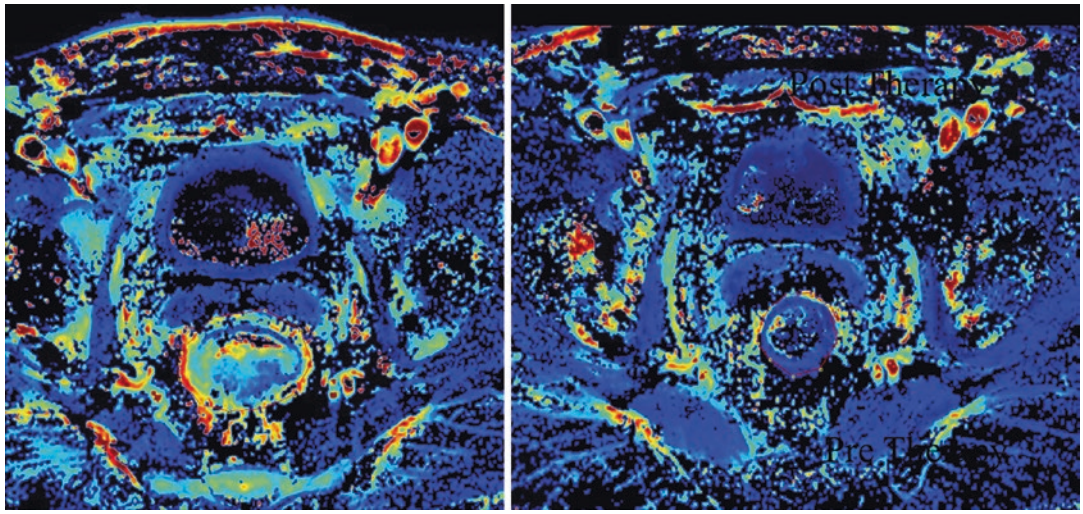


Fig. 7.7 Patient with rectal cancer. K^{trans} maps before (*left*) and after (*right*) chemoradiation therapy. Increased neovascularity is shown on the pretreatment exam that was significantly reduced after treatment

DCE is an MRI method that has been used to assess perfusion in various anatomical regions. It is comprised of the acquisition of dynamic T1-weighted sequences before, during, and after the intravenous injection of a paramagnetic contrast agent (CA). The dynamics of the signal intensity in every pixel of the examined area contain tissue-specific information related to pathophysiology. Main applications include differential diagnosis, grading, and classification of different tumor types [23].

In the clinical routine, radiologists mainly use a subjective evaluation of the shape of the enhancement curves. Although this approach is simple and easy to be performed, it doesn't provide quantifiable markers that can be further processed and used in comparative studies. A variety of analysis techniques have been adopted in the past years for quantifying the DCE-MRI data, which range from simple quantification of the SI enhancement to complex models that measure parameters of the underlying physiology. In the following sections we examine in detail the pharmacokinetic modeling approach that is frequently used in order to provide objective, quantitative measures of tumor physiology based on DCE data analysis.

7.3.1 Reliability of the Pharmacokinetic (PK) Biomarkers

Many factors might affect the reliability of the quantitative biomarkers from a DCE experiment. A first prerequisite is a well-designed DCE protocol, which implies the careful selection of appropriate [24] parameters such as repetition time (TR), echo time (TE), flip angle (FA), and the field of view (FOV). As a general recommendation, TR should be as short as possible in order to ensure high temporal resolution, while TE should be minimized for the elimination of T_2* contribution in the signal of the image. The compromise among spatial and temporal resolution should also be taken in account, as well as the overall acquisition time of the DCE protocol [25].

Concerning the contrast agent, the selection of the type of the tracer is of major importance, as well as the dose injected and the rate of injection [26]. Preprocessing tasks usually include motion correction of the dynamic sequences, where coregistration is carried out with respect to the arterial phase of the DCE, due to the higher signal-to-noise ratio (SNR). Also temporal smoothing of the dynamic curves per pixel can

reduce signal distortions. The selection of ROI for the tissue and artery needs to be annotated from experts, while the role of the hematocrit (HCT) should be taken in account when possible. Critical issues for the accurate quantification of the DCE data include (a) the conversion of SIs of the artery and tissue to CA concentration, (b) the selection of the arterial input function (AIF) region of interest or the assumption of a theoretical one, and (c) the selection of the PK model that the data will be fitted.

7.3.2 Estimation of Contrast Agent Concentration

In some studies, there is a direct processing of the SIs, but this can lead to erroneous results due to nonlinear relationship of SI with concentration, especially in tissues with higher CA concentrations. These nonlinearities are dependent on factors such as native T_1 of the examined tissue and MRI acquisition parameters [27]. It is therefore recommended to convert both tissue and artery SIs to CA concentration for a more robust analysis. The main techniques for measuring T_1 are the inversion recovery [28], the look-locker method using an EPI protocol [29], and the multiple flip angles (FAs) using SPGR protocols [30].

The latter method is widely used in DCE-MRI due to the high SNR and time efficiency offered. It requires two or more FAs for the determination of the pre-contrast relaxation time T_{10} of the tissue. The CA concentration at time t is related to the change in relaxation time via the following formula:

$$C(t) = \frac{1}{r_1} \cdot \left(\frac{1}{T_1(t)} - \frac{1}{T_{10}} \right)$$

where r_1 is the longitudinal relaxivity of the CA, T_{10} is the longitudinal relaxation time prior to CA administration, and $T_1(t)$ is the longitudinal relaxation at time t after the injection of CA.

The MR-SI from a spoiled gradient echo with echo time $TE \ll T_2^*$ is given from the following equation:

$$S_1 = S_0 \cdot \sin(a_f) \frac{1 - e^{-TR/T_{10}}}{1 - \cos(a_f) e^{-TR/T_{10}}}$$

where T_{10} is the longitudinal relaxation time prior to CA administration S_1 is the measured SI, af is the flip angle used, and S_0 is the relaxed signal for a 90° pulse. By acquiring multiple acquisitions (two or more) using different flip angles, the longitudinal relaxation time T_{10} and the relaxed signal S_0 can be estimated, and subsequently the time course of the $T_1(t)$ can also be computed.

7.3.3 Arterial Input Function (AIF)

AIF describes the concentration of the CA during time in the artery that supplies the tissue of interest. The selection of AIF is also critical for the PK modeling procedure and affects the reproducibility and the reliability of the results [31]. In almost all PK models, AIF has to be determined for an accurate analysis. This process may be hampered by a number of issues including the partial volume effect, nonlinear effects, flow artifacts, and patient motion during acquisition. There is a variety of techniques to either directly measure or determine AIF, whereas the most frequently used are discussed below.

Direct measurement of plasma concentration from the field of view (FOV) of the MR image is a method that consists of measuring the AIF from an artery, or from a vein [32], close to the tissue under examination. Having the prerequisite that an artery is included in the FOV of the DCE exam, the SI from an ROI annotated from an experienced operator is converted to CA concentration. The reliability of the AIF measurement in this method is direct dependent on the parameters of the MRI protocol, along with the spatial and temporal resolution of the data set. Moreover, inherent weaknesses of this procedure that can bias the resulted plasma concentration are the partial volume effects [32], inflow effects [33], blood flow pulsatility, and turbulence. Finally, it is important to note that ideally AIF should be determined in an artery close to the tissue under examination and its diameter should exceed the spatial resolution of the dynamic image.

7.3.4 Quantitative Models

Tracer kinetics can be described by systems theory under the assumption that tissues under examination are linear time invariant systems. Under this assumption and considering the intravascular and the extravascular-extracellular space (EES) as two distinct compartments, as shown on Fig. 7.8, the system behavior is described by the following set of equations:

$$C_p(t) = F \cdot R_p(t) \otimes C_a(t)$$

$$C_e(t) = F \cdot R_e(t) \otimes C_a(t)$$

$$C_t(t) = F \cdot \left(R_p(t) + R_e(t)\right) \otimes C_a(t)$$

where $C_a(t)$ is the AIF, $C_p(t)$ is the CA concentration in the vascular bed, $C_e(t)$ is the CA concentration in the EES, $C_t(t)$ is the total tissue concentration, F represents the blood flow, $R(t)$ is the corresponding impulse response function, and $\otimes$ represents convolution. There are a number of models that under several assumptions fit the DCE-MRI data into this set of equations, thus yielding quantitative measurements of tissue physiology.

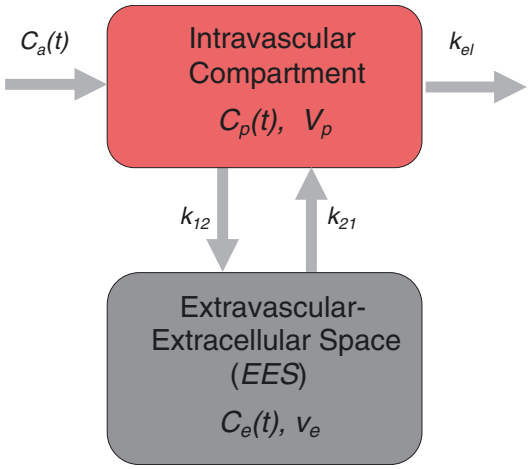


Fig. 7.8 A general representation of the two-compartment model. The first compartment illustrates the plasma with fractional volume v_p and the corresponding concentration $C_p(t)$. The second compartment represents the tissue space, which consists of the extravascular-extracellular space (EES), with corresponding fractional volume v_e and concentration $C_e(t)$. k_{el} is the elimination rate of CA through the kidneys and k_{12}, k_{21} are the exchange rates of CA between the compartments

7.3.4.1 Tofts Model (TM) and Extended Tofts Model (ETM)

One of the most used models in the literature is the Tofts model (TM) [34], a single-compartment model where the plasma space is ignored and the CA is moving to the parenchyma with a rate proportional to the difference of the concentrations between plasma and EES. The model is described by the following equation:

$$\frac{dC_t(t)}{dt} = k^{trans} \cdot \left(C_a(t) - \frac{C_t(t)}{v_e} \right)$$

Using the convolution theorem, the solution of previous equation is given next:

$$R_p(t) = 0$$

$$R_e(t) = k^{trans} / F \cdot e^{-k_{ep} \cdot t}$$

In the above equations, k^{trans} represents the volume transfer constant from plasma space to EES, v_e is the fractional volume of EES, and $k_{ep} = k^{trans}/v_e$ is the transfer constant from EES to the plasma space. The omission of the plasma space was invalid for many tissues and resulted in erroneous PK biomarkers; thus, Tofts extended the original model by introducing the vascular term as an external compartment. The result was to separate the enhancement caused by contrast leakage from that caused by intravascular contrast. The extended Tofts model (ETM) [35] is described by the following equation:

$$R_p(t) = v_p \cdot \delta(t) / F$$

$$R_e(t) = k^{trans} / F \cdot e^{-k_{ep} \cdot t}$$

where v_p is the fractional volume of vascular space. A major weakness of these models is that

k^{trans} can be depicted either as plasma flow in flow-limited cases or as tissue permeability in permeability-limited cases. Another drawback is that TM is accurate only in weakly vascularized tissues, while ETM is also accurate in highly perfused tissues, whereas in intermediate instances the validity of these models is ambiguous [36]. However, what led to widespread usage of these models is the simplicity on their interpretation and the fact that they provide useful biomarkers even in data sets acquired in low temporal resolution.

7.3.4.2 Adiabatic Tissue Homogeneity (ATH) Model

This model [37] belongs to the spatially distributed kinetic models due to the fact that it accounts both temporal and spatial distribution of the CA concentration. The adiabatic approximation is used, which assumes that CA concentration in EES varies slow in comparison with that in intravascular space, and as a result, $C_t(t)$ can be assumed as constant during small-time intervals Δt. Using this assumption a closed-form solution can be derived in the time domain, while its predecessor, the tissue homogeneity (TH) [38], provides closed-form solution only in Laplace space. This is a more complex model compared with TM/ETM, but it also accounts plasma flow rate, extraction fraction, and mean capillary transit time. The impulse response functions for the vascular and parenchyma compartments of this model are given by the following equations:

$$R_p(t) = \left(1 - u\left(t - t_c\right)\right)$$

$$R_e(t) = \left(E \cdot e^{-k_{ep}\left(t - t_c\right)} \cdot u\left(t - t_c\right)\right)$$

7.3.4.3 Two-Compartment Exchange Model (2CXM)

This is a two-compartment model [39] that models plasma and EES as different compartments. Moreover, it allows for separate estimation of permeability and blood flow, thus resulting in a four-dimensional estimated vector [v_e, v_p, F_p, E]. The model is described by the following set of equations:

$$R_p(t) = \left(1 - E_+\right) \cdot e^{-t/T_-}$$

$$R_e(t) = E_+ \cdot e^{-t/T_+}$$

where parameters E_+, T_-, and T_+ are functions of F_p, v_p, v_e, and PS. The main advantage of this model is the separate estimation of regional blood flow and capillary permeability. On the other side, the limitation of the 2CXM is the assumption of the well-mixed tissue compartments, as well as the complexity of the fitting procedure due to the plethora of estimated parameters. The initialization of the estimated vector, the temporal sampling, and the SNR of the data are of major significance for the reliability of the model parameters.

7.3.5 Model-Free Analysis

All model-free methods for DCE-MRI analysis analyze directly the patterns of the dynamic curves, either considering the mean of the ROI curves or analyzing every pixel as a single entity. Widely used markers are the maximum of the SI, the time to peak, and the wash-in and washout rates [40, 41], while several of these parameters have been proposed as measurements that are correlated with response to treatment [42]. Another type of model-free analysis is by considering the shape of the tissue curves and classifying these in predefined tissue types. The shape of the tissue curve can vary across a wide range of patterns, while the most common are:

- Type 1: steady enhancement of the SI
- Type 2: enhancement followed by plateau
- Type 3: enhancement followed by washout

Type 3 curves are usually considered as the most suspicious for detecting cancer; however type 1 and type 2 may also be part of the cancerous tissue. Model-free methods for DCE-MRI for a promising analysis framework have

been gaining ground over the last years. Its main benefit is the increased robustness and the reduced sensitivity to protocol parameters and user interference.

Conclusion

DW-MRI is an important imaging modality that offers unique diagnostic information related to tissue microarchitecture. Hypercellularity, hypoxia, necrosis, and tumor aggressiveness can be evaluated based on diffusion-based imaging biomarkers. DWI can be used to (i) improve lesion conspicuity thus improving sensitivity to the detection of pathology, (ii) assist in differential diagnosis, (iii) improve lesion staging, and (iv) evaluate therapeutic outcome.

DCE-MRI is an imaging modality that combined with post-processing methods can give an added value to the assessment of the exam. A lot of methods have been proposed for the analysis and interpretation of the dynamic curves, with the two major groups being the model-based PK modeling and the model-free approaches using PR techniques. In this section, a review of the methods and their pitfalls has been presented. For the right choice and application of each method, the quality of the data and the examined organ should be taken into account.

References

1. Padhani R, Liu G, Koh D-M, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins DJ, Hammoud DA, Rustin GJS, Taouli B, Choyke PL. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia. 2009;11(2):102–25.
2. Guo AC, Cummings TJ, Dash RC, Provenzale JM. Lymphomas and high-grade astrocytomas: comparison of water diffusibility and histologic characteristics. Radiology. 2002;224(16):177–83.
3. Manikis GC, Nikiforaki K, Papanikolaou N, Marias K. Diffusion Modelling Tool (DMT) for the analysis of Diffusion Weighted Imaging (DWI) Magnetic Resonance Imaging (MRI) data. CGI'16, Proceedings of the 33rd Computer Graphics International. 2016; 97-100.
4. Le Bihan D, Breton E, Lallemand D, Aubin M-L, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168(2):497–505.
5. Leenders KL, Perani D, Lammertsma AA, Heather JD, Buckingham P, Healy MJ, Gibbs JM, Wise RJ, Hatazawa J, Herold S. Cerebral blood flow, blood volume and oxygen utilization: normal values and effect of age. Brain. 1990;113:27–47.
6. Riches SF, Hawtin K, Charles-Edwards EM, de Souza NM. Diffusion-weighted imaging of the prostate and rectal wall: comparison of biexponential and monoexponential modelled diffusion and associated perfusion coefficients. NMR Biomed. 2009;22(3):318–25.
7. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed. 2010;23(7):698–710.
8. Liu C, Bammer R, Acar B, Moseley ME. Characterizing Non-Gaussian diffusion by using generalized diffusion tensors. Magn Reson Med. 2004;51(5):924–37.
9. Lu Y, Jansen JFA, Mazaheri Y, Stambuk HE, Koutcher JA, Shukla-Dave A. Extension of the intravoxel incoherent motion model to non-gaussian diffusion in head and neck cancer. J Magn Reson Imaging. 2012;36(5):1088–96.
10. Bennett KM, Schmainda KM, Bennett R, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. Magn Reson Med. 2003;50(4):727–34.
11. Quirk JD, Bretthorst GL, Neil JJ, Ackerman JJH, Measurement of transmembrane water exchange rate for the in situ rat brain. 9, 2001.
12. Levenberg K. A method for the solution of certain non-linear problems in least squares. Q Appl Math. 1944;2:164–68.
13. Sasaki M, Sumi M, Eida S, Katayama I, Hotokezaka Y, Nakamura T. Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors. PLoS One. 2014;9(11):e112866.
14. Le Bihan D. Intravoxel incoherent motion perfusion MR imaging: a wake-up call. Radiology. 2008;249(3):748–52.
15. Yuan J, Yeung DKW, Mok GSP, Bhatia KS, Wang YXJ, Ahuja AT, King AD. Non-Gaussian analysis of diffusion weighted imaging in head and neck at 3T: A pilot study in patients with nasopharyngeal carcinoma. PLoS One. 2014;9(1):1–8.
16. Suo S, Lin N, Wang H, Zhang L, Wang R, Zhang S, Hua J, Xu J, Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: comparison of different curve-fitting methods. J Magn Reson Imaging. 2015;42(2):362–370.
17. Cavanaugh JE. Unifying the derivations for the Akaike and corrected Akaike information criteria. Stat Probab Lett. 1997;33(2):201–8.
18. Quentin M, Blondin D, Klasen J, Lanzman RS, Miese FR, Arsov C, Albers P, Antoch G, Wittsack HJ. Comparison of different mathematical models of

diffusion-weighted prostate MR imaging. Magn Reson Imaging. 2012;30(10):1468–74.
19. Just N. Improving tumour heterogeneity MRI assessment with histograms. Br J Cancer. 2014;111(12):2205–13.
20. Gaing B, Sigmund EE, Huang WC, Babb JS, Parikh NS, Stoffel D, Chandarana H. Subtype differentiation of renal tumors using voxel-based histogram analysis of intravoxel incoherent motion parameters. Invest Radiol. 2015;50(3):144–52.
21. Jackson A, O'Connor JPB, Parker GJM, Jayson GC. Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging. Clin Cancer Res. 2007;13(12):3449–59.
22. Kontopodis E, Kanli G. Assessing treatment response through generalized pharmacokinetic modeling of DCE-MRI data. Cancer Inform. 2015;14:41–51.
23. Øvrebø KM, Hompland T, Mathiesen B, Rofstad EK. Assessment of hypoxia and radiation response in intramuscular experimental tumors by dynamic contrast-enhanced magnetic resonance imaging. Radiother Oncol. 2012;102(3):429–35.
24. De Naeyer D, Verhulst J, Ceelen W, Segers P, De Deene Y, Verdonck P. Flip angle optimization for dynamic contrast-enhanced MRI-studies with spoiled gradient echo pulse sequences. Phys Med Biol. 2011;56(16):5373–95.
25. Kershaw LE, Cheng H-LM. Temporal resolution and SNR requirements for accurate DCE-MRI data analysis using the AATH model. Magn Reson Med. 2010;64(6):1772–80.
26. Aerts HJWL, Van Riel NAW, Backes WH. System identification theory in pharmacokinetic modeling of dynamic contrast-enhanced MRI: Influence of contrast injection. Magn Reson Med. 2008;59(5):1111–9.
27. Dale BM, Jesberger JA, Lewin JS, Hillenbrand CM, Duerk JL. Determining and optimizing the precision of quantitative measurements of perfusion from dynamic contrast enhanced MRI. J Magn Reson Imaging. 2003;18(5):575–84.
28. Ordidge RJ, Gibbs P, Chapman B, Stehling MK, Mansfield P. High-speed multislice T1 mapping using inversion-recovery echo-planar imaging. Magn Reson Med. 1990;16(2):238–45.
29. Henderson E, McKinnon G, Lee TY, Rutt BK. A fast 3D look-locker method for volumetric T1 mapping. Magn Reson Imaging. 1999;17(8):1163–71.
30. Cheng H-LM, Wright GA. Rapid high-resolution T1 mapping by variable flip angles: accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magn Reson Med. 2006;55:566–76.
31. Yang X, Liang J, Heverhagen JT, Jia G, Schmalbrock P, Sammet S, Koch R, Knopp MV. Improving the pharmacokinetic parameter measurement in dynamic contrast-enhanced MRI by use of the arterial input function: theory and clinical application. Magn Reson Med. 2008;59(6):1448–56.
32. Hansen AE, Pedersen H, Rostrup E, Larsson HBW. Partial volume effect (PVE) on the arterial input function (AIF) in T 1 -weighted perfusion imaging and limitations of the multiplicative rescaling approach. Magn Reson Med. 2009;62(4): 1055–9.
33. Peeters F, Annet L, Hermoye L, Van Beers BE. Inflow correction of hepatic perfusion measurements using T1-weighted, fast gradient-echo, contrast-enhanced MRI. Magn Reson Med. 2004;51(4):710–7.
34. Tofts PS, Kermode AG. Measurement of the blood–brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. Magn Reson Med. 1991;17(2):357–67.
35. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging. 1997;7:91–101.
36. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. Magn Reson Med. 2011;66(3):735–45.
37. Lawrence KSS, Lee T. An adiabatic approximation to the tissue homogeneity model. J Cereb Blood Flow Metab. 1998;18(12):1378–85.
38. Johnson JA, Wilson TA. A model for capillary exchange. Am J Physiol. 1966;210(6):1299–303.
39. Zhou J, Wilson DA, Ulatowski JA, Traystman RJ, van Zijl PC. Two-compartment exchange model for perfusion quantification using arterial spin tagging. J Cereb Blood Flow Metab. 2001;21(4):440–55.
40. Kim JK, Hong SS, Choi YJ, Park SH, Ahn H, Kim CS, Cho KS. Wash-in rate on the basis of dynamic contrast-enhanced MRI: usefulness for prostate cancer detection and localization. J Magn Reson Imaging. 2005;22(5):639–46.
41. Preziosi P, Orlacchio A, Di Giambattista G, Di Renzi P, Bortolotti L, Fabiano A, Cruciani E, Pasqualetti P. Enhancement patterns of prostate cancer in dynamic MRI. Eur Radiol. 2003;13(5):925–30.
42. Padhani AR, Gapinski CJ, Macvicar DA, Parker GJ, Suckling J, Revell PB, Leach MO, Dearnaley DP, Husband JE. Dynamic contrast enhanced MRI of prostate cancer: correlation with morphology and tumour stage, histological grade and PSA. Clin Radiol. 2000;55(2):99–109.

Imaging Biomarker Measurements

8

Bernard E. Van Beers, Benjamin Leporq,
Sabrina Doblas, and Philippe Garteiser

8.1 Introduction

Imaging biomarkers are health or disease markers based on quantitative imaging parameters. With high-throughput computing, it is now possible to extract numerous quantitative features from computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET) images. The conversion of digital medical images into mineable high-dimensional data is called radiomics and is motivated by the concept that biomedical images contain information that reflects underlying pathophysiology [1, 2]. The image measurements are based on size, volume, and shape assessment and on signal intensity and heterogeneity (texture) analysis.

8.2 Size Measurements

The simple clinically used metrics to assess lesion evolution include two-diameter (World Health Organization, WHO) and more recently, one-diameter (Response Evaluation Criteria in Solid Tumors, RECIST) measurements [3–5]. For the last 15 years, the international cancer community has extensively employed the RECIST criteria at CT to assess the response exhibited by patient's tumor on exposure to both marketed and experimental antitumor therapies [6]. The calculated response is categorized as complete response (disappearance of tumor), partial response (change between −100 and −30 %), stable disease (change between −30 and + 20 %), or progressive disease (increase of 20 % or greater). RECIST quantification of response correlates with patient survival and disease-free survival, showing its clinical usefulness [6].

However, RECIST criteria have several shortcomings. First, tumor evolution is linear, rather than polytomous. As cutoffs to define partial response or progressive disease are artificial, quantitative measurements are superior to semiquantitative category assessment for studying tumor progression [6–9].

Second, the reproducibility of manual measurements may be suboptimal and may be improved by semiautomatic size measurements [10]. In a study of large lung tumors, it was shown that the 95 % limits of inter-observer agreement (−39–28 %) of maximum diameter measurements were outside the range of clinical acceptability (<20 % according to the RECIST guidelines) at CT, whereas the corresponding automated measurements (−8.0–11 %) were within clinical acceptable range [11].

B.E. Van Beers (✉) • B. Leporq • S. Doblas
P. Garteiser
Laboratory of Imaging Biomarkers, UMR 1149
INSERM – University Paris Diderot, Paris, France

Beaujon University Hospital Paris Nord,
Paris, France
e-mail: bernard.van-beers@aphp.fr

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_8

Third, RECIST size measurements do not always accurately reflect tumor response, especially when molecular therapies or other targeted therapeutic interventions such as chemoembolization are used [12, 13]. This is explained by the fact that these treatments mainly cause tumor necrosis, with little or no size decrease.

Alternative response criteria have been developed in these cases. These criteria include the mRECIST criteria, in which one diameter of the viable, contrast-enhancing, tumor regions is measured; the European Association for the Study of Liver Disease (EASL) criteria in which two diameters of the enhancing regions are measured; and the Choi criteria in which decrease in tumor size and decrease in tumor density at CT are assessed [14].

In patients with hepatocellular carcinomas, the Choi criteria have been shown to be superior to the RECIST, mRECIST, and EASL criteria to assess treatment response [15]. This underscores the fact that combining signal intensity measurements with size measurements may increase the diagnostic value relative to size measurements alone. With the Choi method, however, the signal attenuation measurements are obtained as the mean value within a region of interest. This region of interest analysis provides only part of the information as tumor heterogeneity is not explicitly described.

8.3 Lesion Segmentation

For more complete quantitative assessment of lesions, feature measurements within the whole lesion volume are needed. Three-dimensional volume segmentation is a critical and challenging component of whole lesion analysis. It is critical because subsequent parameters are generated from the segmented volumes. It is challenging because many tumors have indistinct borders.

Multiple segmentation algorithms have been applied in medical imaging studies. Popular ones are based on boundary or active contour definition [1, 16], region-growing or level-set methods [17, 18], and k-means clustering approaches [19, 20].

Active contours consist of positioning a contour larger than the region to be segmented and iteratively repositioning its points until a convergence criterion is met. The convergence criterion may be based on the geometry of the contour, thereby introducing prior knowledge on the shape of the segmented region [21], on the intensity and spatial variations thereof over the underlying region [22], or on a combination of both types of information. Region-growing approaches, and their advanced counterparts, namely, level-set methods, consist of starting an iterative process on an initial position for the region of interest. This region is then augmented or "grown," by adding neighboring pixels to it. Addition of pixels is conditioned positively if the resulting, larger region remains homogeneous, and negatively if the homogeneity decreases, indicative of a boundary [23]. Finally, k-means clustering approaches rely on Euclidean measures of distances between extracted parameters (pixel intensity or other pixel-wise derived metrics) to generate pixel clusters corresponding to homogeneous regions [24].

Accuracy and reproducibility are important factors to evaluate segmentation algorithms for medical images. However, accuracy is difficult to determine because the reference method is often based on manual segmentation, which is subjective, error prone, and time consuming. Objective volume measurements during surgery are better gold standards but are rarely obtained [17]. In other words, "ground truth" segmentation often does not exist.

Hence, reproducibility is more important than accuracy. Several studies have shown that the reproducibility of semiautomatic segmentation algorithms is superior to that of manual segmentation [11, 17, 18, 25]. A consensus is emerging that optimum reproducible segmentation is achievable with computer-aided edge detection followed by manual curation [2].

8.4 Shape-Based Measurements

Quantitative features describing the geometric shape of a lesion can be extracted from the three-dimensional surface of the rendered volumes.

Measures of compactness, spherical disproportion, sphericity, surface-to-volume ratio, and Zernike moments describe the shape of the lesion [26–28].

8.5 Intensity and Texture Analyses

Intensity and texture analyses can be divided into four families based on the distribution of signal intensity, on the organization of gray level in the spatial domain, on the organization of geometric patterns in the spatial domain, and on analysis performed in the frequency domain.

8.5.1 Analysis Based on the Distribution of Signal Intensity

This analysis is based on first-order statistics which describe the distribution of values of individual voxels without concern for spatial relationships. These are generally histogram-based methods and reduce a region of interest to single values. The parameters include the mean, median, maximum and minimum values, nth centiles, standard deviation, variance, mean absolute deviation, uniformity (uniformity of gray-level distribution), entropy (irregularity of gray-level distribution), skewness (asymmetry of the histogram), and kurtosis.

8.5.2 Analysis Based on the Organization of Signal Intensity in the Image Domain

This analysis provides second-order descriptors which describe statistical interrelationships between voxels with similar or dissimilar contrast values. The spatial distribution of voxel intensities is calculated from gray-level co-occurrence (GLCM) or gray-level run-length texture matrices (GLRLM).

GLCM determines how often a pixel of intensity i finds itself within a certain relationship to another pixel of intensity j (Fig. 8.1). Second-order statistics based on a co-occurrence matrix (GLCM) include autocorrelation, contrast, correlation, cluster prominence, cluster shade, cluster tendency, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy, etc. [29]. The energy (pixel repetition) expresses the regularity of the texture. High energy is observed when the high values in the GLCM are concentrated in some precise locations. It is the case for images with constant or periodic gray-level distributions. A random or noisy image gives a GCLM with more distributed values and a low energy. The contrast is more elevated for GCLM with larger values outside the diagonal, thus for images with local variation of intensities.

The dissimilarity expresses the same characteristic than the contrast, but the weights of inputs of the GCLM increase linearly from the diagonal rather than quadratically for the contrast. These two descriptors are thus often correlated.

The entropy (randomness of the matrix) relies to the spreading of the GCLM diagonal. The entropy is the inverse of energy. These parameters are often correlated.

The homogeneity (uniformity of co-occurrence matrix) inversely evolves with the contrast. Homogeneity is high when the differences between co-occurrences are small. It is more sensitive on the diagonal elements of the GCLM than the contrast which depends on elements outside the diagonal.

The correlation may be described as a measurement of the linear dependency of gray levels of the image. The cluster shade and cluster prominence give information about the degree of symmetry of the GCLM. High values represent low symmetric pattern.

The main difficulty when using GCLM is to fix the parameters because this step needs to be performed case by case. The distance d must reflect the local correlation between the pixels. It is admitted that the correlation is more pertinent for short distances and, typically, d is fixed equal to 1. In practice, GCLM is computed over four orientations (i.e., 0°, 45°, 90°, and 135°) according to Haralick recommendations [29]. The

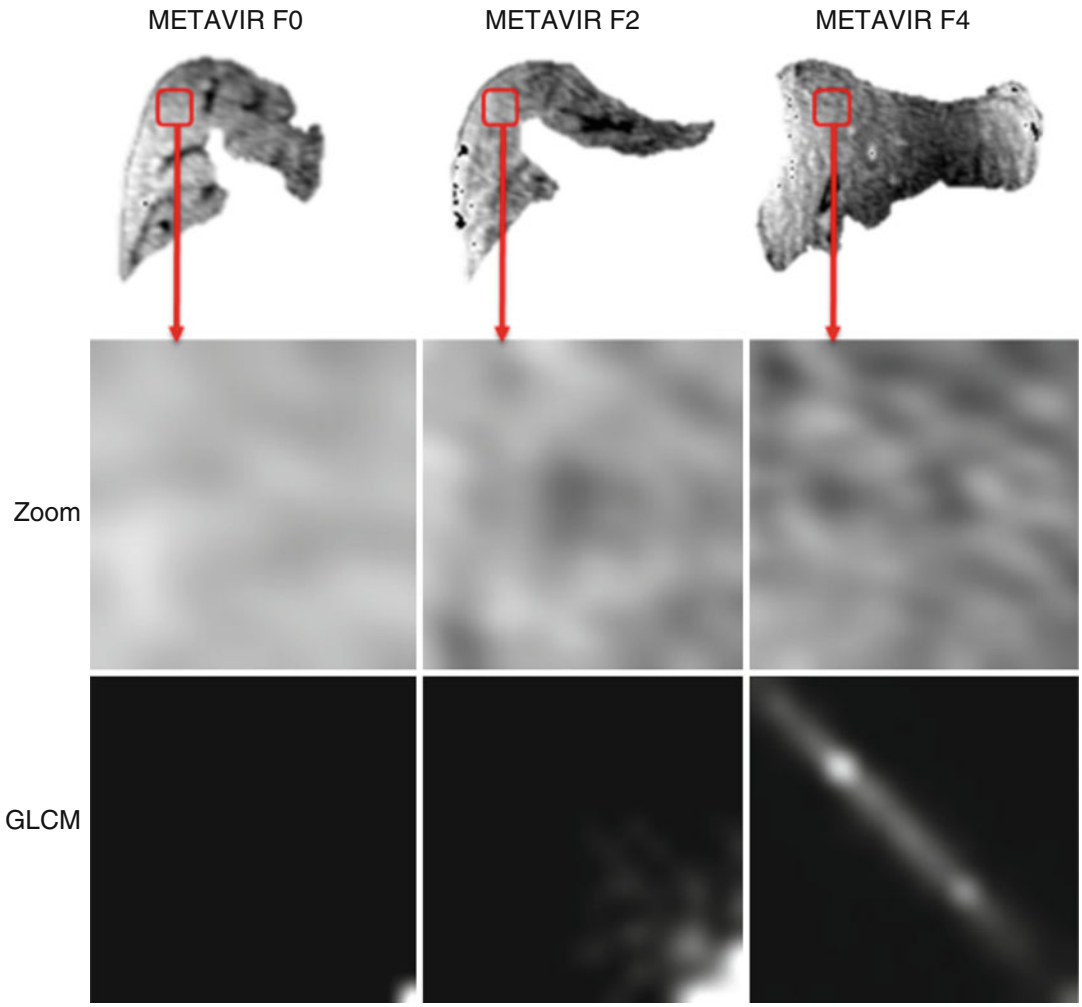


Fig. 8.1 Texture analysis of gadoxetic acid-enhanced MR images in patient with chronic liver disease. This figure shows differences in the GLCM according to the severity of liver fibrosis. Second-order descriptors derived from this matrix can offer quantitative information relevant for the assessment of liver fibrosis

features are computed for each orientation and can be concatenated in a single array of descriptors or averaged to obtain an array of descriptors invariant regarding to the rotation. The choice of the window (i.e., the number of gray levels in the parametric image) is also important and imposes a compromise between the pertinence of the descriptors and the fidelity of the texture.

Another method to derive second-order statistics is the gray-level run-length matrix (GLRLM). A gray-level run is defined as the length in number of pixels of consecutive pixels that have the same gray-level value. From the GLRLM, features can be extracted describing short- and long-run emphasis, gray-level nonuniformity, run-length nonuniformity, run percentage, low gray-level run emphasis, and high gray-level run emphasis [1, 28]. The short-run emphasis characterizes the smoothness of the texture, whereas the long-run emphasis characterizes the coarseness. The run percentage is the ratio between the number of runs over the number of pixels in the image. It characterizes the homogeneity of the texture. The gray-level nonuniformity measures the uniformity of run distribution. It is minimal when the runs are uniformly distributed between

the gray levels. The run-length nonuniformity measures the uniformity of run length and increases with the number of runs of same length.

Other matrices have been proposed to characterize the texture in the spatial domain such as the gray-level size zone matrix (GLSZM). GLSZM does not require computation in several directions, in contrast to GLRLM and GLCM. However, the degree of gray-level quantization has an important impact on the texture classification performance. Similarly to GLRLM, descriptors can be derived from the analysis of this matrix such as the small-zone size emphasis, large-zone size emphasis, low gray-level zone emphasis, high gray-level zone emphasis, small-zone low-gray emphasis, small-zone high-gray emphasis, large-zone low-gray emphasis, large-zone high-gray emphasis, gray-level nonuniformity, zone size nonuniformity, and zone size percentage [30].

8.5.3 Analysis Based on the Organization of Geometric Pattern in the Image Domain

Filter grids can be applied on the images to extract repetitive or non-repetitive patterns. These methods include fractal analysis, wherein patterns are imposed on the images and the number of grid elements containing voxels of a specified value is computed; Minkowski functionals, which assess patterns of voxels whose intensity is above a threshold [31]; and Laplacian transforms of Gaussian band-pass filters that extract areas with increasingly coarse texture patterns from the images [32].

8.5.4 Texture Analysis in the Frequency Domain

These methods use filtering tools such as the Fourier transform, the wavelet decomposition, and the Gabor filter to extract the information. The 2D Fourier transform allows to represent the frequency spectrum on images in which each coefficient corresponds to a frequency in a given orientation. Therefore, the center of the spectra includes the low frequencies and the extremities the high frequencies. An image with a smooth texture will display a spectrum with high values concentrated close to the center, whereas an image with a rough texture will display a spectra with high value concentrated at the extremities. Quantitative information related to the texture can be extracted by decomposing the spectra into sub-bands according to their polar coordinates and calculating the average, energy, variance, and maximum [33]. The Fourier transform can also be applied in local neighbors in the image. It is possible to determine a radial spectrum on windows with increasing size by averaging the coefficient of the Fourier spectrum over all orientations. A principal component analysis is performed to identify the range of frequencies and the size window explaining the variability [34]. The Fourier spectrum only contains frequency information.

In contrast, Gabor filters and the wavelet transforms provide both frequency and spatial information. Gabor filters have the ability to model the direction and frequency sensitivity by decomposing the image spectrum in a narrow range of frequencies and orientations. In the spatial domain, the Gabor filter is a Gaussian function modulated by a complex sinusoid and a Gaussian surface centered on a central frequency F with an orientation θ in the frequency domain. A conventional practice with Gabor filters consists in using filter banks, each centered on a different central frequency and orientation, by covering the whole frequency domain. Each pixel gives a response for each filter. To have a different proportion covered by each filter and to limit the overlap, thus the redundancy of information, Manjunath and Ma have proposed to decompose the spectrum in several scales and orientations [35]. Mean and standard deviation of the filter responses are calculated to extract the texture signature.

Nevertheless, due to the non-orthogonality of Gabor filters, texture attributes derived from these filters can be correlated. It is difficult to determine if a similarity observed between the analysis scales is linked to the property of the

image or to redundancy in the information. Thus for each scale of application, parameters defining the filter must be modified.

This issue is addressed by the use of wavelets, offering a uniform analysis framework by decomposing the image into orthogonal and independent sub-bands. Briefly, the wavelet decomposes the image with a series of functions obtained by translation and scaling from an initial function, called mother wavelet. Wavelet decomposition of an image is the convolution product between the image and the wavelet functions [31].

8.6 Data Reduction

The number of descriptive image features can approach the complexity of data obtained with gene expression profiling. With such large complexity, there is a danger of overfitting analyses, and hence, dimensionality must be reduced by prioritizing the features. Dimensionality reduction can be divided into feature extraction and feature selection. Feature extraction transforms the data in the high-dimensional space to a space of fewer dimensions, as in principal component analysis.

Feature selection techniques can be broadly grouped into approaches that are classifier dependent (wrapper and embedded methods) and classifier independent (filter methods). Wrapper methods search the space of feature subsets, using the training/validation accuracy of a particular classifier as the measure of utility for a candidate subset. This may deliver significant advantages in generalization, though has the disadvantage of a considerable computational expense, and may produce subsets that are overly specific to the classifier used. As a result, any change in the learning model is likely to render the feature set suboptimal. Embedded methods exploit the structure of specific classes of learning models to guide the feature selection process. These methods are less computationally expensive, and less prone to overfitting than wrappers, but still use quite strict model structure assumptions.

In contrast, filter methods evaluate statistics of the data independently of any particular classifier, thereby extracting features that are generic, having incorporated few assumptions. Each of these three approaches has its advantages and disadvantages, the primary distinguishing factors being speed of computation, and the chance of overfitting. In general, in terms of speed, filters are faster than embedded methods which are in turn faster than wrappers. In terms of overfitting, wrappers have higher learning capacity so are more likely to overfit than embedded methods, which in turn are more likely to overfit than filter methods.

A primary advantage of filters is that they are relatively cheap in terms of computational expense and are generally more amenable to a theoretical analysis of their design. The defining component of a filter method is the relevance index quantifying the utility of including a particular feature in the set. The filter-based feature selection methods can be divided into two categories: univariate methods and multivariate methods. In case of univariate methods, the scoring criterion only considers the relevancy of features ignoring the feature redundancy, whereas multivariate methods investigate the multivariate interaction within features, and the scoring criterion is a weighted sum of feature relevancy and redundancy [36–38].

One of the simplest methods relies on the computation of cross correlation matrices, whereby the correlation between each pair of features is computed (Fig. 8.2). The resulting matrix is subsequently thresholded to identify subsets of features that are highly correlated.

A single feature from each subset can then be selected based on maximum relevancy.

8.7 Data Classification

For data mining, unsupervised and supervised analysis options are available. The distinction in these approaches is that unsupervised analysis does not use any outcome variable, but rather provides summary information and graphical representations of the data. Supervised analysis,

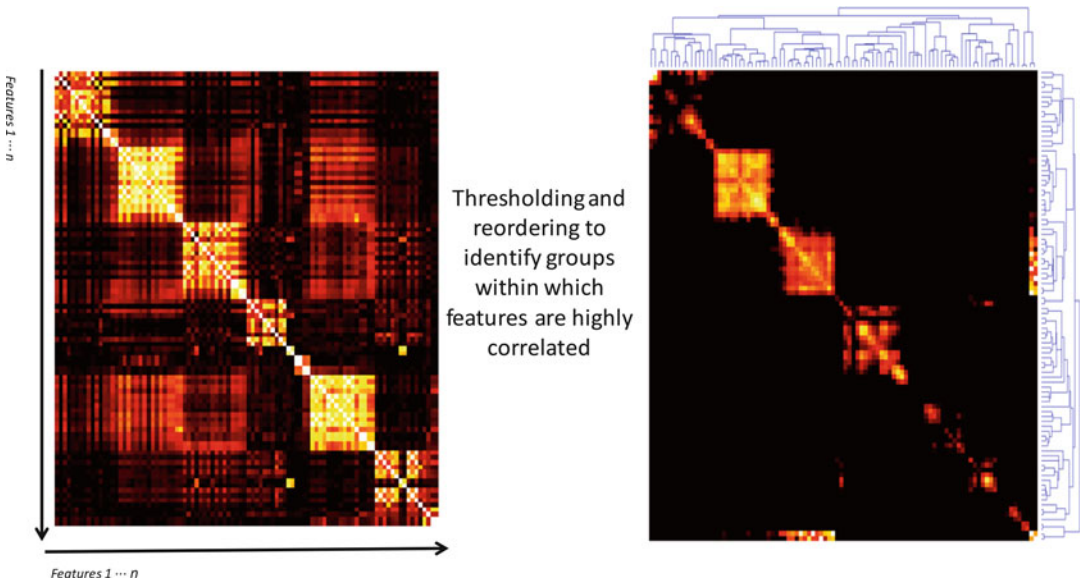


Fig. 8.2 Illustration of the feature selection process. The cross correlation matrix on the *left* is reordered with linkage algorithms on the *right* and thresholded to a given value of correlation coefficient. For data analysis, one feature of each group on the right can be selected based on maximum relevancy, e.g., maximum patient interpatient variability

in contrast, creates models that attempt to separate or predict the data with respect to an outcome or phenotype.

Clustering is the grouping of like data and is one of the most common unsupervised analysis approaches. There are many different types of clustering. Hierarchical clustering, or the assignment of examples into clusters at different levels of similarity into a hierarchy of clusters, is a common type. Similarity is based on correlation (or Euclidean distance) between individual examples or clusters.

Alternatively, k-means clustering is based on minimizing the clustering error criterion which for each point computes its squared distance from the corresponding cluster center and then takes the sum of these distances for all points in the data set.

The data from this type of analyses can be graphically represented using the cluster color map. Cluster relationships are indicated by treelike structures adjacent to the color map or by k-means cluster groups [24, 39] (Fig. 8.3).

Supervised analysis consists of building a mathematical model of an outcome or response variable. The breadth of techniques available is remarkable and includes neural networks, decision trees, classification, and regression trees as well as Bayesian networks [40, 41]. Model selection is dependent on the nature of the outcome and the nature of the training data.

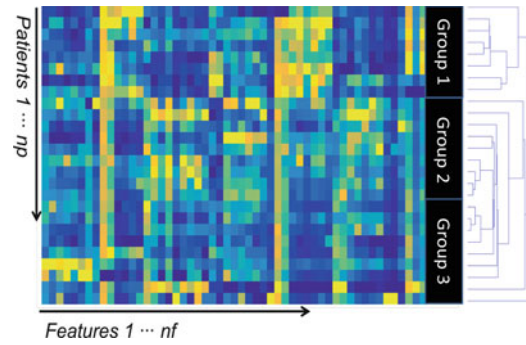


Fig. 8.3 Graphical representation of a radiomics data set. Each patient represents a row of the matrix (np, number of patients), and each column represents one of the features (nf, number of features). First-order imaging parameters based on MR elastography data acquired at several mechanical frequencies in patients with liver cirrhosis and portal hypertension. The hierarchical cluster relationships are indicated by treelike structures on the right of the matrix representation. Alternatively, clustering by a k-means algorithm can be used to group patients into like groups, indicated by groups one to three in the black boxes

Performance in the training data set is always upward biased because the features were selected from the training data set. Therefore, a validation data set is essential to establish the likely performance in the clinic. Preferably, validation data should come from an external independent institution or trial [41]. Alternatively, one may evaluate machine learning algorithms on a particular data set, by partitioning the data set in different ways. Popular partition strategies include k-fold cross validation, leave-one-out, and random sampling [42].

The best models are those that are tailored to a specific medical context and, hence, start out with a well-defined end point. Robust models accommodate patient features beyond imaging. Covariates include genomic profiles, histology, serum biomarkers, and patient characteristics [2].

As a general rule, several models should be evaluated to ascertain which model is optimal for the available data [38, 43]. Recently, Ypsilantis et al. [44] have compared the performance of two competing radiomics strategies: an approach based on state-of-the-art statistical classifiers (logistic regression, gradient boosting, random forests, and support vector machines) using over 100 quantitative imaging descriptors, including texture features as well as standardized uptake values and a convolutional neural network, trained directly from PET scans by taking sets of adjacent intra-tumor slices. The study was performed for predicting response to neoadjuvant chemotherapy in patients with esophageal cancer, from a single 18F-FDG-PET scan taken prior to treatment. The limitation of the statistical classifiers originates from the fact that the performance is highly dependent on the design of the texture features, thus requiring prior knowledge for a specific task and expertise in hand-engineering the necessary features. By contrast, convolutional neural networks operate directly on raw images and attempt to automatically extract highly expressive imaging features relevant to a specific task at hand. In the Ypsilantis et al. study, convolutional neural networks achieved 81 % sensitivity and 82 % specificity in predicting nonresponders and outperformed the other competing predictive models. These results suggest the potential superiority of the fully automated method. However, further testing using larger data sets is required to validate the predictive power of convolutional neural networks for clinical decision-making.

Indeed, it should be noted that machine learning techniques in radiology are still in infancy. Many machine learning studies were done using relatively small data sets. The proposed methods may not generalize well from small data sets to large data sets. To solve the problem, re-training the algorithm will be necessary, but it requires intervention of knowledgeable experts which hinders the deployment of machine learning-based systems in hospitals or medical centers. One possible solution would be utilizing incremental learning and adjusting the computerized systems in an automatic way. In addition, increased large-scale data may bring computational issues to radiology applications. Machine learning techniques employed in these applications may not scale well as training data increases [42].

8.8 Radiomics

Radiomics mines and deciphers numerous medical imaging features. The hypothesis being that these imaging features are augmented with critical and interchangeable information regarding tumor phenotype [28]. Texture is especially important to assess in tumors. Indeed, the tumor signal intensity is very heterogeneous and reflects its structural and functional features, including the number of tumor cells, quantity of inflammation and fibrosis, perfusion, diffusion, and mechanical properties, as well as metabolic activity. Functional parameters which are hallmarks of cancer include sustaining proliferative signaling, resisting cell death, inducing angiogenesis, activating invasion and metastasis, and deregulating cellular energetics [45]. These hallmarks can be assessed with quantitative MR imaging, including perfusion and diffusion MR imaging, MR elastography and susceptibility, and FDG-PET [46, 47].

During the recent years, it became increasingly evident that genetic heterogeneity is a basic feature of cancer and is linked to cancer evolution [48]. This heterogeneity which evolves during time concerns not only the tumor cells but also their microenvironment [49]. Moreover, it has been shown that the global gene expression patterns of human cancers may systematically correlate with their dynamic imaging features [50]. Tumors are thus characterized by regions habitats with specific combinations of blood flow, cell density, necrosis, and edema. Clinical imaging is uniquely suited to measure temporal and spatial heterogeneity within tumors [51], and this information may have predictive and prognostic value.

Spatial heterogeneity is found between different tumors within individual patients (inter-tumor heterogeneity) and within each lesion in an individual (intra-tumor heterogeneity). Intra-tumor heterogeneity is near ubiquitous in malignant tumors, but the extent varies between patients. Intra-tumor heterogeneity tends to increase as tumors grow. Moreover, established spatial heterogeneity frequently indicates poor clinical prognosis. Finally, intra-tumor heterogeneity may increase or decrease following efficacious anticancer therapy, depending on underlying tumor biology [52].

Several studies have shown that tumor heterogeneity at imaging may predict patient survival or response for treatment [53–59].

For instance, in 41 patients with newly diagnosed esophageal cancer treated with combined radiochemotherapy, Tixier et al. showed that textural features of tumor metabolic distribution extracted from baseline 18F-FDG-PET images allowed for better prediction of therapy response than first-order statistical outputs (mean, peak, and maximum SUV) [60].

In 26 colorectal cancer liver metastases, O'Connor et al. showed that three perfusion parameters, namely, the median extravascular extracellular volume, the heterogeneity parameters corresponding to tumor-enhancing fraction, and the microvascular uniformity (assessed with the fractal measure box dimension), explained 86% of the variance tumor shrinkage after FOLFOX therapy [61]. This underscores that measuring microvascular heterogeneity may yield important prognostic and/or predictive biomarkers.

Zhou et al. showed in 32 patients with glioblastoma multiforme that spatial variations in T1 post-gadolinium and either T2-weighted or fluid-attenuated inversion recovery at baseline MR imaging correlated significantly with patient survival [62].

8.9 Limitations of Radiomics

Several issues arise when interpreting imaging data of heterogeneity. First, some voxels suffer from partial volume averaging, typically at interface with non-tumor tissue. Second, there is inevitable compromise between having sufficient numbers of voxels to perform the analysis versus sufficiently large voxels to overcome noise and keep imaging times practical. Most methods of analysis require hundreds to thousands of voxels for robust application. Third, CT, MR imaging, or PET voxels are usually non-isotropic (slice thickness exceeds in-plane resolution). Dimensions are typically 200–2,000 μm for rodent models and 500–5,000 μm for clinical tumors. Compared with genomic and histopathology biomarkers, this represents many orders of magnitude difference in scale, making it difficult to validate image heterogeneity biomarkers against pathology [52].

Variations in image parameters affect the information being extracted by image feature algorithms, which in turn affects classifier performance (Fig. 8.4) [63]. At PET imaging, Yan et al. [64] analyzed the effect of several acquisition parameters on the heterogeneity values. They found that the voxel size affected the heterogeneity value the most, followed by the full width at half maximum of the Gaussian post-processing filter applied to the reconstructed images. Neither the number of iterations nor the actual reconstruction scheme affected the heterogeneity values much.

Because of the information dependence on variations in image parameters, imaging standardization and reproducibility are important issues to determine the effectiveness of image

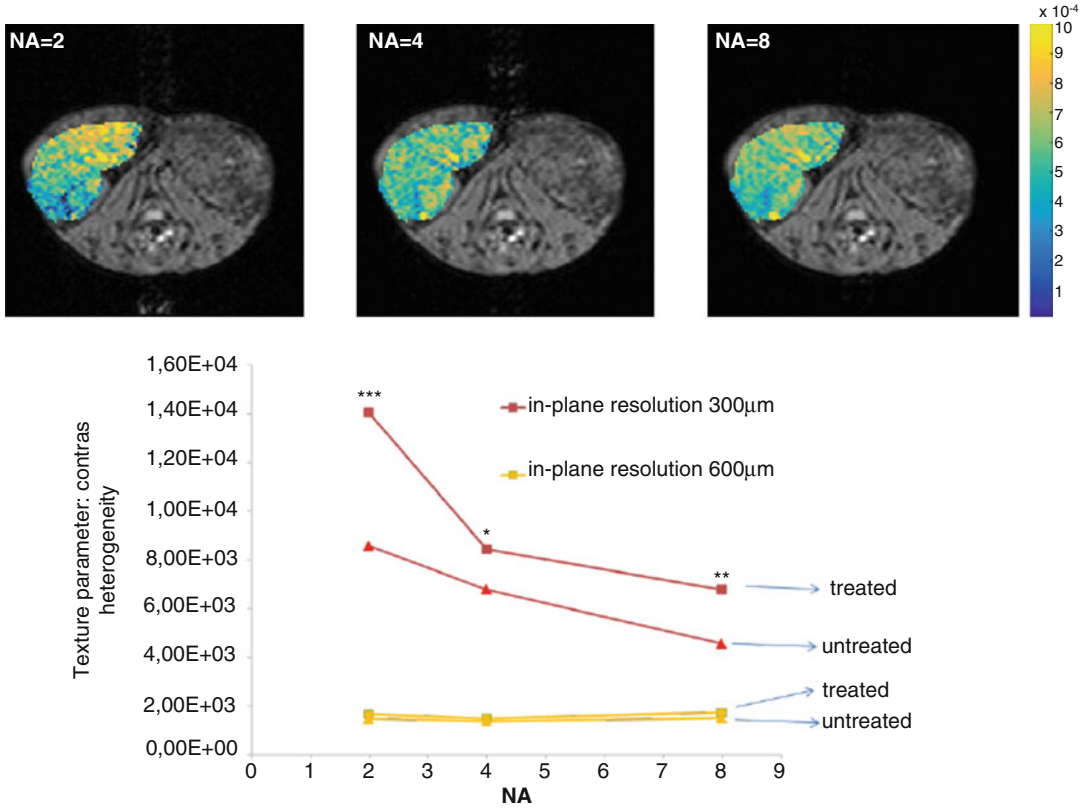


Fig. 8.4 The behavior of a texture parameter acquired on diffusion coefficient maps was assessed in a HepG2 tumor treated with an adipokine that competitively inhibits the fatty acid-binding protein Fabp4. The effects of in-plane resolution and number of averages, were explored. The treated tumor had significantly higher texture on the high-resolution data set regardless of signal-to-noise ratio. On the low-resolution data sets, adipokine treatment did not appear to have an effect. These data show that spatial resolution and signal to noise ratios (manipulated here through varying number of averages (NA)) may affect texture analysis

features being developed and prediction models built to work on those feature values.

Another problem in radiomics and genomics is related to multiple testing issues. In many data sets in these areas, it is not unusual to test the significance of thousands of variables using 50 samples. Any single test may have a low expected false-positive rate; however, the cumulative effect of many repeated tests guarantees that many statistically significant findings are due to random chance (type I errors in statistics should be < 5 %). Chalkidou et al. reported a systematic review of the type I error inflation in texture analysis derived from PET or CT images [65]. After applying appropriate statistical corrections, an average type I error probability of 76 % was estimated with the majority of published results not reaching statistical significance. This underscores that the multiple testing problem may be critical. It has been addressed in statistics in many ways. However, the best way to overcome overfitting and optimism in predictive performance is to evaluate the performance of the model in an external validation cohort, as explained above [66].

Conclusion

Current knowledge suggests that radiomics can enhance individualized treatment selection and monitoring. Furthermore, unlike genomics-based approaches, radiomics is noninvasive and comparatively cost-effective.

Radiomics is thus an innovative and encouraging breakthrough toward the realization of precision medicine. Fast-computing and state-of-the-art software have facilitated the collection and analysis of large amounts of data, while the development of data mining techniques enables researchers to test a large number of hypotheses simultaneously. The high number of image analysis algorithms and image-derived features is promising to unravel complex biology by overcoming the limitations inherent in invasive tissue sampling techniques. However, the high data dimensionality complicates the quantitative analysis, and robust biological and statistical validation is needed before advanced radiomics solutions can be used in the clinics.

References

1. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30(9):1234–48.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures they are data. Radiology. 2016;278(2):563–77. Epub 2015/11/19.
3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer. 1981;47(1):207–14. Epub 1981/01/01.
4. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst. 2000;92(3):205–16. Epub 2000/02/03.
5. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228–47. Epub 2008/12/23.
6. Jain RK, Lee JJ, Ng C, Hong D, Gong J, Naing A, et al. Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies. J Clin Oncol. 2012;30(21):2684–90. Epub 2012/06/13.
7. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. J Natl Cancer Inst. 2007;99(19):1455–61. Epub 2007/09/27.
8. Raymond E, Dahan L, Raoul JL, Bang YJ, Borbath I, Lombard-Bohas C, et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N Engl J Med. 2011;364(6):501–13. Epub 2011/02/11.
9. Sharma MR, Maitland ML, Ratain MJ. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox---point. Cancer Res. 2012;72(20):5145–9; discussion 50. Epub 2012/09/07.
10. Bonekamp D, Bonekamp S, Halappa VG, Geschwind JF, Eng J, Corona-Villalobos CP, et al. Interobserver agreement of semi-automated and manual measurements of functional MRI metrics of treatment response in hepatocellular carcinoma. Eur J Radiol. 2014;83(3):487–96. Epub 2014/01/07.
11. Dinkel J, Khalilzadeh O, Hintze C, Fabel M, Puderbach M, Eichinger M, et al. Inter-observer reproducibility of semi-automatic tumor diameter measurement and volumetric analysis in patients with lung cancer. Lung Cancer. 2013;82(1):76–82. Epub 2013/08/13.
12. Le Cesne A, Van Glabbeke M, Verweij J, Casali PG, Findlay M, Reichardt P, et al. Absence of progression as assessed by response evaluation criteria in solid tumors predicts survival in advanced GI stromal tumors treated with imatinib mesylate: the intergroup EORTC-ISG-AGITG phase III trial. J Clin Oncol. 2009;27(24):3969–74. Epub 2009/07/22.
13. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis. 2010;30(1):52–60. Epub 2010/02/23.
14. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol. 2007;25(13):1753–9. Epub 2007/05/02.
15. Ronot M, Bouattour M, Wassermann J, Bruno O, Dreyer C, Larroque B, et al. Alternative Response Criteria (Choi, European association for the study of the liver, and modified Response Evaluation Criteria in Solid Tumors [RECIST]) Versus RECIST 1.1 in patients with advanced hepatocellular carcinoma treated with sorafenib. Oncologist. 2014;19(4):394–402. Epub 2014/03/22.
16. Michoux N, Vallee JP, Pechere-Bertschi A, Montet X, Buehler L, Van Beers BE. Analysis of contrast-enhanced MR images to assess renal function. MAGMA. 2006;19(4):167–79. Epub 2006/08/15.
17. Hermoye L, Laamari-Azjal I, Cao Z, Annet L, Lerut J, Dawant BM, et al. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. Radiology. 2005;234(1):171–8. Epub 2004/11/27.
18. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One. 2014;9(7), e102107. Epub 2014/07/16.

19. Michoux N, Simoni P, Tombal B, Peeters F, Machiels JP, Lecouvet F. Evaluation of DCE-MRI postprocessing techniques to assess metastatic bone marrow in patients with prostate cancer. Clin Imaging. 2012;36(4):308–15. Epub 2012/06/26.
20. Michoux N, Van den Broeck S, Lacoste L, Fellah L, Galant C, Berliere M, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. BMC Cancer. 2015;15:574. Epub 2015/08/06.
21. Pathak SD, Chalana V, Kim Y. Interactive automatic fetal head measurements from ultrasound images using multimedia computer technology. Ultrasound Med Biol. 1997;23(5):665–73. Epub 1997/01/01.
22. Sebbahi A, Herment A, de Cesare A, Mousseaux E. Multimodality cardiovascular image segmentation using a deformable contour model. Comput Med Imaging Graph. 1997;21(2):79–89. Epub 1997/03/01.
23. Hojjatoleslami SA, Kittler J. Region growing: a new approach. IEEE Trans Image Process. 1998;7(7):1079–84. Epub 2008/02/16.
24. Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. Pattern Recogn. 2003;36:451–61.
25. Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology. 2013;266(3):812–21. Epub 2012/12/12.
26. Tahmasbi A, Saki F, Shokouhi SB. Classification of benign and malignant masses based on Zernike moments. Comput Biol Med. 2011;41(8):726–35. Epub 2011/07/05.
27. Yap FY, Bui JT, Knuttinen MG, Walzer NM, Cotler SJ, Owens CA, et al. Quantitative morphometric analysis of hepatocellular carcinoma: development of a programmed algorithm and preliminary application. Diagn Interv Radiol. 2013;19(2):97–105. Epub 2012/12/13.
28. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006. Epub 2014/06/04.
29. Haralick RMSK, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;3(6):610–21.
30. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Bio-Med Eng. 2014;61(3):630–7. Epub 2013/10/11.
31. Alberich-Bayarri A, Marti-Bonmati L, Angeles Perez M, Sanz-Requena R, Lerma-Garrido JJ, Garcia-Marti G, et al. Assessment of 2D and 3D fractal dimension measurements of trabecular bone from high-spatial resolution magnetic resonance images at 3 T. Med Phys. 2010;37(9):4930–7. Epub 2010/10/23.
32. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging. 2012;3(6):573–89. Epub 2012/10/25.
33. Augusteijn MF, Clements L, Shaw KA. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. IEEE Trans Geosci Remote Sens. 1995;33(3):616–26.
34. Proisy C, Couteron P, Fromard F. Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. Remote Sens Environ. 2007;109:379–92.
35. Manjunath BS, Ma WY. Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell. 1996;18(8):837–42.
36. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005;27(8):1226–38. Epub 2005/08/27.
37. Brown G, Pocock A, Zhao MJ, Lujan M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. J Mach Learn Res. 2012;13(1):27–66.
38. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol. 2015;5:272. Epub 2015/12/24.
39. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863–8. Epub 1998/12/09.
40. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers. J Biomed Inform. 2003;36(4-5):389–99. Epub 2003/12/04.
41. Lambin P, Zindler J, Vanneste BG, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. Adv Drug Deliv Rev. 2016. pii:S0169-409X(16)30008-4. Epub 2016/01/18.
42. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933–51. Epub 2012/04/03.
43. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. Sci Rep. 2015;5:13087. Epub 2015/08/19.
44. Ypsilantis PP, Siddique M, Sohn HM, Davies A, Cook G, Goh V, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. PLoS One. 2015;10(9), e0137036. Epub 2015/09/12.
45. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74. Epub 2011/03/08.
46. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for

PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1:122S–50. Epub 2009/06/24.
47. Van Beers BE, Daire JL, Garteiser P. New imaging techniques for liver diseases. J Hepatol. 2015;62(3):690–700. Epub 2014/12/03.
48. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;501(7467):338–45. Epub 2013/09/21.
49. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature. 2013;501(7467):346–54. Epub 2013/09/21.
50. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol. 2007;25(6):675–80. Epub 2007/05/23.
51. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. Radiology. 2013;269(1):8–15. Epub 2013/09/26.
52. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21(2):249–57. Epub 2014/11/26.
53. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A. 2008;105(13):5213–8. Epub 2008/03/26.
54. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology. 2012;264(2):387–96. Epub 2012/06/23.
55. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. Radiology. 2013;266(1):177–84. Epub 2012/11/16.
56. Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. J Nucl Med. 2015;56(1):38–44. Epub 2014/12/17.
57. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med. 2013;54(1):19–26. Epub 2012/12/04.
58. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114(3):345–50. Epub 2015/03/10.
59. Fehr D, Veeraraghavan H, Wibmer A, Gondo T, Matsumoto K, Vargas HA, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A. 2015;112(46):E6265–73. Epub 2015/11/19.
60. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52(3):369–78. Epub 2011/02/16.
61. O'Connor JP, Rose CJ, Jackson A, Watson Y, Cheung S, Maders F, et al. DCE-MRI biomarkers of tumour heterogeneity predict CRC liver metastasis shrinkage following bevacizumab and FOLFOX-6. Br J Cancer. 2011;105(1):139–45. Epub 2011/06/16.
62. Zhou M, Hall L, Goldgof D, Russo R, Balagurunathan Y, Gillies R, et al. Radiologically defined ecological dynamics and clinical outcomes in glioblastoma multiforme: preliminary results. Transl Oncol. 2014;7(1):5–13. Epub 2014/04/29.
63. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? J Nucl Med. 2015;56(11):1642–4. Epub 2015/08/22.
64. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med. 2015;56(11):1667–73. Epub 2015/08/01.
65. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One. 2015;10(5), e0124165. Epub 2015/05/06.
66. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Br J Surg. 2015;102(3):148–58. Epub 2015/01/30.

9 Detecting Measurement Biases: Sources of Uncertainty, Accuracy, and Precision of the Measurements

Jose Miguel Carot and Andrea Conchado

9.1 Introduction

Measurement in health sciences helps practitioners to predict relevant clinical outcomes through different treatments and groups of patients. With this aim, clinical endpoints assess subjective health status, functional status, well-being, and quality of life from the point of view of a patient participating in a study or a clinical trial. Clinical endpoints include all types of patient-based outcome measures, such as questionnaires and scales, to report the patient's subjective experience of health and the consequences of illness, such as symptom severity scales, satisfaction, distress, or difficulty. There is a consensus that these clinical endpoints should be used as the primary endpoints of all clinical research. However, the exclusive use of clinical endpoints poses some challenges for predicting clinical outcomes. For instance, survival is regarded as the main gold standard clinical endpoint, though its use as a primary clinical endpoint is highly unpractical due to its infrequent occurrence. Similarly, other unambiguous clinical outcomes, such as stroke or occurrence of predefined infections, have also been suggested as primary endpoints, although they usually do not offer clear and unambiguous information about the success of medical interventions [40].

These difficulties in the use of primary clinical endpoints have fostered the development of biomarkers in recent years. Biomarkers are objective and quantifiable measurements reflecting "an interaction between a biological system and a potential hazard" [48]. Instrumental or surgical outcomes, all causes of mortality, or indicators of laboratory procedures are good examples of widely used biomarkers. Biomarkers have frequently been used in clinical trials as surrogate endpoints, that is to say, as a stand-in (not a replacement) of clinical endpoints. However, biomarkers used as surrogate endpoints must be statistically examined to check whether they are able to predict precisely and accurately clinical outcomes, being *precise* a synonym of *reliable* and *accurate* a synonym for *valid* [42]. There is a generalized consensus that the quality of any measurement instrument should be described by its reliability and validity, although other domains regarding measurement have been recently proposed.

This chapter will deal with the ability of biomarkers to obtain reproducible measures (reliability), to measure the construct of interest (validity), and to detect small differences between scores (responsiveness). Other properties, such

J.M. Carot (✉) • A. Conchado
Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, Camino de Vera s/n, Edificio 7A, 46022 Valencia, Spain
e-mail: jcarot@eio.upv.es

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_9

as the ability to obtain meaningful measures (interpretability), the degree of acceptance by patients (acceptability), and the extent of effort demanded of staff for its use (feasibility), fall outside the scope of this study and deserve research and analysis in their own right [16].

9.2 What Makes a Biomarker Accurate and Precise?

Good measurement in health and clinical sciences is not just a matter of numbers; it is much more complex than that. Any biomarker is subject to uncertainty, which requires that there is not just one truth about the clinical outcome we try to predict. Instead, there are infinite values, ranging in the so-called confidence intervals, within the true value of the clinical outcome in question where it is very likely to be contained. These confidence intervals are usually built consistently with the data and the knowledge of the aspect being measured [26].

Uncertainty in health care may be caused by a large number of different factors, which can be classified into their fundamental sources, issues, and locus, according to the three-dimensional taxonomy proposed by Han, Klein, and Arora [22]. The first dimension, sources of uncertainty, includes the lack of knowledge about the probabilities of a particular event, the ambiguity caused by imprecisions, the presence of conflicting opinions or lack of information, and the complexity of the phenomena being studied. The second dimension, uncertainty issues, deals with the difficulties associated to data analysis for each patient, such as diagnosis and prognosis procedures, causal explanations, and treatment recommendations. From a system-centered approach, structures and processes of care have also been considered in this dimension, as well as other psychosocial and existential uncertainties, more centered in the patient. The last dimension, locus of uncertainty, tries to clarify the respective informational needs and roles of patients and clinicians. It is not necessary for practitioners to discuss the nature of uncertainty in biomarkers, nor is it necessary for them to take control of all potential sources of error. But they, as most professionals working in the field of science usually do, must be aware of the presence of uncertainty in his (or her) analysis.

Several sources of uncertainty have been identified for imaging biomarkers, consistently with the usual stages of any measurement process: acquisition, preparation, processing, measuring, biological variation and calibration. For instance, measurements taken during the preparation stage can be influenced by procedures related to noise reduction or spatial hyperesolution. In previous research, there are a number of studies aimed to identify and correct these sources of uncertainty and their consequences on the accuracy of data. To properly identify "uncertainty," we must define the components that constitute the origin of this concept. Taylor and Kuyatt [41] distinguished between those which can be assessed through statistical methods and those assessed by other means. This classification represents a breakpoint with regard to the traditional division suggested by the classical test theory into a "true" or underlying component which is unknown and some degree of an "error" component. The true value of the outcome variable is what we try to analyze, whereas the error term is related to the imprecision in the measure that frustrates our efforts of obtaining a pure true score. This error term may be due to randomness (called "noise"), inherent to any measurement process dealing with human behavior, or the presence of a systematic error or bias. Random errors may arise for a number of reasons, like changing conditions while taking measurements, inaccuracy of observers, and lack of technical equipment sensitivity or imprecise definitions of what has to be measured. Random errors do not include human or technical mistakes, such as reading a wrong value or disruption while performing a controlled experiment. On the contrary, systematic error or bias refers to ratings that depart systematically from true values. The presence of bias in a biomarker involves a shift in all measures, in such a way that the mean value of the outcome variable is constantly different from the true (and unknown) value in the population of study.

Although it is not possible to predict potential bias in clinical trials in advance, appraising bias control is a highly recommendable practice to avoid erroneous conclusions in clinical endpoints. In fact, except when the outcome is mortality, the size of the bias cannot be predicted from data [45]. Some studies have pointed out that an inadequate or inexistent randomization in published trial reports may be linked to more positive assessments of intervention effects. Similarly, a clear association has been found between studies funded by industrial organizations and pro-industry conclusions [19].

In the assessment of clinical endpoints, the design of the clinical trials should be blind and randomly allocated to intervention whenever possible, as these flaws may bias their results. Statistical methodologies for design of experiments may also help to prevent experiments from bias through orthogonal factorial experiments, instead of modifying one factor per experiment. From a psychological point of view, one should bear in mind that a considerable amount of cognitive work is done by patients before reporting about their health status. For instance, patients may be attempting to guess and agree with the research interests of their practitioners (acquiescence); they may avoid extreme values when filling health questionnaires using Likert scales (end aversion, social desirability), or even ticking responses one after another (straight line) [14]. Other background variables may affect the assessments made by patients, such as gender, age, or socioeconomic status. Even the psychological mood when providing self-assessments may have important effects on the response. For instance, it has been studied how moving out of hospital confinement alone generates a positive systematic bias in the degree of improvement perceived by patients. In contrast, patients with depressed mood reported a disproportionate negative perception of their health status [16].

More than thirty different variations of bias have been defined in medical imaging, which can be classified into two main categories: Sampling and Magnitude bias. Selection or sampling bias refers to the selection of patients or case studies. This source of bias is especially relevant when radiologists try to identify the impact of an imaging biomarker status on a particular clinical endpoint while adjusting for age, gender or other personal features. These strata defined by radiologists may be relevant for research, though not balanced or homogeneous in their sample, when compared with other potential predictors. In order to avoid this sampling bias, multivariate models should be used to simultaneously adjust for all predictors [16]. On the other hand, magnitude bias becomes crucial when radiologists try to identify a Region of Interest (ROI) with background noise using Magnetic Resonance Imaging (MRI). By means of this technology, a list of eigenvalues and eigenvectors, representing principal diffusivities and principal directions, are calculated in each voxel. Usually these eigenvalues are sorted according to their increasing magnitude, resulting in an undesired sorting bias. As a remedial measure, different statistical solutions have been proposed, though there are no available and generalized methods to quantify and assess the presence of this bias in the distribution of eigenvalues obtained with MRI [3]. However, radiologists are not expected to eliminate all sources of bias. In doing so, the consequence could be to limit the generalizability of results and the scope of their conclusions [41]. Studies with such methodological background would be absolutely unrealistic, and consequently useless for the potential readers of their scientific contributions.

The point of bias caused by psychological factors, not controlled by practitioners, is generally true for clinical endpoints. However, also in imaging biomarkers, similar drawbacks are often found. This is specially the case of reproducible clinical trials where measures are taken using different, but related, biomarkers (internal consistency), several persons of the staff take measures (observer variation), a lapse of time is necessary to perceive the effects of a treatment or an intervention (test–retest), or the effectiveness of a new method is tested against other existing clinical procedures (method comparison). The following section will describe how to deal with the vari-

ability associated to each of these sources of potential bias.

9.3 How Reliable Are the Measures Provided by Biomarkers?

Before proceeding to the definition of reliable biomarkers, let us first examine the difference between repeatable and reproducible clinical trials. Repeatable studies are appropriate for analyzing agreement among patients, as they refer to the variation in repeated measurements made on the same patients under identical conditions, for instance, using the same set of biomarkers. In contrast, reproducibility studies are adequate to examine to what extent the variation in measurements made on a patient under changing conditions may affect the reliability of a biomarker. These changing conditions include using different biomarkers (internal consistency), comparison across time (test–retest), observer variation, and method comparison [36].

In the previous section, we discussed the difference between the true underlying score and the error term that compose any measure provided by a biomarker. This point is especially important for the idea of reliable biomarkers, since reliability is defined as the ratio of the true score variance, with regard to the observed score variance, assumed to be the sum of the true and error components. This ratio provides a relative parameter that reaches unity when the total variance equals true variance and zero when measurement errors cope with all observed variance. Measurement errors cannot be avoided; neither should they be excluded from the analysis. Reliable biomarkers provide small measurement errors, so that differences between patients can be relatively well identified. On the contrary, when large measurement errors are obtained using a particular biomarker, reliability will be low. In these cases, true differences between patients will be difficult to distinguish, as they may be due to randomness, rather than a true difference in clinical outcomes. The general idea is that reliable biomarkers will be better clinical endpoints as they will provide values closer to the true underlying clinical outcome we try to examine.

Defining reliability as a unit-free ratio of variances constitutes an interesting aid to practitioners for assessing their biomarkers. Nunnally and Bernstein [34] suggested a generic cut-off in 0.70 to consider that a measurement instrument has an appropriate reliability, although suggested a minimum value of 0.90 when results will be used for decisions about individuals on the basis of his (her) score. The first lower limit fixed in 0.70 for reliable measurement instruments was confirmed by Streiner [38] for health measures in studies based on groups of patients in clinical trials.

9.3.1 Internal Consistency

As it will be described later, in order to ensure content validity of any measurement instrument, we need a consistent set of biomarkers that sample the entire domain and not include biased biomarkers that refer to other constructs. Since classical test theory assumes that this set of items is a random, but representative, sample of the population of all biomarkers available to represent the construct of the study, these biomarkers should be correlated highly with one another. In other words, we need the measurement instrument to have internal consistency.

The first method of measuring internal consistency was suggested by Cronbach [9], through his well-known coefficient α, which can be written as

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum \sigma_k^2}{\sigma_{\text{total}}^2}\right]$$

where k represents the number of biomarkers, $\sum \sigma_k^2$ is the sum of variances of all of the items, and σ_{total}^2 is the variance of the total score provided by the instrument.

This parameter has been widely used in research for several decades, although it has been often misused. As a result [12], some concerns about this reliability measure. First, it must be

remarked that this measure is a characteristic depending on the scores obtained in a particular sample of patients. It is not good practice to make statements about the reliability of an instrument under all circumstances, based exclusively on the value of Cronbach's alpha. Besides, the set of selected biomarkers must represent a unidimensional construct so that the analysis of the internal consistency makes sense. For that reason, this coefficient should not be used if it is suspected that biomarkers represent a multidimensional structure. Finally, the value of Cronbach's alpha depends on the number of biomarkers being studied. It appears that we may increase its value through the addition of new biomarkers to the instrument. Thus, this general reference of achieving a minimum value of 0.7 to ensure reliability may be distorted when large sets of biomarkers are examined. In fact, values of the Cronbach's alpha greater than 0.95 may point out to redundancy between biomarkers. The usual situation where we may find such high values for this measure is unidimensional scales with too many items correlating and overlapping among them. Consequently, the recommendable values for this measure should range among 0.7 and 0.95 to ensure internal consistency [39].

An alternative method for assessing internal consistency is examining inter-item correlations, based on the correlations of each item to the scale as a whole, omitting that item. Minimum correlations of 0.20 are expected to be obtained among biomarkers in order to conclude that they are internally consistent. For dichotomous items, Kuder and Richardson [30] suggested an alternative coefficient as the equivalent of Cronbach's alpha.

9.3.2 Observer Variation

Like internal consistency, observer variation constitutes an important source of variability in reproducibility studies. It may be caused by technical failures, misinterpretation of data, or skipping abnormalities while taking measurements for biomarkers. Measurements made by different observers usually present a high degree of variability due to the presence of bias between them. In contrast, measurements made by the same observer are more similar [3]. Meanwhile, the measurement errors associated to each observer can also have different standard deviations, if one of the observers is likely to obtain more precise measurements. The consequence in these circumstances is a higher intraclass correlation coefficient (ICC) within the same observer, when compared to the ICC corresponding to different observers.

The intraclass correlation coefficient (ICC) represents the correlation between one measurement of a biomarker (either a single rating or a mean of several ratings) and another measurement for that biomarker, but under different conditions [18, 37]. The ICC coefficient ranges from −1 to 1, like the Pearson correlation coefficient, and is defined as a ratio of true variance versus observed variance. However, the Pearson coefficient measures the similarity of the relative measures taken by two observers, whereas the ICC indicates the average similarity of the patients on both ratings.

The first step in calculating the ICC is to identify our research interests as regards to the role of the observers in the clinical trial. This coefficient refers to a family of analysis of variance, and there are three possible approaches for assessing observer variation accordingly (Cases 1, 2, and 3) [37]. In the first approach, we consider that each biomarker is measured by different observers (Case 1). Secondly, we can consider that a particular group of observers take measures for each biomarker (Case 2). Finally, following the third approach, all possible observers take measures in each biomarker (Case 3). These different approaches are specified in the mathematical formulation of the ICC coefficient.

If the study is designed according to Case 1, a one – way ANOVA is the most appropriate statistical model. In this analysis, total variability is decomposed into a Between – targets Mean Square (BMS) and a Within – targets Mean Square (WMS). On the contrary, if the study is designed as a Case 2 or 3 analysis, the within – target sum of squares is partitioned into a between – Judges Mean Square (JMS), referred to observ-

ers, and a Residual Mean Square (EMS). As mentioned, the only difference between both Cases is the assumption that observers are randomly sampled (Case 2) or fixed (Case 3).

For determining the effect of different observers in the variability of biomarkers, Case 1 is not interesting, as it does not allow the study of observer variation. If our interest lies in a particular selection of observers (Case 2), we should consider that biases between them are constant. Then, we assume that the effect of observers on the biomarker is fixed. Assessing differences between observers is essential for the interpretation of the results in most studies and will be an appropriate approach in most of studies. In these cases, allow for a random subject effect and fixed observed effects. Thus, a two-way mixed effects model will be adequate for obtaining the ICC.

Those who need to draw inferences about a wider population of observers should consider that biases between observers are constant. Perhaps the origin of these biases is inherent to the method itself, or the practitioner is interested in extrapolating his (or her) conclusions to other studies about the same clinical outcome. At that point, we consider that the observers in our study are a random sample of all potential observers from a larger population, and we assume that observers have random effects on the biomarker. In these cases, a two-way random effect model seems more appropriate, as random subject effects and random observer effects are both allowed [4]. However, the practitioner must beware the need for greater number of measurements (at least higher than two per observer) for analyzing the random effect of observer variation.

Differences between the measurements taken by different observers can be easily assessed for continuous data, as shown. In contrast, when data are categorical the question becomes more difficult to assess. Landis and Koch [31] suggested a set of tests for interobserver bias as generalized kappa-type statistics. However, Brennan and Silman [6] warned about the difficulties of dealing with data in categorical form. A more pragmatic approach based on raw data is recommendable, rather than simplistically calculating the χ^2 statistic.

9.3.3 Test–Retest Reliability

Test–retest reliability assesses whether a biomarker yields the same results on repeated measurements separated by a lapse of time, provided that the rest of the conditions of the clinical trial have not changed as regards the clinical endpoint being measured. This analysis allows the researcher to distinguish between a true change in the biomarker and another one occurring on the basis of chance or systematic bias. A generic interval between 2 and 14 days is the recommendable length of time that should elapse between both measurements [39]. This period of time is large enough for not being able to recall the biomarkers' previous measurements. However, it is not so long that actual changes in the clinical output may have occurred.

The use of Pearson correlations for test–retest reliability, or regression analysis, is now considered outdated as it can seriously exaggerate the impression of reliability [2]. It has been also argued that the statistical strength of the association between two measurements may not be due exclusively to agreement (Bland and Altman 1986). In fact, a change of units of measurement would not affect correlation, but it could have a significant effect in the assessment of agreement. Results from two repetitions of a measurement may correlate highly but be systematically different. For this reason, the intraclass correlation coefficient is usually a more appropriate indicator of test–retest reliability. Another alternative is to examine intrasubject variation in a graphical representation, as suggested by Bland and Altman [6], where differences in the measurements of a biomarker are plotted against the mean of both measurements.

9.3.4 Method Comparison

Whereas the ICC is commonly used for analyzing observer variation and test–retest reliability,

the Bland and Altman plot tends to be largely applied for comparing two methods of measurement of the same biomarker. This plot is used for examining whether two methods of measurement are so similar that one of them could accurately replace the other [5]. This graphic plots on the x-axis the mean of a patient's measurements using both methods $(A+B)/2$, against the difference between both methods $(A-B)$ on the y-axis. Notice that we should always subtract the same method's measurement from the other's $(A-B)$ in order to obtain consistent results.

This approach provides the bias and the limits of agreement for the bias. In simple terms, the bias is the overall mean difference between both methods of measurement, whereas the standard deviation of this bias is the estimate of the error. It is important to make a first inspection of the plot, in order to identify visible relationships between this bias across the x-axis and the difference of measurements on the y-axis. If the variability of the paired differences is uniform along the range of measurements, we may estimate the limits of agreement.

These limits of agreement are marked in the Bland and Altman plot using dashed lines and represent a range of values in which the "true" agreement will lie for approximately 95 % of the sample. In other words, the limits of agreement provide an interval within which 95 % of future differences between methods is expected to fall. The hypothesis of zero bias can be statistically examined by a paired t-test of the measurements from each method. However, the bias and limits of agreement must be additionally assessed from a qualitative perspective. If both fall within the lower and upper cut-offs defined by the practitioner as satisfactory, both methods may be used interchangeably. In contrast, if bias exceeds at least one cut-off, there may be an over- or underestimation of the true clinical endpoints measured through each method. Consequently, in these cases both methods cannot be considered identical and should not be used interchangeably [23, 27].

After checking the assumption of normal distribution for paired differences, we can compute the mean and standard deviation of the paired differences using the following expression:

$$\text{Mean difference} \pm 1.96\ \text{SD}(\text{differences})$$

If the paired differences are normally distributed, we may compute the standard error of the limits of agreement as

$$\frac{1.71\ \text{SD}(\text{differences})}{\sqrt{n}}$$

where n represents the number of subjects.

On the contrary, whenever a relationship between the bias and the differences in the measurements using different methods is found in the plot, a transformation (of the raw data) may be successfully employed. However, in those cases the limits of agreement will be asymmetric and the bias will not be constant. For instance, it is a common problem to find that the variability of the differences increases when the value being measured is larger. This problem can be solved taking the difference in the logarithm of the methods [4].

9.4 How Well Do Biomarkers Represent the Constructs Being Measured?

In the previous section, we discussed different approaches for examining reliability. However, a reliable biomarker does not involve its validity. The point is this: the reliability of a biomarker indicates the stability of measures under changing conditions, whereas validity is defined as the extent to which a biomarker measures what it intends to measure [29]. This point bears repeating because many biomarkers have been suggested to be valid indicators of some particular clinical outcomes, based on their values for the ICC or the Bland and Altman plot.

However, in the search for credibility, there is a tendency to accept almost any biomarker with high reliability coefficients, as an appropriate biomarker of a particular clinical outcome. Additional analysis of validity must be performed in those cases, as the description of the biomark-

ers needs to confirm the link between the clinical outcome and the biomarker. The following section will describe different types of validity, according to the approach of the analysis.

9.4.1 Content Validity

The content validity defines the degree to which a sample of items constitutes an adequate definition of the construct to be measured. In other words, it refers to how adequately the selected items (or the selected biomarkers) cover the topics that were specified in the scope definition of the instrument.

The content validity also includes face validity, which is particularly focused on the clinical credibility of a measure, thanks to its clarity and completeness with regard to the research area. There is a general consensus that content validity is largely a matter of expert judgment. Frequently, patients and experts are asked to critically review the content validity of an existing instrument, through formal focus groups, cognitive interviews, and, occasionally, tests of linguistic clarity [32]. Thus, a high number of composite indicators have been defined, in order to reflect the degree of agreement between experts. A widely used measure for reporting content validity is the content validity index (CVI) [33]. However, this index may be computed through two alternative methods, based on the universal agreement among experts, or an average between the item-level CVIs, sometimes leading to different conclusions [35]. Thus, statements about the content validity of an instrument should be based on exhaustive conceptualizations of constructs, well-written items, and carefully selected and trained experts regarding the underlying constructs and the rating task to be developed [12, 32].

9.4.2 Construct Validity

A construct is a theoretical concept, which is intended to be measured through a biomarker, or set of biomarkers. In the field of health and clinical sciences, the constructs of interest are often described as clinical outcomes. Generally, constructs are defined as latent variables, such as major depression or distress. These constructs are usually measured through several manifest variables such as patient outcomes, reported by the patient or observed by the clinicians.

Confirmatory factor analysis (CFA) provides empirical evidence of the convergent and discriminant validity of theoretical constructs, while adjusting for measurement error [7]. Convergent validity refers to the evidence that theoretically similar, or overlapping, constructs are highly correlated, while discriminant validity focuses on the lack of correlation between indicators of theoretically distinct constructs. The most elegant model for addressing the simultaneous study of convergent and discriminant analysis is the analysis of the multi-trait–multi-method (MTMM) matrices, as described by Campbell and Fiske [8].

9.4.3 Criterion Validity

Criterion validity examines to what extent scores provided by the instrument agree with a definitive "gold standard" reference of measurement in the same field of study. The COSMIN group arrived to the conclusion that no gold standard exists for health-related or patient-reported outcomes. Thus, as it cannot be assured that the comparison with other instrument is really gold, it may be more appropriate to focus on construct validity [33]. On the contrary, if the estimation of criterion validity may be supported by a gold standard reference, the type of data should be considered. If both instruments have continuous scores, correlational studies may offer interesting information. However, if one of the instruments has a dichotomous scale, the AUC provided by the ROC curve constitutes a more appropriate method [36].

9.5 How Accurately Can Biomarkers Detect Small Clinical Effects?

Responsiveness represents the ability of a biomarker to detect clinically important changes [20]. However, before conducting a responsive-

ness study, it is important to address the issues of the validity and reliability of the biomarker in question. Internal responsiveness characterizes the ability of a biomarker to detect change over a determined lapse of time, whereas external responsiveness refers to the extent to which a change in a biomarker is linked to a corresponding change in other biomarker, acting as a reference of the clinical output [25]. A biomarker may be both reliable and valid, but unresponsive to change. Conversely, they may show poor reliability and validity but excellent responsiveness. Responsiveness, reliability, and validity are necessary requirements for biomarkers designed primarily to measure change over time. Besides, an important distinction must be made between discriminative biomarkers, designed for distinguishing between patients, and evaluative instruments, within individual assessments over time [21]. For evaluative measures, it is crucial for any instrument to be able to detect changes that are actually occurring.

In order to determine whether an instrument is responsive, two key parameters must be defined. Firstly, the smallest detectable change (SDC) (or smallest detectable difference (SDD)) represents the minimal change that must be surpassed to ensure real change. Small effects usually require larger sample sizes for clinical studies to be detected. The size of the SDD depends on the ability of the biomarker to detect small differences, but also on the sensitivity of the statistical methods. However, choosing methods for data analysis with high statistical power will not always improve responsiveness of a biomarker, as this decision must be based on the observed or expected distribution of the data.

Secondly, the SDC provides a reference to quantify the minimal clinically important difference (MCID) as a valid and clinically relevant cut-off level that patients and their physicians consider clinically important. However, both measures cannot be used interchangeably, as there is a meaningful difference between them [43]. If the MCID is greater than the SDD, small effects could be detected by the biomarker, but it will make no sense for the patient, since they will not notice it. On the contrary, if the SDD is greater than the MCID, more statistical power will be required through the increase of the sample size so that smaller effects can be detected [1].

Directly related to this measure is the measurement error, quantified by the standard error of measurement (SEM). It is defined as the standard deviation of errors of measurement and is directly related to the error score variance [24]. Thus, it represents the amount of variance of repeated measures within a patient. A biomarker is clinically worthwhile when its measurement error does not surpass the cut-off value for the minimal clinically important difference.

Rather than reliability and validity, there is no consensus about how responsiveness should be quantified. This lack of agreement has led to a proliferation of statistics for representing responsiveness, which complicates the comparison of studies using different measures. In fact, employing an inadequate coefficient among all available measures for responsiveness can hide the real change in the biomarker, through its inclusion in the noise term with the variability of other non-controlled factors [17]. In fact, it is not surprising to find different conclusions from studies comparing similar biomarkers, depending on the measure chosen to assess responsiveness. Besides, researchers interested in evaluating the capacity of a biomarker to detect change must be careful not to select systematically patients who are generally expected to improve, rather than remain stable or even worsen.

9.5.1 Internal Responsiveness

The most commonly used methods for assessing internal responsiveness are based on the mean or general change in a group of patients in a concrete period of time.

Traditionally, the t-test for paired data tests the null hypothesis in such a way that there is no change in the mean score of a biomarker between two moments of measurement (for instance, before and after an intervention or a treatment). This statistic may be valid for comparative purposes in a single study, although it has a strong dependence on the size of the sample. Thus, this measure should be complemented by the use of effect size

statistics, based on the ratio between the magnitude of health status changes, and its variability:

$$t = \frac{\text{Mean change}}{\text{SD}_{\text{difference}} / \sqrt{n}}$$

Alternatively, responsiveness is often assessed through effect sizes, which represent the change in mean scores on a particular biomarker, divided by the standard deviation of the measure at baseline (called standardized effect size (SES)). This statistic provides direct information on the magnitude of the change, expressed as a variation. Thus, this SES is then expressed in relative units that allow comparing responsiveness between different biomarkers. A value of SES greater than 0.8 may be considered large, whereas moderate or medium effects range from 0.2 to 0.8, and effect sizes lower than 0.2 are considered small or minor [9]. When using this parametric measure, it should be considered that it does not make sense in nonnormally distributed data. In addition, we are unable to interpret whether it reflects a real change in the score of the biomarker or just the variability of the baseline score:

$$\text{SES} = \frac{\text{Mean change}}{\text{SD}_{\text{baseline measurement}}}$$

Another possibility for assessing responsiveness is the standardized response mean (SRM), which divides the mean change in the score of the biomarker by the standard deviation of the patient's changes in the scores, rather than baseline scores [28]. This measure is considered to be more informative than the effect size, since those biomarkers with high variability of patient's changes in scores will lead to low values of the SRM. Again, values above 0.8 are considered to be large, and values between 0.5 and 0.8 are moderate, and values lower than 0.2 are regarded as small:

$$\text{SRM} = \frac{\text{Mean change}}{\text{SD}_{\text{change}}}$$

The modified standardized response mean (MSRM) suggested by Guyatt et al. [21], also known as the *responsiveness statistic*, was built on the previous measure and is described by some as the superior responsiveness. This measure divides the minimal clinically important difference (MCID) by the error of the quadratic mean (MSE) obtained through a repeated measure analysis of variance for stable patients. When there are only two measures for each patient (i.e., before and after an intervention or a clinical treatment), the MSE is the standard deviation of change scores for stable patients. Nevertheless, this index is not so widely used as it was initially expected, due to the difficulty to define accurately the MCID for some particular biomarkers and patient populations [20]. Similarly, values of 0.2, 0.5, and 0.8, or greater, are usually referred to represent small, moderate, and large responsiveness, respectively. The use of the same cut-off for so different responsiveness measures points out to some confusion in the interpretation of these measures:

$$\text{MSRM} = \frac{\text{MCID}}{\sqrt{2\text{MSE}}}$$

All this measures based on effect sizes share a common limitation. The observed change in the measure may not be related to a relevant change in the health status of the patient. We cannot state that a change has happened in the clinical or health status of a person with just two measures (before and after a treatment or an intervention). Thus, it is not recommendable to rely solely on internal measures of responsiveness, as it could lead to erroneous conclusions about the biomarker in question. External clinical measures should be also provided as they provide an external criterion to assess changes at the individual patient level [25].

9.5.2 External Responsiveness

The use of receiver operating characteristic (ROC) curves was initially suggested by Deyo and Centor [13]. These curves examine the proportion of true changes detected through a biomarker with regard to a clinical endpoint

(sensitivity) and the proportion of stable patients according to the biomarker, who are truly not changing (specificity). The novelty of this measure of responsiveness is that some external criteria are now required, in order to determine the meaning of "true" change. This external standard may be based on the patient's or the clinician's judgment of change and constitutes a reference to measure the ability of the biomarker to detect change. The ROC curve provides a very useful plot of the false-positive rates $\left(1-specificity\right)$ versus the true positive rates $\left(sensitivity\right)$. The area under the curve (AUC) is calculated using both measures and represents the biomarker's responsiveness, in terms of the relationship between a biomarker and an external standard for change. The ROC curves present only two limitations. Firstly, the external clinical score must be dichotomized, losing valuable information on the magnitude of actual change in the clinical endpoint. Secondly, separate curves must be obtained in order to assess responsivity of an instrument capable to detect improvement and deterioration of the patient's health.

Otherwise, correlational analysis may appear as an interesting option to analyze whether a biomarker is responsive regarding an alternative clinical endpoint. However, this measure has an important dependence on the distribution of data. Nonlinearly related variables, or extreme values, may affect the interpretation of results obtained with this statistic. Husted et al. [25] suggested using a generalization of regression models, as they provide a natural solution to examine external responsiveness in biomarkers. Similarly, García de Yébenes Prous et al. [17] recommended to perform a repeated measures analysis of variance (ANOVA) with an intrapatient factor (measured in two moments at least) and a factor of change between patients (with two or more levels: improvement/no improvement).

Conclusions

The present chapter has tried to put together a variety of approaches to analyze reliability, validity, and responsiveness of biomarkers to be used in clinical trials as surrogate endpoints. The wide diversity in the available measures for reporting these properties should make the reader aware about the importance of a thorough search in advance to any data analysis. Selecting the most appropriate measure of reliability, or the most suitable approach to examine validity, depends not only on the quantitative or qualitative nature of the data but also on the statistical distributions that represent these data. The same measure of responsiveness may not fit two different biomarkers about the same clinical endpoint, if one of these biomarkers is nonnormally distributed. Thanks to this variety of statistical needs for data analysis, we may choose among a long list of measures for assessing each characteristic of biomarkers. A good statistical assessment of a biomarker does not require computing all available measures of reliability, validity, and responsiveness. Instead, the reader should be able to select the measure that fits more accurately to the statistical distribution of the data and contemplate different and complementary approaches to assess biomarkers.

Thus, the assessment of imaging biomarkers should consider the statistical distribution of data, as well as the presence of potential outliers or confounding effects. Whenever possible, every measure regarding biomarkers should be followed by a confidence interval, so that information about the uncertainty in the estimation of parameters can be assessed by the reader. However, the qualitative characteristics of the patients who participate in the clinical trial should also be contemplated. There are a number of studies in the existing literature, showing the properties of a biomarker (or set of biomarkers), while setting aside the particular characteristics of the patients from which is was obtained. Thus, it may seem that the biomarker will always present some particular degree of reliability, validity, or responsiveness, as reported by previous research. But the reader should appreciate that these levels were reached using a particular sample of patients. A new application of the biomarker for measuring the same clinical endpoint will provide different values for those properties.

References

1. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower ex. Arthritis Care Res. 2001;45(4):384–91.
2. Bartko JJ. The intra-class correlation coefficient as a measure of reliability. Psychol Rep. 1966;19(1):3–11.
3. Basser PJ, Pajevic S. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. Magnetic Resonance in Medicine. 2000;44(1): 41–50.
4. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol. 2008;31(4): 466–75.
5. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;8(1):307–310.
6. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60.
7. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ (Clin Res Ed). 1992;304(6840):1491–4.
8. Brown TA. Confirmatory factor analysis for applied research. New York: Guilford Publications; 2015.
9. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. American Psychological Association; 1959.
10. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale: Erlbaum; 1988.
11. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
12. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. Educ Psychol Meas. 2004;64:391–418.
13. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res. 1992;5(4):194–7.
14. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis. 1986;39(11):897–906.
15. Donovan JL, Frankel SJ, Eyles JD. Assessing the need for health status measures. J Epidemiol Community Health. 1993;47(2):158–62.
16. Ensor JE. Biomarker Validation: Common Data Analysis Concerns. The Oncologist. 2014;19(8):886–91.
17. Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating patient-based outcome measures for use in clinical trials: a review. Health Technol Assess. 1998;2(14):74.
18. García De Yébenes Prous MJ, Salvanés FR, Ortells LC. Responsiveness of outcome measures. Reumatol Clín (Engl Ed). 2008;4(6):240–7.
19. Doros G, Lew R. Design based on intra-class correlation coefficients. Am J Biostat. 2010;1(1):1–8.
20. Gluud LL. Bias in clinical intervention research. Am J Epidemiol. 2006;163(6):493–501.
21. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol. 1989;42(5):403–8.
22. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171–8.
23. Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making Int J Soc Med Decis Making. 2011;31(6):828–38.
24. Hanneman SK. Design, analysis and interpretation of method-comparison studies. AACN Adv Crit Care. 2008;19(2):223.
25. Harvill LM. Standard error of measurement. Educ Meas Issues Pract. 1991;10(2):33–41.
26. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000;53(5):459–68.
27. ISO. Evaluation of measurement data – guide to the expression of uncertainty in measurement. International Organization for Standardization Geneva ISBN. 2008;50(Sept):134.
28. Jensen AL, Kjelgaard-Hansen M. Method comparison in the clinical laboratory. Vet Clin Pathol. 2006;35(3):276–86.
29. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. Med Care. 1992;30:917–25.
30. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm Ser Res Fundam. 2008;65(23):2276–84.
31. Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika. 1937;2(3):151–60.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
33. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–5.
34. Martuza VR. Applying norm-referenced and criterion-referenced measurement in education. Boston: Allyn & Bacon; 1977, Incorporated.
35. Mcdowell I. Measuring health : a guide to rating scales. Stat Med. 1989;8. doi:10.1093/aje/155.10.899.
36. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10(1):22.
37. Nunnally JC, Bernstein IH. The theory of measurement error. Psychometric Theory. New York, NY: McGrawHill.

38. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489–97.
39. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236–40.
40. Shrout PE, Fleiss JL. Intra-class correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.
41. Sica GT. Bias in Research Studies. Radiology 2006;238(3):780–89.
42. Streiner DL. Construct validity of the relationship profile test : a self-report measure of dependency-detachment construct validity of the relationship profile test : a self-report measure of dependency – detachment. J Pers Assess. 2003;80(1):99–103.
43. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
44. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5(6):463.
45. Taylor BN, Kuyatt CE. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technology. 1994;1297:2.
46. Thomas EJ, Peterson L a. Medical errors and adverse events in health care. J Gen Intern Med. 2003;18:6.
47. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol. 2010;63(1):28–36.
48. WHO International Programme on chemical safety biomarkers in risk assessment: validity and validation. 2001. Retrieved from http://www.inchem.org/documents/ehc/ehc/ehc222.htm.
49. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Sterne JAC. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601–5.

Validating the Imaging Biomarker: The Proof of Efficacy and Effectiveness

10

George C. Manikis, Nickolas Papanikolaou, and Celso Matos

10.1 Introduction

Medical imaging is an active and developing area, providing significant anatomical, functional, and molecular information in a wide range of clinical and research studies. Medical imaging techniques like ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) in conjunction with advanced acquisition protocols have given rise to creating and establishing quantitative biomarkers to assess biological processes and clinical end points [1]. Imaging modalities such as CT, MRI, and PET and their associated imaging biomarkers contribute significantly to the design of oncology therapeutic trials [2]. In contrast to biomarkers obtained by gathering tissue sample from a patient (i.e., genes or proteins from molecular studies), imaging biomarkers have the advantage of being noninvasive, repeatable over time, and relatively comfortable for patients (Table 10.1).

Particularly in MRI, a significant breakthrough has been achieved in the last decades where significant imaging biomarkers have been developed for functional information assessment. Diffusion-weighted imaging (DWI) and diffusion tensor imaging (DTI) give an insight in the complexity of the tissue environment via the motion of water molecules. Dynamic contrast-enhanced (DCE) and dynamic susceptibility contrast (DSC) imaging techniques, acquired after intravenous administration of specific contrast agents, depict tissue perfusion and the microvascular environment. Magnetic resonance spectroscopy (MRS) is conducted for measuring the concentration of several biochemical metabolites in the tissue, reflecting unique information on its chemical composition.

10.2 Road for Biomarker Development

A significant number of imaging biomarkers are presented in the literature as potential indicators of biological processes and for monitoring the response to therapy. However few of these are used routinely in the daily clinical practice failing to replace the established histological "gold standards" [9]. This issue relies on the fact that integrating imaging biomarkers into the clinical practice requires a stepwise well-structured procedure in which a couple of criteria must be first followed. As reported in [10], these criteria underline the importance of a biomarker to (a) provide accurate, precise, and feasible measurements, (b) be associated with a clinical end point, and (c) perform in a specific context of its

G.C. Manikis, MSc
Computational BioMedicine Laboratory (CBML), Institute of Computer Science (ICS) Foundation for Research and Technology – Hellas, Nikolaou Plastira 100, Heraklion, Greece

N. Papanikolaou, PhD • C. Matos, MD, PhD (✉)
Champalimaud Center, Av. Brasilia, Doca de Pedrouços, 1400-038, Lisbon, Portugal
e-mail: celso.matos@fundacaochampalimaud.pt

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_10

Table 10.1 Repeatability assessment of the imaging biomarkers calculated from different MRI modalities, organs and diseases

Imaging biomarker	Organ, disease	Metric	Reference
DWI: ADC, nADC	Hepatic metastases	CV_{ADC}: 10.1 % CV_{nADC}: 8.3 %	Deckers et al. [3]
DCE-CT: Arterial flow, blood volume, permeability	Gastroesophageal junction cancer	ICC_{AF}: 0.88 ICC_{BV}: 0.89 ICC_{PERM}: 0.91	Lundsgaard Hansen et al. [4]
DWI: ADC, D, D*, f	Hepatic metastases	CV_{ADC}: −14.7 to 13.8 % $CV_{ADChigh}$: −15.5 to 9.46 % CV_{f}: −75.3 to 241 % CV_{D*}: −89 to 2120 % CV_{D}: −20.8 to 25.3 %	Andreou et al. [5]
DCE-MRI: Ktrans Kep Ve Vp	Renal cell carcinoma	ICC_{Ktrans}: 0.686 ICC_{Kep}: 0.906 ICC_{Ve}: 0.764 ICC_{Vp}: 0.657	Wang et al. [6]
T1 relaxometry	Liver	$CV_{T1liver}$: 0.9–2.5 %	Aronhime et al. [7]
DWI: ADC, f, D, D*, fD*, DDCa, a, DDCk, kurtosis	Pediatric tumors	CV_{ADC}: 3.3 % CV_{f}: 41 % CV_{D}: 2.5 % CV_{D*}: 35.1 % CV_{fD*}: 38.1 % CV_{DDCa}: 4.3 % CV_{a}: 3.5 % CV_{DDCk}: 6.1 % CV_{K}: 52.7 %	Jerome et al. [8]

proposed use. In the context of developing biomarkers in general, the Institute of Medicine broadly classifies these criteria into three distinct but interrelated areas: the analytical validation, qualification, and utilization. Qualification and utilization are major parts for assessing the clinical impact of a biomarker, whereas analytical validation plays the most important role in determining the technical performance of it.

Analytical validation is a prerequisite step before qualification and utilization as it provides the concepts and methods for evaluating the validity and the performance characteristics of a biomarker from the technical perspective. The analytical validation process is not limited to accuracy metrics such as the sensitivity and specificity but the measurement of efficacy and effectiveness. Once technical performance is ensured for a potential biomarker, clinical questions need to be answered. Sensitivity and specificity of a potential biomarker in a human population and its cutoff points for establishing clinical decision-making systems and many other areas that are out of the technical scope of this chapter are major concerns that need to be addressed. Thorough studies for qualification and utilization are presented in [10–12].

The subsequent chapters will give an overview of the analytical validation criteria and address the computational models and measurement systems required for assessing the ability of a biomarker to meet the validation criteria. When met, established proofs of efficacy and effectiveness of imaging biomarkers will facilitate the use of these in drug development, therapy assessment, and patient care.

10.3 Proof of Efficacy and Effectiveness Through Analytical Validation

On the one hand, the proof of efficacy enrolls the necessity to assess the ability of a biomarker in measuring tissue characteristics from an acquired image reproducibly, reliably, and accurately. On the other hand, the proof of effectiveness is

mainly involved in analyzing the ability of a biomarker to be clinically relevant, thus to be consistent as a potential indicator for clinicians, physicians, and drug developers to make proper and informed decisions. The Quantitative Imaging Biomarkers Alliance (QIBA) [13], organized by the Radiological Society of North America (RSNA), devotes its attention to the development of technical performance analysis methods and metrics in response to the aforementioned concerns for efficacy and effectiveness. As reported in [14] and [15], this technical infrastructure, provided by the QIBA Metrology Working Group, was grouped into three primary validation fields: measurement bias and linearity, reproducibility, and repeatability.

10.3.1 Standardization: A Prerequisite Before Validation

A major challenge to the biomarker validation is the lack of standard approaches to data gathering, analysis, and representation. Technical issues in standardization impair consistency, accuracy, repeatability, and reproducibility of a candidate biomarker, thus making the validation a difficult process. Two main categories related to standardization are highlighted in [12]: (a) the standardization of image acquisition and (b) the standardization of image analysis. The first category reflects the large variety of ways in which images of the same sequence are acquired. In case of DWI-MRI, optimization of the b-value distribution demonstrated improved accuracy and repeatability of the derived parameters from four mathematical models [16]. The second category refers to the mathematical framework in which the images are analyzed, visualized, and presented quantitatively. A representative study in [17] highlights the substantial influence of a region of interest (ROI) size and position into the quantification of the perfusion characteristics of DCE-MRI data.

To overcome these issues, standardized acquisition protocols and image analysis framework need to be addressed, strengthening reproducibility across different conditions. In case of DWI-MRI data, a thorough study in [18] underlines the standards required for stakeholders to conduct multicenter studies, evolving diffusion imaging biomarkers to be clinically useful with significant impact on drug development and patient treatment. Several consensus studies have been held on the standardization of perfusion MRI [19], PET, and CT [20].

10.3.2 Bias and Linearity

The ultimate goal of a potential imaging biomarker is to provide unbiased measurements in order to gain insight into the pathophysiology of a patient and to contribute significantly in the design of therapeutic trials. Accuracy and bias are often used synonymously [21] and in general describe the difference between a measurement and the true value of the same examined object. To this understanding, bias of an imaging biomarker can be evaluated either at cross-sectional studies in which the biomarker is measured at one time point or during a longitudinal study where changes are measured over multiple time points.

In both studies, inherent in the estimation of bias is to know the true value of the examined object. In case of in vivo imaging studies applied to patients, determining bias implies a challenging process because tissue characteristics and structural changes due to a disease (i.e., necrosis, fibrosis, cellularity, vasculature network architecture, etc.) are linked indirectly to their related imaging biomarkers [22] and gold standards from histopathology are necessary. Alternatively, validated tissue-mimicking materials (i.e., phantoms) can be well-defined references replacing real data for assessing bias [23].

Bias measurement typically begins with the qualitative representation of the measured value(s) against the referenced, otherwise the true or expected value, in a single plot. This visualization can be enhanced with confidence bounds that reflect the variability of the multiple measures and additional boxplots with outliers plotted. Quantitative assessment of bias includes metrics for estimating the squared difference between a measurement and the true value and

thresholds based on agreement intervals varying within the 95 % of the total differences.

Linearity is a crucial indicator for estimating the change between the measurements and the true values over the range of true values. In other words, linearity can be measured in longitudinal studies and when combined with analysis of bias can potentially provide quantitative information which is directly proportional to the true value change. A single scatterplot depicting a pairwise analysis of measured values of a biomarker versus its true values will have a slope equal to one and intercept at zero when there is no bias and perfect linearity over the range of true values. With known linearity degree of an imaging biomarker in a longitudinal study, measurement interval limits are determined, ensuring that its measurements reliably indicate clinically important true changes. Simple regression analysis can be also followed to describe the statistical relationship between the multiple measurements of a biomarker and its true values over the range of them.

10.3.3 Reproducibility

Reproducibility refers to test conditions, assessing the same imaging biomarker at short intervals, when studies are conducted using different experimental conditions. These experimental conditions include multiple measurements of the same imaging biomarker but with different vendors, measuring systems, operators, and locations that may compromise the reproducibility of the results. A reproducibility study is said to be valid when at least one of the above criteria is met, and a potential biomarker is robust enough and reproducible when repeated measurements with no variation of the same subject are provided under diverse conditions.

Measurements in reproducibility studies can be derived from both synthetic and real data. Synthetic data can be retrieved from phantoms that mimic the tissue characteristics [24, 25], whereas real data can be provided by single lesion of a patient or a group of patients with similar characteristics (i.e., individuals with same disease) [26–29]. However, patient's comfort and safety are crucial aspects that limit the ability to perform multiple repeated studies. Radiation exposure in CT and PET and the use of intravenous injection of contrast agents in DCE-MRI and DSC-MRI are typical features that limit the number of scans in a reproducibility test. On the other hand, phantoms often fail to illustrate the complexity and characteristics of the human tissue, thus leading to overestimated measurements for reproducibility [15].

A statistical analysis framework is required for assessing reproducibility qualitatively and quantitatively. Once a reproducibility experiment has been performed, useful statistics act as indicators of the variability between the different measurements. A technical representation of the reproducibility metrics is analytically outlined in [30]. Based on [30], reproducibility is often measured by two statistical metrics: the concordance correlation coefficient (CCC) and the reproducibility coefficient (RDC). Pairwise qualitative assessment is given by scatterplots and Bland-Altman plots. Moreover, distribution analysis based on histograms, Q-Q plots, and boxplots is an important tool to visually inspect the agreement between the biomarker estimates delivered from different conditions.

10.3.4 Repeatability

The potential ability of an imaging biomarker to act as an indicator of a biological process and for monitoring the response to therapy can be severely influenced by the lack of reproducibility and repeatability. These two terms are commonly confused, with varying degrees of consistency on their terms found in the literature [14]. Repeatability tests, often named as test-retest studies (Fig. 10.1), refer to the variation in repeat measurements of the same imaging biomarker under identical conditions. In contradiction to the experimental conditions in a reproducibility test, measurements are taking place under the same vendors, measuring systems, operators, and locations. A crucial prerequisite in repeatability tests is the short period of time required for

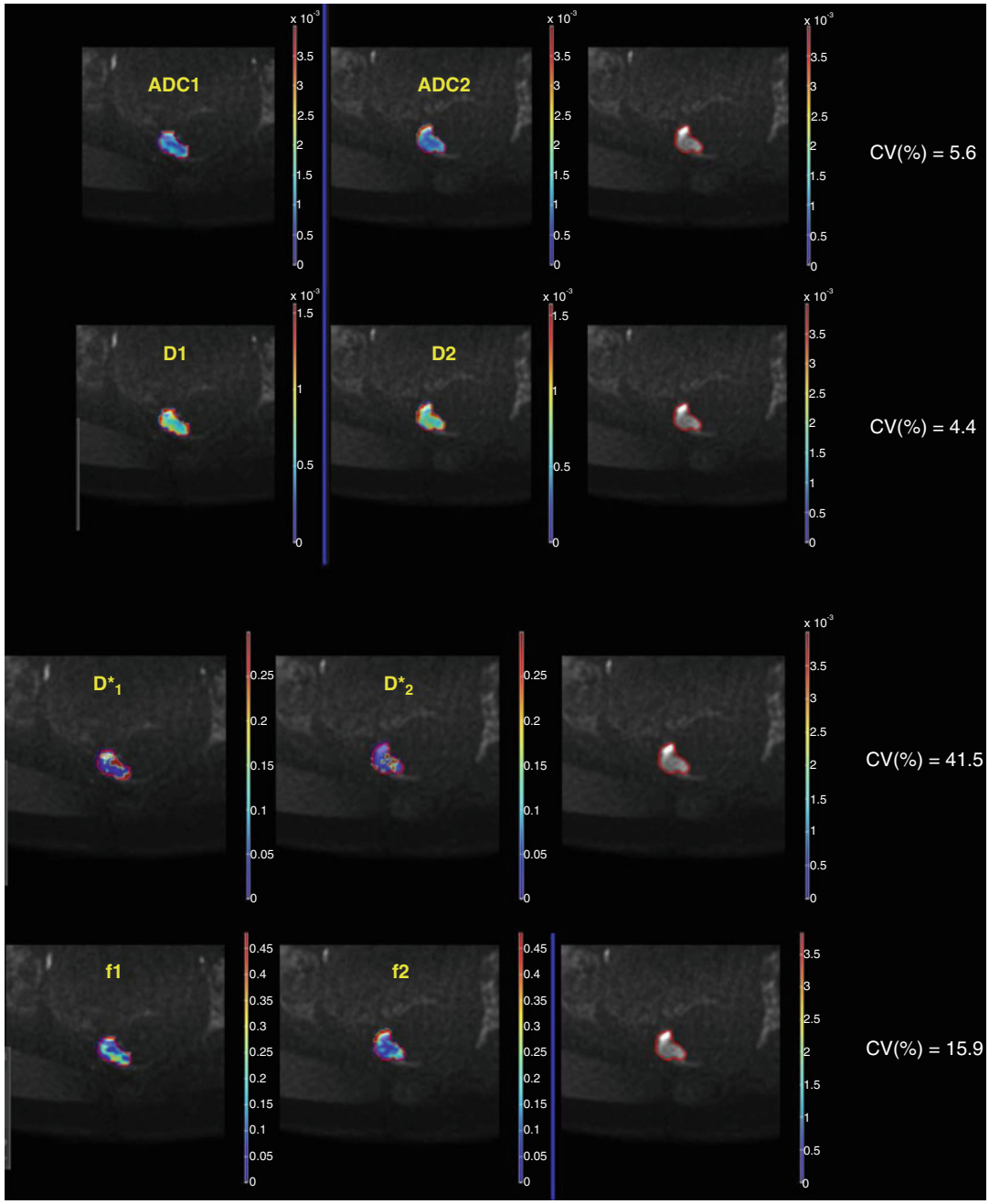


Fig. 10.1 Test-retest study in a patient with rectal cancer. Two identical scans were performed within the same imaging session to assess short-term repeatability. Variation coefficients (CV%) are sown for various DWI biomarkers including ADC, D, D*, and f. Post-processing was done with the DMT software (ICS, FORTH, Heraklion, Greece) [31]

measurements to be taken, in order to ensure that changes in the biomarker are not caused by inherent technical or physiological variation.

Repeat image sets for repeatability evaluation studies are relatively rarely obtained. Recently, efforts have been made into generating data for assessing repeatability of applied segmentation techniques for brain volume measurements [32]. Efforts for addressing the need of publicly available data for repeatability tests are also made by the National Cancer Institute's Cancer Imaging Program and its project RIDER [33].

Repeatability tests are also sensitive to standardization. A number of repeatability studies have been published describing the technical aspects and performance of imaging procedures in both phantoms and patients using different acquisition protocol and modalities. Studies in [34] and [35] showed that DSC-MRI perfusion-related parameters and especially the cerebral blood volume (CBV), when standardized protocols are followed, can be significantly repeatable in patients with glioblastoma multiforme. Another study in [36] reported that repeated hypoxia PET scans with [^{18}F]HX4 provide spatially stable results in two different group of patients: patients with head and neck cancer and patients with lung cancer. Recently, a correlation and repeatability study in [37] tried to assess the repeatability power and correlation degree between parameters from DCE-MRI and intravoxel incoherent motion (IVIM) DWI-MRI. In another study by Intven M et al. [38], repeatability of ADC was calculated in order to distinguish treatment-induced changes from measurement variations in patients undergoing chemoradiation therapy for rectal cancer, when comparing pre- and posttreatment ADC values. The mean tumor ADC value was 1.15×10^{-3} mm^2/s (SD 0.07×10^{-3} mm^2/s), while the repeatability coefficient of the ADC value was 9.8 %. Metens et al. [39] showed that a significantly better repeatability of liver and spleen diffusion coefficients and lower intersubject ADC variability can be obtained when using respiratory- and cardiac-triggered diffusion sequences. Aliu et al. [40] evaluated the variability and repeatability of repeated magnetic resonance imaging (MRI) measurements in normal breast tissues between and within subjects. The CV between subjects of fibroglandular tissue density and enhancement at visit 1 and visit 2 ranged from 47 to 63 %. The CV within subjects was 13 % for FGT density, 22 % for FGT enhancement, and 11 % for ADC.

From the technical point of view, the influence of bias between repeat measurements is assumed to be neglected. Therefore, in a repeatability study, the degree of agreement between the measurements depends only on the standard deviation within the examined subject. Standard repeatability statistical metrics include the widely used intraclass correlation coefficient (ICC), the repeatability coefficient (RC), and the within-subject variance (σ^2_w) and coefficient of variation (WCV), respectively. A detailed technical representation of these metrics can be found in [30].

References

1. Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Subramaniam RM, Lenchik L. Clinical utility of quantitative imaging. Acad Radiol. 2015;22(1):33–49.
2. Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, Buatti JM, Fennessy FM, Eikman EA, Kumar V, Forster KM, Wahl RL, Lieberman FS. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging. 2012;30(9):1301–12.
3. Deckers F, De Foer B, Van Mieghem F, et al. Apparent diffusion coefficient measurements as early predictive markers of response to chemotherapy in hepatic metastasis: a preliminary investigation of reproducibility and diagnostic value. J Magn Reson Imaging. 2014;40:448–56.
4. Lundsgaard Hansen M, Fallentin E, Axelsen T, Lauridsen C, Norling R, Svendsen LB, Nielsen MB. Interobserver and intraobserver reproducibility with Volume Dynamic Contrast Enhanced Computed Tomography (DCE-CT) in gastroesophageal junction cancer. Diagnostics (Basel). 2016;6(1).6(1). pii: E8. doi:10.3390/diagnostics6010008.
5. Andreou A, Koh DM, Collins DJ, Blackledge M, Wallace T, Leach MO, Orton MR. Measurement reproducibility of perfusion fraction and pseudodiffusion coefficient derived by intravoxel incoherent motion diffusion-weighted MR imaging in normal liver and metastases. Eur Radiol. 2013;23(2):428–34.
6. Wang H, Su Z, Ye H, Xu X, Sun Z, Li L, Duan F, Song Y, Lambrou T, Ma L. Reproducibility of dynamic

contrast-enhanced MRI in renal cell carcinoma: a prospective analysis on intra- and interobserver and scan-rescan performance of pharmacokinetic parameters. Medicine (Baltimore). 2015;94(37):e1529.
7. Aronhime S, Calcagno C, Jajamovich GH, Dyvorne HA, Robson P, Dieterich D, Fiel MI, Martel-Laferriere V, Chatterji M, Rusinek H, Taouli B. DCE-MRI of the liver: effect of linear and nonlinear conversions on hepatic perfusion quantification and reproducibility. J Magn Reson Imaging. 2014;40(1):90–8.
8. Jerome NP, Miyazaki K, Collins DJ, Orton MR, d'Arcy JA, Wallace T, Moreno L, Pearson AD, Marshall LV, Carceller F, Leach MO, Zacharoulis S, Koh DM. Repeatability of derived parameters from histograms following non-Gaussian diffusion modelling of diffusion-weighted imaging in a paediatric oncological cohort. Eur Radiol. 2016. doi:10.1007/s00330-016-4318-2.
9. Poste G. Bring on the biomarkers. Nature. 2011;469(7329):156–7.
10. Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Lenchik L, Subramaniam RM. Methods and challenges in quantitative imaging biomarker development. Acad Radiol. 2015;22(1):25–32.
11. Waterton JC, Pylkkanen L. Qualification of imaging biomarkers for oncology drug development. Eur J Cancer. 2012;48(4):409–15.
12. (ESR), European Society of Radiology. ESR statement on the stepwise development of imaging biomarkers. Insights Imaging. 2013;4(2):147–52.
13. Quantitative Imaging Biomarkers Alliance. http://rsna.org/QIBA.
14. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, Guimaraes AR, Filice R, Zhang Z, Sullivan DC, QIBA Terminology Working Group. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res. 2015;24(1):9–26.
15. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, Kondratovich M, McShane LM, Reeves AP, Barboriak DP, Guimaraes AR, Wahl RL, RSNA-QIBA Metrology Working Group. Metrology standards for quantitative imaging biomarkers. Radiology. 2015;277(3):813–25.
16. Merisaari H, Jambor I. Optimization of b-value distribution for four mathematical models of prostate cancer diffusion-weighted imaging using b values up to 2000 s/mm(2): simulation and repeatability study. Magn Reson Med. 2015;73(5):1954–69.
17. Braunagel M, Radler E, Ingrisch M, Staehler M, Schmid-Tannwald C, Rist C, Nikolaou K, Reiser MF, Notohamiprodjo M. Dynamic contrast-enhanced magnetic resonance imaging measurements in renal cell carcinoma: effect of region of interest size and positioning on interobserver and intraobserver variability. Invest Radiol. 2015;50(1):57–66.
18. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJ, Taouli B, Choyke PL. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia. 2009;11(2):102–25.
19. Essig M, Nguyen TB, Shiroishi MS, Saake M, Provenzale JM, Enterline DS, Anzalone N, Dörfler A, Rovira À, Wintermark M, Law M. Perfusion MRI: the five most frequently asked clinical questions. AJR Am J Roentgenol. 2013;201(3):W495–510.
20. Antunovic L, Rodari M, Rossi P, Chiti A. Standardization and quantification in PET/CT imaging: tracers beyond FDG. PET Clin. 2014;9(3):259–66.
21. Frey EC, Humm JL, Ljungberg M. Accuracy and precision of radioactivity quantification in nuclear medicine images. Semin Nucl Med. 2012;42(3):208–18.
22. Le Bihan D, Iima M. Diffusion magnetic resonance imaging: what water tells Us about biological tissues. PLoS Biol. 2015;13(7):e1002203.
23. Obuchowski NA, Barnhart HX, Buckler AJ, Pennello G, Wang XF, Kalpathy-Cramer J, Kim HJ, Reeves AP, Case Example Working Group. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. Stat Methods Med Res. 2015;24(1):107–40.
24. Malyarenko D, Galbán CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, Ross BD, Chenevert TL. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging. 2013;37(5):1238–46.
25. Li Q, Gavrielides MA, Sahiner B, Myers KJ, Zeng R, Petrick N. Statistical analysis of lung nodule volume measurements with CT in a large-scale phantom study. Med Phys. 2015;42(7):3932–47.
26. Miyazaki K, Jerome NP, Collins DJ, Orton MR, d'Arcy JA, Wallace T, Moreno L, Pearson AD, Marshall LV, Carceller F, Leach MO, Zacharoulis S, Koh DM. Demonstration of the reproducibility of free-breathing diffusion-weighted MRI and dynamic contrast enhanced MRI in children with solid tumours: a pilot study. Eur Radiol. 2015;25(9):2641–50.
27. Padhani AR, Khan AA. Diffusion-weighted (DW) and dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) for monitoring anticancer therapy. Target Oncol. 2010;5(1):39–52.
28. Zhang X, Pagel MD, Baker AF, Gillies RJ. Reproducibility of magnetic resonance perfusion imaging. PLoS One. 2014;9(2):e89797.
29. Griffanti L, Rolinski M, Szewczyk-Krolikowski K, Menke RA, Filippini N, Zamboni G, Jenkinson M, Hu MT, Mackay CE. "Challenges in the reproducibility of clinical studies with resting state fMRI: an example in early Parkinson's disease". Neuroimage. 2016;124(Pt A):704–13.
30. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M, Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC, QIBA Technical Performance Working

Group. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27–67.
31. Manikis GC, Nikiforaki K, Papanikolaou N, Marias K. Diffusion Modelling Tool (DMT) for the analysis of Diffusion Weighted Imaging (DWI) Magnetic Resonance Imaging (MRI) data. CGI'16, Proceedings of the 33rd Computer Graphics International. 2016; 97–100.
32. Maclaren J, Han Z, Vos SB, Fischbein N, Bammer R. Reliability of brain volume measurements: a test-retest dataset. Sci Data. 2014;1:140037.
33. Armato SG, Meyer CR, Mcnitt-Gray MF, McLennan G, Reeves AP, Croft BY, Clarke LP, RIDER Research Group. The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. Clin Pharmacol Ther. 2008;84(4):448–56.
34. Jafari-Khouzani K, Emblem KE, Kalpathy-Cramer J, Bjørnerud A, Vangel MG, Gerstner ER, Schmainda KM, Paynabar K, Wu O, Wen PY, Batchelor T, Rosen B, Stufflebeam SM. Repeatability of cerebral perfusion using dynamic susceptibility contrast MRI in glioblastoma patients. Transl Oncol. 2015;8(3):137–46.
35. Prah MA, Stufflebeam SM, Paulson ES, Kalpathy-Cramer J, Gerstner ER, Batchelor TT, Barboriak DP, Rosen BR, Schmainda KM. Repeatability of standardized and normalized relative CBV in patients with newly diagnosed glioblastoma. AJNR Am J Neuroradiol. 2015;36(9):1654–61.
36. Zegers CM, van Elmpt W, Szardenings K, Kolb H, Waxman A, Subramaniam RM, Moon DH, Brunetti JC, Srinivas SM, Lambin P, Chien D. Repeatability of hypoxia PET imaging using [^{18}F]HX4 in lung and head and neck cancer patients: a prospective multicenter trial. Eur J Nucl Med Mol Imaging. 2015;42(12):1840–9.
37. Hectors SJ, Wagner M, Besa C, Bane O, Dyvorne HA, Fiel MI, Zhu H, Donovan M, Taouli B. "Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI?". J Magn Reson Imaging. 2016. doi:10.1002/jmri.25194.
38. Intven M, Reerink O, Philippens ME. Repeatability of diffusion-weighted imaging in rectal cancer. J Magn Reson Imaging. 2014;40(1):146–50.
39. Metens T, Absil J, Denolin V, Bali MA, Matos C. Liver apparent diffusion coefficient repeatability with individually predetermined optimal cardiac timing and artifact elimination by signal filtering. J Magn Reson Imaging. 2016;43(5):1100–10.
40. Aliu SO, Jones EF, Azziz A, Kornak J, Wilmes LJ, Newitt DC, Suzuki SA, Klifa C, Gibbs J, Proctor EC, Joe BN, Hylton NM. Repeatability of quantitative MRI measurements in normal breast tissue. Transl Oncol. 2014;7(1):130–7.

The Final Step: Imaging Biomarkers in Structured Reports

11

Enrique Ruiz-Martinez,
Jose Damian Segrelles Quilis, Fabio García Castro,
Luis Martí-Bonmatí, and Ángel Alberich-Bayarri

11.1 Introduction

William Morton wrote the first radiology report in 1896. It was written in narrative mode and with no consensual content (Fig. 11.1). In 1923 Charles D. Enfield emphasized that the radiologist provided detailed descriptions, without expressing his opinion on the clinical relevance of such descriptions, which could benefit the patient [1]. However, the methodology for reporting medical studies has not changed with the same speed as technology, work environment or medical discoveries. Imaging studies are often an integral part of a patient's evaluation, and the associated radiology report serves as the primary method of communication between healthcare providers and, increasingly, patients.

E. Ruiz-Martinez (✉) • L. Martí-Bonmatí
Á. Alberich-Bayarri
La Fe Polytechnics and University Hospital, Valencia, Spain
e-mail: ruiz_enrmar@gva.es; marti_lui@gva.es; alberich_ang@gva.es

J.D.S. Quilis
Universitat Politècnica de València, Valencia, Spain
e-mail: dquilis@dsic.upv.es

F.G. Castro
QUIBIM (Quantitative Imaging Biomarkers in Medicine), Valencia, Spain
e-mail: fabiogarcia@quibim.com

The report is a document in which the imaging modality used, the radiological findings and the diagnosis should be compulsorily incorporated. Moreover, the findings must be written in a concise but complete way and should answer the questions that originated the exploration. The methodology and style of writing the radiology report relies on the professional experience of the physician and his/her abilities in taking findings and correlating them with other clinical information, e.g. laboratory data or anatomical pathology data. This means that this methodology requires writing resources besides technical knowledge, as the traditional report, i.e. in free text, does not present any structure. This form of reporting is learned through the period of medical residencies.

Most hospitals decide to divide the report in sections: patient data, characteristics of examination and indications for the exploration, comparison with other studies (optional), findings and conclusions. In this sense, the Radiology Society of North America (RSNA) established an initial consensus regarding the content of the report (Table 11.1) [17].

However, this way of reporting is subjective and in some cases does not answer the clinical question or does not have an impact in improving patient care. Furthermore, there are many discrepancies between reports in clinical routine, even with the same professional and the same diagnosed disease. Jeffrey Sobel analysed 822

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_11

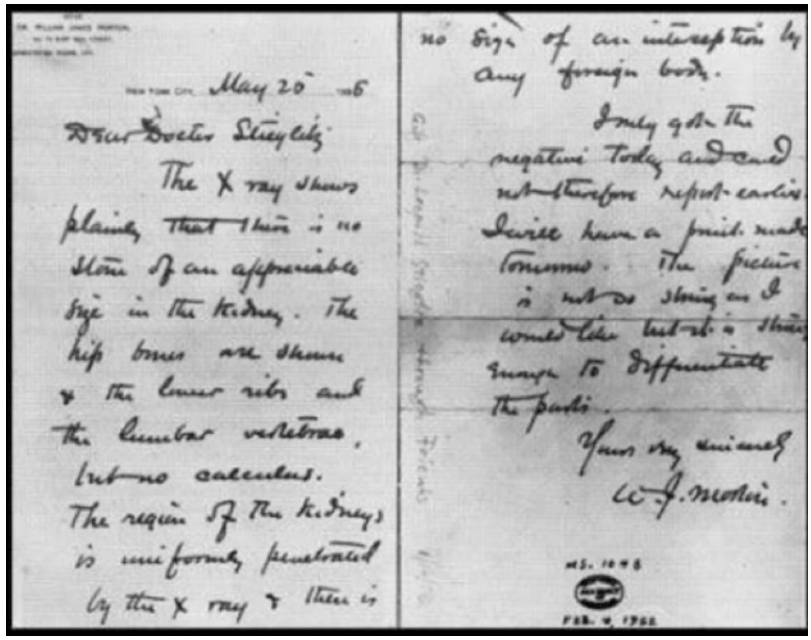


Fig. 11.1 The first recorded radiology report in 1896 by William Morton, MD

reports in 1996 and found that the radiologist used 14 terms for describing interstitial oedema/infiltration and 23 terms for the presence of an abnormality [2]. With the objective of solving this critical situation, Armas R.R. proposed in 1998 that an effective report should not contain abbreviations or neologisms [3]. The main drawback of the traditional way of reporting is that the radiologist is prone to fall in a stream-of-consciousness writing. That is, the physician makes a customization of the organization and content of the report for each case. This inherent variability in unstructured radiology reporting generates reports with different degrees of completeness and effectiveness. Armas enumerated the main properties of the report: clear, correct, concise, complete, consistent and confidence focused. Lafortune M. marked the steps to transform the radiology report in a clear and structured document in 1998 [4]:

- The report should be useful to the requesting doctor and to the patient.
- The report must answer the clinical reason for the benefit of the patient.
- The text must be readable, comprehensible, brief and concise.
- The report should avoid unnecessary long sentences and prolific language.
- The report should consistently focus on important features of the case.

Since 1996 the scientific and radiological community has made an effort to structure the radiology report. Structuring the report may lead to a quicker diagnosis, improving the communication between radiologists and between radiologists and clinicians, increasing report completeness and effectiveness, reducing costs, and raising the satisfaction levels of the clinicians, and most importantly, the report will consistently focus on important features of the study case [5]. Studies in the last decade show that the radiologist and the clinician prefer structured reporting systems [6–9].

There are many initiatives to promote structured reporting, among which the project of the European Society of Radiology (ESR) and the RSNA (www.radreport.org) stands out. These institutions created a library with more than 200

Table 11.1 Components of the radiology report

Section	Contents
Administrative information	Imaging facility
	Referring provider
	Date of exploration
	Time of exploration
Patient identification	Name
	Identifier (e.g. social security number)
	Birthdate
	Gender
Clinical data	Medical history
	Risk factors
	Allergies
	Reason for exam, including clinical need
Imaging modality	Time of image acquisition
	Image equipment
	Image acquisition parameters
	Contrast materials and other drugs administered
	Radiation dose (depends on modality)
Summary or impression	Key observations
	Inferences
	Conclusions, including any recommendations
Signature	The date and time of electronic signature for each responsible provider, including attestation statement for physicians supervising trainees, if applicable

report templates in English and approximately 50 templates translated into other languages. The templates try to serve as samples of "best practice" to lead radiologists through the process of report generation [17]. Moreover, each template includes metadata about the author, title, subject, brief description and date. The ESR and RSNA mapped the information in templates in standardized biomedical ontologies. The best example to illustrate structured reports is the Breast Imaging Reporting and Data System (BI-RADS), as the Food and Drug Administration of the United States (FDA) requires this system to be used for all mammography reports. This report has helped reduce variability in diagnostics and improved transparency in the communication between radiologists and clinicians [19].

Modern speech recognition software has popularized structured radiology thanks to its automation features [10–12]. Several speech recognition software packages allow creating fields that can contain text by default and/or that can contain a choice of possibilities for the radiologist to select [10]. This technology provides the chance to implement structured reports in radiology departments [12–15]. However, nowadays, there are few radiology departments where the structured report is the standard in clinical routine. In the words of Reiner, "adoption to date has been tepid" [16]. The strengths and weaknesses of structured reports will be assessed in depth in this chapter. In addition, we analyse which initiatives carry out this project.

Due to advances in technology, the way radiologists report and their work environment have changed in a drastic manner. Since the introduction of digital radiology terms like picture archiving and communication system (PACS) or radiology information system (RIS) have gained much importance in daily routine. Although this chapter does not intend to explain the information systems of a hospital in depth, we will briefly describe the different information systems that radiologists use.

The radiology information system (RIS) allows to maximize the resources available to carry out the examinations, and it also facilitates the introduction of patient data, the exploration scheduling and the patient care control and helps identify the professionals and radiologists who perform the exploration. The RIS also allows the interconnection with the digital radiology system and has to be flexible enough to connect to the hospital information system (HIS) and to the image storage system (PACS).

The PACS manages radiological images after they are acquired by any of its supported medical imaging machine types. It has two main functions: storing the images and sending them to the

required workstation. The functional unit of a PACS is the study, which consists of one or more series, each formed by one or more images. The possibility of communication between all the devices that form a PACS is made possible by the standardization of products. Digital Imaging and Communications in Medicine (DICOM) is the standard in medical imaging [19]. This protocol defines the services that each equipment or device is able to implement, independently of the manufacturer. The most important feature of the PACS is the interaction and integration with the RIS. The integration between health information systems is achieved through information exchange protocols, for example, the well-known Health Level 7 (HL7) protocol [20].

Using this technology, engineers dedicated to healthcare ICT created the DICOM Structured Reporting (DICOM-SR). DICOM-SR is defined by how it is constructed more than by what it contains [21]. In this way, DICOM-SR structures data hierarchically into a tree of nodes. Each node has a concept name with a value. The concept name is coded from standard medical terminologies such as the Systematized Nomenclature of Medicine (SNOMED), Radiology Lexicon (RadLex) or Logical Observation Identifiers Names and Codes (LOINC), and its code is unique and language independent. Finally, the values of a node may be one of several types: text, numeric, concept coded or reference to other DICOM objects.

DICOM-SR can be simple documents, without the need to present the content into a human-readable form. Each document encodes only the content, but does not define how it has to be presented. However, it forces a restriction: the content must be unambiguous [22]. For this reason, DICOM-SR may be compared to Extensible Markup Language (XML). The XML files contain tags with meaning but without any form of presentation. So, following this parallelism, if the XML files need some presentation tool like Cascading Style Sheets (CSS), the DICOM-SR also needs an application to make its content legible.

With the objective of reducing variability, DICOM-SR includes the Information Object Definitions (IODs). Their function is similar to Document Type Definition (DTD) in XML files: to create "well-formed" documents. IODs specify the valid combinations of atomic components.

The benefits of using DICOM-SR are listed here:

- Better communication with the clinician
- More precise coding of the diagnosis, minimizing the number of refused reports
- Less typing time
- Automatic coding and language independent
- Consistent and complete information
- Reports stored next to the study images
- References to regions of interest (ROI) (organs, tissues, etc.)
- Automatic references to previous versions of the document
- No need for dictation of patient data

Nowadays the DICOM Structured Report is still in research phase. Only the Radiation Dose Structured Report (DSR) is used in hospitals in daily routine to control the radiation dose administration in patients. Recently, Rosa Medina Garcia et al. published a paper detailing a system to diagnose breast cancer through the DICOM Structured Report [23]. The main reason for the lack of implementation of structured reports is the absence of support for these files in current PACS systems.

On the other hand, a new need arises due to the integration of imaging biomarkers in clinical routine. The final result should be able to be seen and reviewed on the available information systems by radiologists and clinicians [24]. Current PACS systems do not integrate a DICOM-SR viewer, so the extra information added in DICOM files can't be exploited by the users. Moreover, there is no way of performing data mining on imaging biomarkers. A possible solution for this issue is the utilization of a framework like JasperReports or Crystal Reports that allow to create flexible reports [25, 26]. These frameworks provide a feature for creating customized templates, where the input data may be the imaging biomarkers extracted from a study. Supported

input types include XML files, databases (e.g. MySQL, Oracle or PostgreSQL), JavaBeans or comma-separated values (CSV) files.

Thanks to these frameworks, a user with programming knowledge is able to automatically create documents in report form using JAVA or C# libraries. This way, the radiologist is able to review the extracted imaging biomarkers in a result report that can improve the final diagnostic, thanks to the added quantitative information. The final report will be human-readable and stored in the PACS system as a DICOM file associated to the corresponding study. In this chapter we will introduce a possible solution to create imaging biomarker reports.

11.2 Data Mining on an Unstructured Radiology Report

Radiology reports in free text are highly variable, so if we try to create a database to exploit the information provided in unstructured reports, the end result will probably be negligible and frustrating. A possible solution is based on the application of natural language processing techniques to perform an extraction of the information included in an unstructured report.

Radiology reports present a particular lexicon, such that the identification of specific terms in the text is of great importance to improve the performance of information recovery systems [25]. An important issue related with a radiology report is that the negation of medical terms is often used to indicate the absence of a certain disease or injury. In fact, in a radiology report more than half of the terms used are negated, i.e. the radiologist uses the negation to indicate a healthy condition of the tissue. In this context, standard techniques of natural language processing fail to recover the required information in an efficient manner. Research groups work with negation recognition systems to solve this difficulty in other types of text with a similar problem than the one found in radiology reports [26]. Another important need for the natural language system is to assign weights to the most critical and important words in the report, using a dictionary of typical terms. This dictionary is restricted to the terms used by radiologists.

Using these techniques, we can develop systems with an efficient processing engine and report indexing. Natural language systems allow to manage the high number of radiology reports generated daily by a hospital without affecting the system performance. Additionally, to the indexing process, the extraction of statistical information and analysis results should also be included in order to perform a quality control of the data and the structured reports introduced into the system.

A prototype system has been implemented at La Fe Polytechnics and University Hospital in Valencia. This solution uses the extracted RIS reports as inputs. The reports are analysed by the system and stored it in a customized database. To visualize the results, a web application with a similar interface as Google search is used. This application allows searching by simple words or by concatenated words in the reports. The results show a list of reports that contains the terms of the search. Recursively, the user can indicate the most significant reports to refine the search. The final result list can be downloaded as a CSV file.

However, natural language processing suffers from limitations when performing data mining on radiology reports. There are various reasons. First, as we indicated above, there is the issue of negation [28, 29]. This is a problem that has been widely studied, but it is not a trivial one. Scientific literature indicates that a 90 % of sensitivity and specificity can be achieved [27]. Unfortunately, these rates are insufficient in clinical care. Second, natural language processing depends on the nature of the data. And third, natural language processing requires a synonym vocabulary to search efficiently.

11.3 Structured Report

According to Weiss and Langlotz, there are three structured report types [10]. The first model is written in free text, but it is split into sections; nowadays this is the most widely used model.

The second type is modelled with templates with a highlighted format; a sample is proposed by Naik et al. [6]. The third type of report uses a standardized language or lexicons, such as RadLex does. This type of report is possible thanks to technologies such as DICOM. Apart from these types, the RSNA promotes their recommended best practices to adapt the structured report to each individual or to each centre [17, 30]. In this section of the chapter, we focus on the DICOM structured report and on the template-dependent structured report.

The structured report offers many benefits. First, high quality and accuracy (a critical point) as it reduces ambiguity and variability in terms with multiple definitions [13]. Second, a structured report is able to help research, quality improvement and clinical decision support. Third, immediate information accessibility, automation and exchange between centres in order to optimize teleradiology. Fourth, the structured report facilitates the communication between radiologist and clinician, due to the fact that the clinician is able to distinguish essential from secondary information [45]. Fifth, thanks to a systemised approach, the structured report prevents discovering only one lesion and omitting the rest. It helps improving the general view of patient care.

Step by step, the structured report is gaining acceptance in clinical routine, but we can't turn a blind eye to the principal obstacles that prevent its definite implantation. The main problem is resistance to change. The time and energy required to learn and a positive attitude towards change are essential, but not always found, especially in radiologists of advanced age [13]. A higher accuracy is usually an indicator of structured reporting, but authors have questioned this assertion [37]. In fact, the largest study on structured reporting found that it suffered from a lower accuracy compared to the traditional report [7]. However, this might be partly due to limitations of the reporting system used. Another potential problem is the negative impact in the radiologist workflow. The creation of a structured report requires an increased visual attention, so the radiologist has less time to spend seeing and analysing the study images, which may disrupt the diagnosis of lesions or pathologies [10]. Furthermore, it may be difficult to summarize or provide an overview when facing complex diseases, due to the fact that structured radiology reports are split into several sections.

An important drawback of the actual structured report systems, from the point of view of the radiologist, is that it is more laborious to generate a structured report than a traditional one, leaving less time for the radiologist to examine the study images, and leading to a lower efficiency. In this sense, Sistrom and Honeyman-Buck demonstrated that the structured report and the report with free text show similar efficiency and accuracy rates for transmitting case-specific interpretative content [31]. An important aspect to keep in mind is that radiologists and clinicians must be involved in the creation of templates for structured reports [32]. Therefore, the structured reporting systems are not yet widely available in PACS or radiologist's workstations.

Some authors reported that the use of checklists may improve the accuracy of the report when using templates [33]. Anyway, there are different solutions with a different degree of restriction to create structured report systems using templates [33–35]. However, some authors indicate that the restriction level of a template depends on the diagnostic of the reported disease [36, 37]. One critical concern that radiologists have about structured reports is that they may result in a vast data increase in the information systems. However, due to all information being coded into unique fields, the reports can be indexed. This way the computational load can be reduced. On the other hand, information generated from unstructured reports does not allow any form of indexing, which implies an excessive computational load for the information system servers.

The American College of Radiology (ACR) has promoted a system over the past few decades for indexing the images and cases collected by radiologists. The ACR Index presents attractive features for image indexing with the purpose of using them as teaching materials. First, this index offers anatomic and pathologic identifiers using decimal numbers. In a very simple way, the numbers before the decimal point indicate the anatomic location, and the numbers after the

decimal point indicate the pathologic entity. Second, the ACR Index is human-readable. Each digit denotes a specific term in the taxonomy. Third, the code scheme employed by the ACR Index ensures that teaching materials are not modified by radiologists or institutions. Due to the Internet and its capabilities, the teaching materials have now an online version. Lexicons like the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) can be used to organize the information in electronic medical records with a more extensive amount of terms. The ACR Index contains only a few thousand unique terms, offering much less detail than other terminology systems [38].

The RSNA generated the Medical Imaging Resource Community (MIRC) project, offering an online tool for the creation of electronic teaching files and other forms of image libraries. Moreover, the user can annotate information to the images. This way, a need arises for a more complete and computerized index than the original ACR Index. RadLex was developed by the RSNA to fill these gaps, with the purpose of creating a complete terminology on medical imaging. RadLex' main goal is to create a terminology that can be used to annotate, index and retrieve content from MIRC. To avoid useless efforts, the RSNA and the College of American Pathologists agree to use SNOMED-CT terms as a starting point for the lexicon. Furthermore, various standards organizations, such as DICOM or Integrating the Healthcare Enterprise (IHE) have participated in the RadLex project [22, 39].

RadLex includes the anatomic and pathologic codes available in the ACR Index. In addition, it also integrates other types of terms: equipment, procedures and imaging techniques used in image acquisition, difficulty of image interpretation and image quality. Other important advantage of RadLex is the possibility of updating with new concepts, including other popular medical lexicons, such as SNOMED-CT, International Classification of Diseases (ICD)-9, Current Procedural Terminology (CPT) or BrainInfo. Nowadays, RadLex includes more than 68,000 English terms, which are available in a variety of forms: table format (downloaded directly to an Excel file), Ontology Web Language (OWL) files and as a database. RadLex terms also are available online or using the MIRC authoring tool. The RadLex project must be continuously updated to guarantee that new concepts are incorporated and to maintain the cross-links with other vocabularies.

All the information included in a report cannot be captured by a standardized lexicon. Radiologists need a narrative text to express unusual elements, to integrate a large and complex set of observations or to describe parts of the image that may not be relevant to patient care at that moment, but that should still be documented. Therefore, this is a feature that should be integrated into any reporting software. In addition to potential improvements in the quality of care, the use of lexicons has other practical advantages. The lexicons allow us to avoid different interpretations and ambiguities, since each code referring to a concept has a specific meaning for a particular coding. Hence, coding is essential for a subsequent exploitation, avoiding ambiguous interpretations that can lead to confusion. Yi Hong et al. analysed the frequency use of RadLex terms employed by radiologists when using radiology reports based on templates [44]. About 2,509 unique reporting elements were extracted from a list of 70 reports, and they were afterwards matched with RadLex. The results indicated that there was a 41 % of perfect correlation and a 26 % of partial correlation and that 33 % of the terms were uncorrelated to RadLex. Using multidimensional scaling analysis (MDS), it was discovered that 13 % of the 33 % that represented the unmatched terms were combinations of two or more RadLex terms. So, it was demonstrated that a significant overlap exists between terms of structured reports using templates and the RadLex.

DICOM is the adopted standard, and it is used in most radiology departments in hospitals. Due to the growth in new medicine areas, DICOM has expanded its standard to meet the new needs created. One of the latest additions to the DICOM standard is including structured reports using DICOM-SR. This type of DICOM file allows adding semantic information to medical images.

DICOM-SR can be used in areas with high heterogeneity and with very different data types.

For this reason, a generic specification for structured reports is highly complex. The information in a structured report requires a pattern to describe exhaustively the casuistry of the radiology report. This point is fundamental: the structure of the report must be subject to the final user needs. For example, the radiologist detects and evaluates findings on medical images, and the traumatologist assesses the need of a surgical intervention. In addition to structuring, the unification of terms that defines the information in a report can be different and complex.

The coding of concepts can help improve the structure of reports. In this sense, there are many different codes and tools that meet different needs. The most widely used ones are the ICD-9 and ICD-10 (International Classification of Diseases), which are lexicons to code diagnosis and procedures; the Systematized Nomenclature of Medicine (SNOMED); the Logical Observation Identifier Names and Codes (LOINC), a database created with the purpose of facilitating the exchange and development of a result pool; and RadLex, explained previously in detail. The great diversity of lexicons adds to the complexity of unifying concepts in a standard way in order to generate the information from a DICOM tag value.

The DICOM standard includes a code to generate structured reports in the medical field. DICOM Structured Reporting (DICOM-SR) allows the integration of the most important lexicons in medical imaging, and also it allows the inclusion of customized lexicons [22]. DICOM-SR itself establishes the guidelines to follow when preparing a DICOM structured document, and it allows a standard coding of information on the report in DICOM format (sets of "Data Elements" that match a given IOD).

DICOM-SR is able to structure semantic information. The representation of a DICOM-SR will be defined by the IODs; in DICOM-SR there are only three specified types of IODs that define a structured document: Basic Text, Enhanced SR and Comprehensive SR. DICOM-SR objects include a tree structure, which represent the structured report information as "Data Elements". This tree structure is known as the "SR Document Content" and contains the definition of the structured report. The "SR Document Content" has a tree structure in which each node keeps a relationship with its parent node and has a defined data type, as shown in Fig. 11.2. This figure shows an "SR Document Content" that describes the discovery of a round malignant tumour of 1.5 cm.

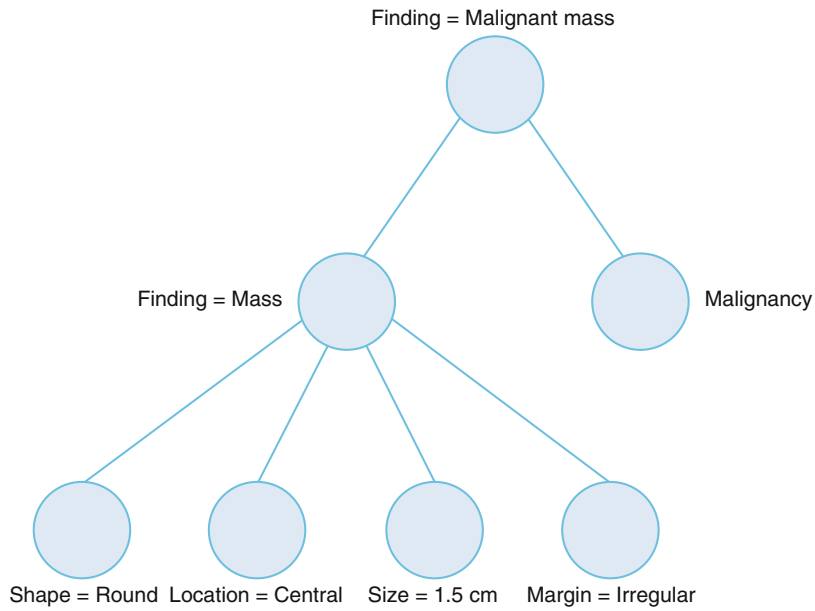


Fig. 11.2 General structure in a content tree of DICOM-SR

Moreover, the "SR Document Content" can have different structures depending on the use of DICOM-SR. There are templates for DICOM-SR that define the needed structures and constraints, such as lexicons included or the relationship between nodes. DICOM-SR also offers the possibility of validating the structures determined by these templates, as it is done with XML schemas. The templates are defined in the DICOM standard. These templates are in the process of specification and are not yet completed. At present, there are works underway to migrate DICOM-SR documents to XML and also for generating XML schemas from DICOM-SR templates [40, 41].

Therefore, DICOM-SR is a useful tool to generate structured reports because it fits perfectly into the DICOM world. It allows generation, processing and validation of these reports in a simple manner using current XML-based technologies. A structured DICOM-SR document can be encoded into XML documents, so that the way of handling and using of these documents will not be different from any XML document.

The Health Level Seven (HL7) standard is also included in its standard natively, using a structured document management called Clinical Data Architecture (CDA) [20, 42]. This format was incorporated in 1997 to structure the semantic content of clinical documents. All content of a CDA is defined through the coding of information using the XML standard. However, DICOM-SR is a complementary approach to structured documents for HL7, because currently HL7 has not proposed a rigorous structure with CDAs, i.e. it does not include templates for defining a structured report.

Another important aspect to consider is measurements taken by a radiologist from medical images using the PACS viewer. These annotations and measurements provide critical information to support the observations in the report. With the purpose of integrating them into DICOM-SR and to avoid expressions like "5 cm mass found, best seen in image 65", the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG) launched an initiative called Annotation and Image Markup (AIM) [43]. With AIM, the radiologist is able to specify what type of information he has captured when making an annotation or drawing a shape. AIM saves the position of the regions of interest (ROIs), geometry properties, anatomic entities, image observations and calculations. AIM provides a means to add this information to DICOM-SR or XML files, so it can be included directly or indirectly in structured reports.

The Breast Imaging Reporting and Data System (BI-RADS) is the best example of structured report radiology. BI-RADS includes a limited range of breast diseases, which is a good practice, and it is well suited for structured reporting. BI-RADS has five editions: 1993, 1995, 1998, 2003 and 2013 [46–50]. There are specific guides for mammography, ultrasound and magnetic resonance. The most important improvement is a better clarity and consistency in reporting, improving patient care and clinical practice. The Institute of Medicine in its 2005 report recognized that BI-RADS assessment provides an important tool for diagnosing mammography [50]. Moreover, the standardized language aids in the education of resident radiologists, also offering a more consistent practice routine [51]. The structure of BI-RADS is designed for coherent and rational examination of mammographic findings. These properties facilitate the resident training. Basset et al. indicated that 98 % of radiology residents of the United States and Canada used BI-RADS in their mammographic reports [52]. Scientific papers show that BI-RADS training can decrease variability and improve performance in spite of interobserver variability due to heterogeneity of recommendations and disease categories [52–58]. Mammography research has increased thanks to the structured report for mammography evaluation proposed by BI-RADS [18].

Over the last few years, the ACR is promoting an initiative to develop tailored lexicons, such as the Liver Imaging Reporting and Data System (LI-RADS), the Lung CT Screening Reporting and Data System (Lung-RADS) or the Head Injury Imaging Reporting and Data System (HI-RADS) [59–61]. The purpose of LI-RADS is twofold: to ensure 100 % specificity when

hepatocellular carcinoma is diagnosed using images and to improve the coherency and consistency of radiology reports. Lung-RADS is designed to standardize lung cancer reports in screening explorations with CT. Using Lung-RADS reduces confusion and facilitates monitoring the outcome. Pinsky PF et al. demonstrated in a screening trial a notable reduction on the false-positive result rate. However, sensitivity also decreased [62].

HI-RADS is being developed to standardize reports and data collection of imaging in patients with traumatic brain injury (TBI). This way, radiologists will be able to apply a consistent terminology related to the diagnosis of TBI, with these purposes: reducing errors and variability in image interpretation, enhancing communication with clinicians, facilitating research and improving the patient outcome in the long run. The ACR is preparing an Imaging White Paper establishing the guidelines for TBI diagnosis depending on the modality of the acquired images. Furthermore, the National Institute of Neurological Disorders and Stroke is building a computer system that will allow a clinician to review a head CT or a brain MRI to obtain a standardized report in a similar fashion to BI-RADS and LI-RADS. This tool could increase the quality of reports for imaging studies obtained in TBI patients.

Prostate Imaging Reporting and Data System (PIRADS) aims to enable a consistent interpretation and communication in the radiology report for prostate multiparametric magnetic resonance imaging (mpMRI) [63, 64]. It is adopted currently by the ACR and the European Society of Urogenital Radiology (ESUR). There are many studies about the efficiency and efficacy of PI-RADS; Portalez et al. demonstrated a significant increment of positive biopsies when radiologist used PI-RADS scoring in patients with at least one negative transrectal ultrasound-guided biopsy [65]. Junker et al. correlated positively high values of PI-RADS with the malignancy of the tumour [66]. Finally, Schimmöller et al. studied the inter-reader variability in PIRADS, concluding that the achieved agreement was satisfactory [67]. Both institutions have launched in 2015 a second PI-RADS version. This version is based on these studies and the acquired clinical experience, providing complete information about acquisition, interpretation of images and prostate mpMRI reporting.

11.4 Imaging Biomarkers in Reports

Imaging biomarkers in clinical assistance must provide information to radiologists and clinicians. Nowadays, there are many solutions in the scientific scenario of biomarker platforms [68–74]. These platforms' main objective is to integrate biomarkers in clinical routine, but they have an important flaw: they lack an efficient way of reporting the final results. In this section we present a user-friendly way of reporting imaging biomarker results to the radiologist using DICOM-SR.

In this process, we assume that we work with an imaging biomarker validated by radiologists and clinicians. In addition, this biomarker generates results that are stored in a database or in data files with a standard format, such as XML.

Assuming these assumptions are true, it is advisable to follow these steps when designing an imaging biomarker report. First, the formation of a roundtable with the involved stakeholders: radiologists, clinicians and biomedical engineers to discuss the methodology and content of the report. It is important to include patient data, processing date, image acquisition modality, significant parametric images and parameter values. Optionally, the report may include reference values for the parameters. Due to the fact that imaging biomarkers can generate a large number of results, it is necessary to decide which images and parameters will be included in the report. This step is critical, because the chosen parameters should add value for the radiologist at the time of diagnosis and for the clinician when deciding a course of action. Second, a report draft must be outlined and validated by all the stakeholders. Once validated, and depending largely on the technology used for the implementation of the biomarker

and the method selected for storing results, the engineers have to choose which technology should be used in order to create the report. Finally, seeking a high readability, a quick interpretation and the dismissal of superfluous information, the length of the report should not exceed one sheet.

There are several alternatives to create a report dynamically, for example, Crystal Reports, JasperReports, Business Intelligence and Reporting Tools (BIRT) or Pentaho [25, 26, 75, 76]. Originally, these software packages were developed to generate reports from a variety of sources, so that organizations could achieve a better understanding of their business. However, their potential is applicable in other fields. In this case, these tools allow to create templates of reports for the visualization of imaging biomarker results. The features of these tools are listed in Table 11.1:

	Crystal reports	JasperReports	BIRT	Pentaho
Versions				
Free	×	✓	✓	×
Payment	✓	✓	✓	✓
Report designer				
WYSIWYG editor	✓	✓	✓	✓
HTML client	✓	×	×	×
Palette	✓	✓	✓	✓
Data explorer	✓	✓	✓	✓
Property editor	✓	✓	✓	✓
Chart builder	✓	✓	✓	✓
Report preview	✓	✓	✓	✓
Script editor	✓	✓	✓	×
Outline	✓	✓	✓	✓
Expression builder	✓	✓	✓	✓
Reusability				
Report templates	✓	✓	✓	✓
Report elements	✓	✓	✓	✓
Subreports	✓	✓	✓	×
Cross tab	✓	✓	✓	×
Compiled report definition	×	×	✓	×
Parameterized reports	✓	✓	✓	✓
Data source				
JDBC	✓	✓	✓	✓
XML	✓	✓	✓	✓
POJO	✓	✓	✓	✓
Flat file	×	✓	×	×
Scripted	✓	✓	×	×
EJB	×	×	✓	×
Hibernate	×	×	✓	✓
Connection pool support	✓	✓	×	✓
Data set				
Query builder	✓	✓	✓	✓
Joined data set	✓	✓	✓	×
Customization	✓	✓	✓	×

Formatting				
Conditional visibility	✓	✓	×	×
Value mapping	✓	✓	×	×
Highlighting	✓	✓	×	✓
Interactive reports				
Hyperlink	✓	✓	×	×
Charts	✓	✓	✓	✓
Localization				
Strings in the report definition	✓	✓	✓	✓
Locale-aware data formatting	✓	✓	✓	×
Dynamic formatting	✓	✓	✓	×
Deployment				
For web application	✓	✓	×	×
For stand-alone JAVA application	×	✓	✓	×
Support and development	✓	✓	✓	✓
Documentation	✓	✓	×	✓
Integration				
Report designer	✓	✓	✓	✓
Design engine API	✓	✓	×	✓
Report engine API	✓	✓	✓	✓
Charting engine API	✓	✓	✓	✓
Performance	×	×	×	✓
Extensibility				
Custom data access	✓	✓	✓	✓
Custom report items	✓	✓	×	×
Custom chart types	✓	✓	✓	✓
Scripting	✓	✓	✓	×
Output format				
HTML	✓	✓	✓	✓
PDF	✓	✓	✓	✓
XLS	✓	✓	✓	×
DOC	✓	✓	×	×
CSV	✓	✓	✓	×
RTF	✓	✓	✓	✓
TXT	✓	✓	✓	✓
XML	✓	✓	✓	×

A prototype for the integration of imaging biomarkers in the PACS has been developed at the La Fe Polytechnics and University Hospital (Valencia, Spain). First, XML files were established as the data inputs of the report. Second, the templates used to generate the PDF reports were created with JasperReports Designer. Third, a software was implemented, with JAVA and the DCM4CHEE library, to convert PDF files to DICOM files and send them to the PACS. Figure 11.3 shows an example of an imaging biomarker report on multiple sclerosis lesions.

La Fe Hospital Universitari i Politècnic

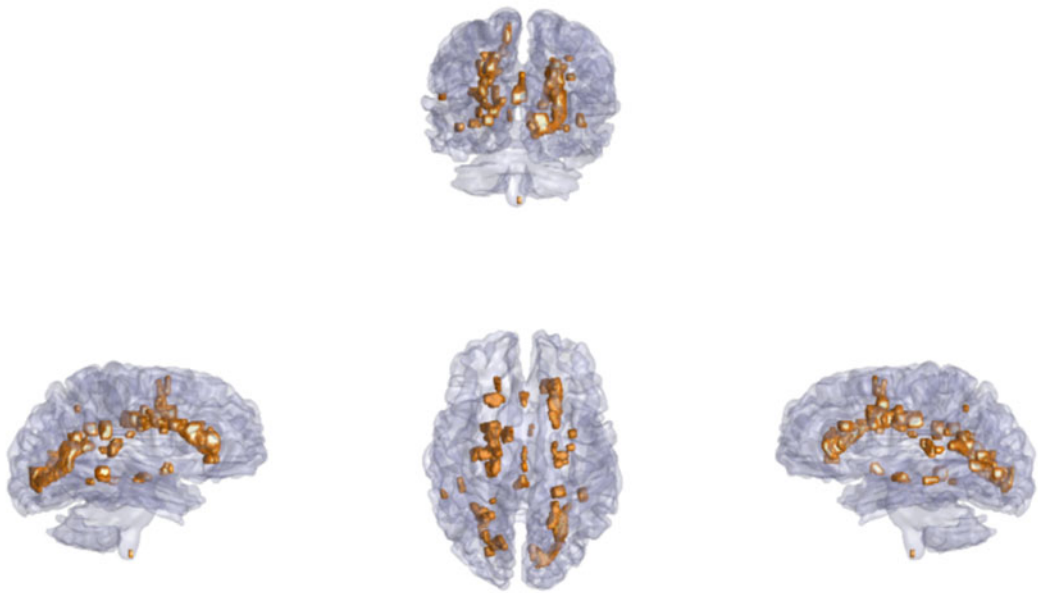
GIBI 30

quibim QUantitative Imaging Biomarkers Medicine

http://www.quibim.com
info@quibim.com

Brain
MS Lesions

Imaging Center		Patient Name	
Modality	MR	Patient ID	
Study Description		Patient Sex	
Study Date		Birthdate	



Lesion Study			
Lesion number	27	Dominant Lesion Volume [mL]	1.43
Total Lesion Volume [mL]	5.88	Dominant Volume / Total [%]	24.36
Total Relative Volume [%]	1.01	Dominant Volume / Total IC [%]	0.24
Mean Lesion Volume [mL]	0.21	Dissemination [cm]	5.35
Mean Lesion Area [mm^2]	99.06	Entropy [ua]	2.18

Data from this quantification report should be considered as the results of research with an evidence level 2 (Centre for Evidence-based Medicine) in phase of clinical approval.
QUIBIM S.L. - Quantitative Imaging Biomarkers in Medicine. Avenida Fernando Abril Martorell 106, Torre A, Biopolo La Fe, Valencia (SPAIN)

Fig. 11.3 Imaging biomarker report of multiple sclerosis lesions

Conclusions

The structured report offers opportunities to improve the efficiency of radiology departments and all of its stakeholders: clinicians, administrators, engineers and healthcare authorities. Radiologists must change their traditional way of reporting and they must be convinced of the advantages that structured reporting has to offer. Nowadays, the technology industry is developing tools to produce structured reports without compromising completeness, accuracy or workflows. Templates suggested by the RSNA, ontologies like RadLex and the DICOM Structured Report technology may become the backbone that supports achieving the final objective: improve patient care.

References

1. Enfield CD. The scope of the roentgenologist's report. JAMA. 1923;80:999–1001.
2. Sobel JL, Pearson ML, Gross K, Desmond KA, Harrison ER, Rubenstein LV, Rogers WH, Kahn KL. Information content and clarity of radiologists' reports for chest radiography. Acad Radiol. 1996;3:709–17.
3. Armas RR. Qualities of a good radiology report. Am J Roentgenol. 1998;170:1110.
4. Lafortune M, Breton G, Baudouin JL. The radiological report: what is useful for the referring physician? J Can Assoc Radiol. 1988;39:140–3.
5. Langlotz CP. Structured reporting: are we there yet? Radiology. 2009;253:23–5. doi:10.1148/radiol.2531091088.
6. Naik SS, Hanbidge A, Wilson SR. Radiology reports: examining radiologist and clinician preferences regarding style and content. AJR Am J Roentgenol. 2001;176:591–8.
7. Johnson AJ, Chen MY, Swan JS, Applegate KE, Littenberg B. Cohort study of structured reporting compared with conventional dictation. Radiology. 2009;253:74–80. doi:10.1148/radiol.2531090138.
8. Plumb AA, Grieve FM, Khan SH. Survey of hospital clinicians' preferences regarding the format of radiology reports. Clin Radiol. 2009;64:395–6. doi:10.1016/j.crad.2008.11.009.
9. Bosmans JML, Weyler JJ, De Schepper AM, Parizel PM. The radiology report as seen by radiologists and referring clinicians. Results of the COVER and ROVER surveys. Radiology. 2009;259:184–95. doi:10.1148/radiol.10101045.
10. Weiss DL, Langlotz CP. Structured reporting: patient care enhancement or productivity nightmare? Radiology. 2008;249:739–47. doi:10.1148/radiol.2493080988.
11. Park J, Pillarisetty VG, Brennan MF, Jarnagin WR, D'Angelica MI, Dematteo RP, G Coit D, Janakos M, Allen PJ. Electronic synoptic operative reporting: assessing the reliability and completeness of synoptic reports for pancreatic resection. J Am Coll Surg. 2010;211:308–15. doi:10.1016/j.jamcollsurg.2010.05.008.
12. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, Sexton B, Hyzy R, Welsh R, Roth G, Bander J, Kepros J, Goeschel C. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med. 2006;355:2725–32.
13. Schwartz LH, Panicek DM, Berk AR, Li Y, Hricak H. Improving communication of diagnostic radiology findings through structured reporting. Radiology. 2011;260:174–81. doi:10.1148/radiol.11101913.
14. Reiner B, Knight N, Siegel EL. Radiology reporting, past, present, and future: the radiologist's perspective. J Am Coll Radiol. 2007;4:313–9.
15. Boland GW. Voice recognition technology for radiology reporting: transforming the radiologist's value proposition. J Am Coll Radiol. 2007;4:865–7.
16. Reiner BI. Optimizing technology development and adoption in medical imaging using the principles of innovation diffusion. II. Practical applications. J Digit Imaging. 2012;25:7–10. doi:10.1007/s10278-011-9409-7.
17. Kahn Jr CE, Langlotz CP, Burnside ES, Carrino JA, Channing DS, Hovsepian DM, Rubin DL. Toward best practices in radiology reporting. Radiology. 2009;252:852–6. doi:10.1148/radiol.2523081992.
18. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, Mendelson EB, Wilcox PA, Butler PF, D'Orsi CJ. The ACR BI-RADS experience: learning from history. J Am Coll Radiol. 2009;6:851–60. doi:10.1016/j.jacr.2009.07.023.
19. Digital Imaging and Communications in Medicine. http://dicom.nema.org. Accessed 20 Aug 2015.
20. Health Level Seven. http://www.hl7.org. Accessed 20 Aug 2015.
21. Hussein R, Engelmann U, Schroeter A, Meinzer HP. DICOM structured reporting: part 1. Overview and characteristics. Radiographics. 2004;24:891–6.
22. Clunie D. DICOM structured reporting. 1st ed. Bangor: Pixelmed; 2000.
23. Medina García R, Torres Serrano E, Segrelles Quilis JD, Blanquer Espert I, Martí-Bonmatí L, Almenar Cubells A. A systematic approach for using DICOM structured reports in clinical processes: focus on breast cancer. J Digit Imaging. 2015;28:132–45. doi:10.1007/s10278-014-9728-6.
24. Martí-Bonmatí L, Alberich-Bayarri A, García-Martí G, Sanz-Requena R, Pérez Castillo C, Carot-Sierra JM, Manjón-Herrera JV. Biomarcadores de imagen, imagen cuantitativa y bioingeniería. Radiologia. 2012;54:269–78. doi:10.1016/j.rx.2010.12.013-.
25. Crystal reports. http://www.crystalreports.com Accessed 10 Sep 2015.
26. Jasperreports. http://jaspersoft.com. Accessed 10 Sep 2015.

27. Díaz-Galiano MC, Martín-Valdivia MT, Ureña-López LA. Query expansion with a medical ontology to improve a multimodal information retrieval system. Comput Biol Med. 2009;39:396–403. doi:10.1016/j.compbiomed.2009.01.012.
28. Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. J Am Med Inform Assoc. 2007;14:304–11.
29. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2001;34:301–10.
30. Wiki of RSNA. http://reportingwiki.rsna.org. Accessed 10 Sept 2015.
31. Reiner BI. The challenges, opportunities, and imperative of structured reporting in medical imaging. J Digit Imaging. 2009;22:562–8. doi:10.1007/s10278-009-9239-z.
32. Reiner B, Siegel E. Radiology reporting: returning to our image-centric roots. AJR Am J Roentgenol. 2006;187:1151–5.
33. Sistrom CL, Langlotz CP. A framework for improving radiology reporting. J Am Coll Radiol. 2005;2:159–67.
34. Reiner BI. Customization of medical report data. J Digit Imaging. 2010;23(4):363–73. doi:10.1007/s10278-010-9307-4.
35. Johnson AJ. All structured reporting systems are not created equal. Radiology. 2012;262:726. doi:10.1148/radiol.11111679.
36. Bosmans JM, Peremans L, Menni M, De Schepper AM, Duyck PO, Parizel PM. Structured reporting: if, why, when, how— and at what expense? results of a focus group meeting of radiology professionals from eight countries. Insights Imaging. 2012;3:295–302. doi:10.1007/s13244-012-0148-1.
37. Sistrom CL, Honeyman-Buck J. Free text versus structured format: information transfer efficiency of radiology reports. AJR Am J Roentgenol. 2005;185:804–12.
38. SNOMED International. What is SNOMED CT. http://www.snomed.org/snomedct/what_is.html. Accessed 25 Sept 2015.
39. IHE. Integrating the healthcare enterprise. http://www.ihe.net. Accessed 25 Sept 2015.
40. Zhao L, Lee KP, Hu J. Generating XML schemas for DICOM structured reporting templates. J Am Med Inform Assoc. 2005;12:72–83.
41. Lee KP, Hu J. XML schema representation of DICOM structured reporting. J Am Med Inform Assoc. 2003;10:213:23.
42. Robert H. HL7 Clinical document architecture, release 2.0. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=258. Accessed 25 Sept 2015.
43. Rubin DL, Mongkolwat P, Kleper V, Supekar KS, Channin DS. Medical imaging on the Semantic Web: annotation and image markup. AMIA Annu Symp Proc. 2008;6:626:30.
44. Hong Y, Zhang J, Heilbrun ME, Kahn Jr CE. Analysis of RadLex coverage and term co-occurrence in radiology reporting templates. J Digit Imaging. 2012;25:56–62. doi:10.1007/s10278-011-9423-9.
45. Al-Hawary MM, Francis IR, Chari ST, Fishman EK, Hough DM, Lu DS, Macari M, Megibow AJ, Miller FH, Mortele KJ, Merchant NB, Minter RM, Tamm EP, Sahani DV, Simeone DM. Pancreatic ductal adenocarcinoma radiology reporting template: consensus statement of the Society of Abdominal Radiology and the American Pancreatic Association. Radiology. 2014;270:248–60. doi:10.1148/radiol.13131184.
46. American College of Radiology. Breast imaging reporting and data system® (BI-RADS®). Reston: American College of Radiology; 1992.
47. American College of Radiology. Breast imaging reporting and data system® (BI-RADS®). 2. Reston: American College of Radiology; 1995.
48. American College of Radiology. Breast imaging reporting and data system® (BI-RADS®). 3. Reston: American College of Radiology; 1998.
49. American College of Radiology. Breast imaging reporting and data system® (BI-RADS®). 4. Reston: American College of Radiology; 2003.
50. American College of Radiology. Breast imaging reporting and data system® (BI-RADS®). 5. Reston: American College of Radiology; 2013.
51. Nass SJ, Bal J. Improving breast imaging quality standards. Washington, DC: National Academy of Science; 2005.
52. Berg WA, D'Orsi CJ, Jackson VP, et al. Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? Radiology. 2002;224:871–80.
53. Baker JA, Kornguth PJ, Floyd Jr CE. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. AJR Am J Roentgenol. 1996;166:773–8.
54. Houssami N, Boyages J, Stuart K, Brennan M. Quality of breast imaging reports falls short of recommended standards. Breast. 2007;16:271–9.
55. Geller BM, Barlow WE, Ballard-Barbash R, Ernster VL, Yankaskas BC, Sickles EA, Carney PA, Dignan MB, Rosenberg RD, Urban N, Zheng Y, Taplin SH. Use of the American College of Radiology BIRADS® to report on the mammographic evaluation of women with signs and symptoms of breast disease. Radiology. 2002;222:536–42.
56. Lehman C, Holt S, Peacock S, White E, Urban N. Use of the American College of Radiology BIRADS® guidelines by community radiologists: concordance of assessments and recommendations assigned to screening mammograms. AJR Am J Roentgenol. 2002;179:15–20.
57. Taplin SH, Ichikawa LE, Kerlikowske K, Ernster VL, Rosenberg RD, Yankaskas BC, Carney PA, Geller BM, Urban N, Dignan MB, Barlow WE, Ballard-Barbash R, Sickles EA. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. Radiology. 2002;222:529–35.

58. D'Orsi CJ, Hall FM. BI-RADS lexicon reemphasized. AJR Am J Roentgenol. 2006;187:W557.
59. American College of Radiology. Liver imaging reporting and data system version 2013.1. http://www.acr.org/Quality-Safety/Resources/LIRADS/. Accessed 1 Oct 2015.
60. American College of Radiology. Lung CT screening reporting and data system version 2013.1. http://www.acr.org/Quality-Safety/Resources/LungRADS/. Accessed 1 Oct 2015.
61. American College of Radiology. Head injury imaging reporting and data system version 2013.1. http://www.acr.org/Quality-Safety/Resources/HIRADS/. Accessed 1 Oct 2015.
62. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, Kazerooni E. Performance of lung-RADS in the national lung screening trial: a retrospective assessment. Ann Intern Med. 2015;162:485–91. doi:10.7326/M14-2086.
63. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V, Futterer JJ, Fütterer JF. ESUR prostate MR guidelines 2012. Eur Radiol. 2012;22:746–57. doi:10.1007/s00330-011-2377-y.
64. Dickinson L, Ahmed HU, Allen C, Barentsz JO, Carey B, Futterer JJ, Heijmink SW, Hoskin PJ, Kirkham A, Padhani AR, Persad R, Puech P, Punwani S, Sohaib AS, Tombal B, Villers A, van der Meulen J, Emberton M. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. Eur Urol. 2011;59:477–94. doi:10.1016/j.eururo.2010.12.009.
65. Portalez D, Mozer P, Cornud F, Renard-Penna R, Misrai V, Thoulouzan M, Malavaud B. Validation of the European society of urogenital radiology scoring system for prostate cancer diagnosis on multiparametric magnetic resonance imaging in a cohort of repeat biopsy patients. Eur Urol. 2012;62:986–96. doi:10.1016/j.eururo.2012.06.044.
66. Junker D, Schäfer G, Edlinger M, Kremser C, Bektic J, Horninger W, Jaschke W, Aigner F. Evaluation of the PIRADS scoring system for classifying mpMRI findings in men with suspicion of prostate cancer. Bio Med Res Int. 2013:252939. doi:10.1155/2013/252939.
67. Schimmöller L, Quentin M, Arsov C, Lanzman RS, Hiester A, Rabenalt R, Antoch G, Albers P, Blondin D. Interreader agreement of the ESUR score for prostate MRI using inbore MRIguided biopsies as the reference standard. Eur Radiol. 2013;23:3185–90. doi:10.1007/s00330-0132922y.
68. Hacklander T, Martin J, Kleber K. Informatics in radiology (infoRAD): an open source framework for modification and communication of DICOM objects. Radiographics. 2005;25:1709–21.
69. Goldberg IG, Allan C, Burel JM, Creager D, Falconi A, Hochheiser H, Johnston J, Mellen J, Sorger PK, Swedlow JR. The open microscopy environment (OME) data model and xml file: open tools for informatics and quantitative analysis in biological imaging. Genome Biol. 2005;6:R47.
70. Martone ME, Tran J, Wong WW, Sargis J, Fong L, Larson S, Lamont SP, Gupta A, Ellisman MH. The cell-centered database project: an update on building community resources for managing and sharing 3D imaging data. J Struct Biol. 2008;161:220–31.
71. Kvilekval K, Fedorov D, Obara B, Singh A, Manjunath BS. Bisque: a platform for bioimage analysis and management. Bioinformatics. 2010;26:544–52. doi:10.1093/bioinformatics/btp699.
72. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics. 2007;5:11–34.
73. Prodanov D. Data ontology and an information system realization web-based management of image measurements. Front Neuroinform. 2011;25:25. doi:10.3389/fninf.2011.00025.
74. Muehlboeck JS, Westman E, Simmons A. TheHiveDB image data management and analysis framework. Front Neuroinform. 2014;7:49. doi:10.3389/fninf.2013.00049.
75. BIRT Home. http://www.eclipse.org/birt. Accessed 10 Sept 2015.
76. Reporting | Pentaho community. http://community.pentaho.com/projects/reporting. Accessed 10 Sept 2015.

12 Pearls and Pitfalls in Gold Standards and Biological Correlation

David J. Lomas and Edmund Godfrey

12.1 Introduction

Imaging biomarkers in medicine are almost exclusively developed as indicators of an underlying biological process related to disease. As discussed elsewhere, they may be used for disease diagnosis, outcome prediction and to assess treatment efficacy. In these roles they may be used as surrogate markers of the underlying process but more typically as markers of longer term clinical outcomes. A frequent motivating factor is obtaining information regarding a disease process quickly – avoiding the delay required for the disease process to cause longer term clinical endpoints. Ideally the development of a biomarker requires formal validation against a reference or "gold" standard to prove that the biomarker truly reflects the underlying biological process. This is an important but often imperfectly addressed step for a wide range of reasons. The absence of, inaccuracy of or difficulty obtaining a suitable reference standard (particularly if invasive) is often part of the original rationale for developing the biomarker. This chapter addresses the challenges of biomarker validation against reference or "gold" standards illustrating this with a range of biomarker examples including three imaging biomarkers in current practice: (A) MR elastography as a surrogate marker of liver fibrosis stage, (B) CT colonography detected adenomas as an imaging biomarker of colon cancer risk and (C) CT perfusion metrics as imaging biomarkers of tumour blood supply.

12.2 Surrogacy and Biological Plausibility

Imaging biomarkers became popular following a randomised drug trial using X-ray evaluation of rheumatoid arthritis treatment. In this trial [4] an erosion score biomarker was used as a surrogate for the active synovitis component of rheumatoid disease to indicate disease activity and specifically disease progression. This biomarker was not "validated" against any other standards but was used in a qualified way (see later) and proved successful at demonstrating drug efficacy over a relatively short time period. The linkage between the biomarker and the underlying process here is biologically plausible as the pathology relating active synovitis and adjacent bone erosion is well understood.

However, there are pitfalls in the relatively empirical assumption of biological plausibility. Another drug trial [7] used the increased

D.J. Lomas (✉)
Department of Radiology,
University of Cambridge, Cambridge, UK
e-mail: djl15@radiol.cam.ac.uk

E. Godfrey
Department of Radiology,
Cambridge University Hospitals
NHS Foundation Trust,
Cambridge, UK
e-mail: edmund.godfrey@addenbrookes.nhs.uk

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_12

incidence of ventricular tachycardia (VT) on 24 h ECG tape recordings as a surrogate biomarker for sudden cardiac death. It appeared plausible that a decreased incidence of VT would be a biomarker of a decreased incidence of sudden cardiac death which was assumed to result from an arrhythmia. In practice, the drug used did decrease the incidence of VT on 24 h recordings, but paradoxically it significantly increased the rate of sudden cardiac death in the treated group. In this case, the choice of surrogate marker proved incorrect, and clearly the underlying pathophysiology of sudden cardiac death was not well enough understood. In practice implanted defibrillators proved to be the appropriate treatment option for this group of patients.

More than a decade after the CAST trial, the diabetic therapy rosiglitazone was fast-tracked through the drug approval system on the basis that it reduced Hb1Ac – itself a well-validated surrogate marker of mean blood sugar levels. It was assumed that a drug reducing blood sugar levels would reduce cardiovascular and other risks of complications related to diabetes. After introduction it was eventually noticed that the drug increased the rate of cardiac events and it was withdrawn [45]. In this case the surrogate marker (Hb1AC) was being used incorrectly as a surrogate for complications of diabetes and the side-effects of the intervention (rosiglitazone) were not properly considered.

Often the choice of surrogate marker is not driven by the underlying pathophysiology but instead by the serendipitous availability of the marker. Wherever possible surrogate markers should be fully validated against a reference or "gold" standard to prove they really are indicators of the underlying disease process or feature. The preceding examples emphasise the importance of fully understanding the relationship between a surrogate marker and the underlying disease processes. They also indicate the risks of linking a biomarker to a clinical endpoint before full validation and/or qualification of the biomarker has been demonstrated. Where biomarkers are being used to monitor or guide therapeutic interventions, they also illustrate the importance of considering whether the intervention may have unanticipated side-effects that are not reflected by the biomarker itself.

It is frequently the case that the selection of a biomarker of an underlying disease process or feature does not directly address the desired clinical questions, which are usually to make a diagnosis or predict survival or response to therapy. Frequently, these important clinical endpoints are influenced by multiple factors and not just a single underlying disease process. Often the underlying process for which an imaging biomarker is sought is an intermediate biomarker itself. In some cases, the underlying disease process or feature is known to be causal, but in many cases this is unproven, and at best there may be only evidence of an association between the process and the actual disease of interest.

Finally, the dynamic range of a biomarker and its ability to discriminate abnormal from normal is important. Some functional imaging biomarkers may appear attractive qualitatively such as using the hepatic uptake (and related T1 signal change) of the hepatobiliary agent gadoxetic acid as a marker of liver function. However, in this case, it has been shown that there is large normal phenotypic variation of the required transporter protein OATP4 [43], making measurements based on this approach difficult to interpret, although there are many publications suggesting this biomarker role [47, 67, 72]. Therefore, biomarkers require a good dynamic range and the ability to clearly discriminate between normal and abnormal and responders and nonresponders. Reducing unwanted biomarker variation also means the biomarker itself has to have acceptable reliability and repeatability as well as good precision and accuracy as discussed in earlier chapters. An assessment of the biological plausibility for our three key examples follows.

12.2.1 MR Elastography

MRE is a non-invasive technique that provides an absolute measure of bulk tissue stiffness in vivo. Initial MRE evaluation has been based on its performance against the reference standard of histopathological assessment of hepatic fibrosis.

Clinical examination and palpation at the bedside [25]. and during surgery provides the initial evidence for the relationship between increased macroscopic liver stiffness and fibrosis. Histopathological evaluation both macroscopically and microscopically has also shown an association between the presence of increased collagen deposition and progressive hepatic functional impairment [29]. There is substantial research associating the extent and type of liver fibrosis with clinical endpoints [14]. There is also evidence from other body organs that abnormally increased fibrosis related to collagen deposition is associated with increased organ stiffness. Overall there is good evidence from several sources that increased liver stiffness is associated with increased liver fibrosis and that this is likely to be causally related [17, 61].

12.2.2 CT Colonography

Evaluation of the performance of CT colonography has been based on the reference standard of optical endoscopy polyp detection. Colonic adenoma detection itself is here partly being used as a biomarker of risk of progression to colorectal cancer. It is now widely accepted that most colorectal cancers arise via the adenoma-carcinoma sequence. Evidence for this is drawn from a variety of different sources. Firstly, the geographical variation of adenoma prevalence correlates with that of colorectal cancer [10]. Prior to the practice of colonoscopic polypectomy, the natural history of polyps was observed [64], with carcinomas being found at the site of previously demonstrated adenomas. It is not uncommon to find foci of invasive malignancy within adenomas [18] and vice versa [13]. Finally, adenomas and carcinomas have a similar anatomical distribution, favouring the distal colon and rectum [9]. These factors led to the proposal of the Vogelstein model for colorectal carcinogenesis [15]. Adenoma detection at optical colonoscopy is therefore very plausible as the "gold standard" surrogate marker for cancer risk and for the assessment of non-invasive colonic imaging tests such as CT colonography.

12.2.3 CT Perfusion Metrics

CT perfusion is a tracer method based on serial imaging of the first-pass and subsequent tissue uptake of an injected bolus of an extracellular iodine-based contrast agent. Based on simple area under the time-concentration curves and multi-parametric pharmacokinetic modelling, several metrics can be derived that are related to blood flow, tissue perfusion, permeability and tracer exchange between vascular and extracellular tissue compartments. Qualitatively, the "enhancement" of suspected tumours has long been used as an indicator of a blood supply and conversely the loss of enhancement as an indicator of effective treatment, for example, in GIST [8]. A blood supply is known to be essential for providing nutrients to support the growth and dissemination of most tumours. For this reason the assessment of tumour perfusion (typically the amount of blood reaching the tumour in mls/min/ml of tissue) and related parameters are considered important for predicting disease progression and the potential for tumour therapy delivered via the blood supply. However other factors apart from the amount of blood being delivered to the tumour (i.e. perfusion) are known to influence tumour growth and outcomes and in particular the nature and distribution of the blood vessels, their origin and their efficiency in nutrient delivery to the tumour as well as level of oxygenation are key factors that may vary for a given level of "perfusion".

12.3 Selecting a Suitable Reference Standard for Validation

The reference standard should ideally be an established method of measuring exactly the same process or feature that the biomarker is supposed to be measuring. As already discussed this is often intrinsically difficult and may be impractical to undertake in vivo in humans. Furthermore this is predicated on the correct selection of the biomarker in relation to the underlying disease

process as discussed earlier. Any limitations of the biomarker or confounding factors also need to be taken into account when comparing with a reference standard.

12.3.1 Liver MRE

Direct mechanical measurements of liver stiffness have proven difficult to obtain in vivo and are probably only practical during open abdominal surgery. In practice, *histopathology fibrosis staging* has been widely used as the reference standard for the initial evaluation of MRE. The histopathological evaluation of liver fibrosis has been validated against clinical endpoints [14]. Although the link between hepatic fibrosis and liver stiffness is plausible as outlined above, the histopathological evaluation of liver disease typically includes the evaluation of many other features (e.g. inflammation, proliferative rate, vascular impairment, cellular necrosis, immunohistochemistry). This has led to different disease-specific staging systems which can be difficult to compare directly with a single metric of liver stiffness. An alternative approach is to focus on the collagen component of fibrosis alone and employ semi-automated computerised quantitative assessments of the degree of fibrosis, for example, using trichrome stains for collagen [46].

12.3.2 CT Colonography

CTC has been assessed against the double contrast barium enema (DCBE) in a variety of studies including the SIGGAR trial [26]. DCBE was used as a gold standard given that it was widely available, acceptable to patients, gave results almost immediately and was plausible given what was known regarding colorectal cancer pathogenesis. Given the mounting evidence that DCBE is inferior in accuracy to both colonoscopy and CTC [3], it is no longer in widespread clinical use. Its poor performance and lack of availability makes it unsuitable for continued use as a gold standard.

Optical colonoscopy detection of adenomas with biopsy is now considered the gold standard reference test for colorectal adenomas and colorectal cancer. Apart from surgical resection, it is the only way of gaining histological confirmation of the presence of neoplasia. New tests such as CT colonography are therefore validated against optical colonoscopy which is an attractive gold standard given its wide clinical use, immediate results and patient acceptability.

12.3.3 CT Perfusion Metrics

Defining a reference standard for human in vivo measurements of tumour perfusion has proven extremely difficult and identifying suitable standards challenging. The best widely accepted reference standard for tissue perfusion is the use of injected microspheres with subsequent quantitative analysis of the tissue either by microscopy or more recently radioactivity emission. This is not practical to perform in man but animal studies in kidney [35], heart [63] and liver [38] have provided indirect evidence that CT-based perfusion parameters in humans are likely to be valid. Alternative reference standards include the histopathological assessment of blood vessels (e.g. microvessel density) and the use of other tracer-based measurements (^{15}O water PET).

In the 1990s, tumour microvessel density (MVD) was proposed as a practical alternative reference standard that could be evaluated in tumour biopsy or resection samples and therefore used for validation and potentially in clinical practice. Biologically, this approach appears highly plausible, i.e. that the amount of vasculature within a tumour would be a marker of overall tumour perfusion. However, in many human (and animal) tumours where this relationship has been studied, it does not appear to be valid. It is thought that the chaotic nature and variable permeability of tumour vessels makes the performance and efficiency of tumour perfusion relatively unrelated to MVD, and this has been demonstrated in the endometrial tumours [24] and colorectal tumours [12]. This also probably explains why

MVD does not appear to be associated with therapeutic efficacy [28].

PET-based radiotracers such as ^{82}Rb, ^{13}N-ammonia and ^{15}O water PET studies offer another potential approach to perfusion validation in vivo. These approaches have almost exclusively been applied to myocardial perfusion, and they are both expensive and complex requiring whole-body PET systems, specialised generators and cyclotron facilities to generate the short-lived ^{15}O water radiotracer.

12.4 Imperfect Standards: Effects

In practice, it is virtually impossible to have a reference "gold" standard for biomarker validation that is perfect with 100 % precision, sensitivity, specificity and accuracy. Typically, the standard (as well as the biomarker) will be influenced by other factors apart from the underlying disease process or feature of interest. These factors can reduce specificity and sensitivity and introduce bias or create confounding depending on the particular situation in which the biomarker is being used [69]. In addition, the intrinsic variability (i.e. repeatability and reliability) of the reference standard if large will reduce precision and may limit or prevent biomarker validation. An often-overlooked pitfall is that if the biomarker of interest actually has better performance than the reference standard, then it will be impossible to prove this without further evaluation and its real value may be missed. This issue has been highlighted by the inappropriate but frequent use of serum creatinine as a "gold standard" for evaluating new biomarkers of acute kidney injury [68].

12.4.1 Liver MRE and Histopathology

Although histopathology has been validated against clinical outcomes in chronic liver disease and a panel of conventional microscopy features along with immunohistochemical markers are widely used to diagnose and evaluate liver disease, there are recognised limitations. The most obvious is sampling error as the majority of liver biopsy specimens are small core samples that may be influenced by regional variation in the liver. This is well appreciated in imaging with spatial variation of fat and fibrosis being relatively common findings in routine examinations on US, CT and MRI (Fig. 12.1). Is there proof that fibrosis observed at histology does correlate with liver stiffness assessed by other methods? There is biopsy and mechanical testing-based evidence this relationship does hold [40], and of course clinical palpation of tumours and other lesions supports this linkage. However, is the relationship a linear one? Is there evidence that the actual physical amount of collagen correlates with stiffness changes or is there a threshold or plateau effect that we are unaware of in the relationship? These latter questions remain largely unanswered and are clearly difficult to address in vivo. Another key limitation is the assumption that the fibrosis grading – typically a 5- or 6-point scale reflects just increasing amounts of collagen in the sample. In practice other features such as the appearance of the collagen (e.g. "bridging" pattern) and not just the amount influences the histopathology grading level. A further complication with the reference standard is that it is a largely subjective standard relying on human experience and observation. There are several studies indicating that substantial

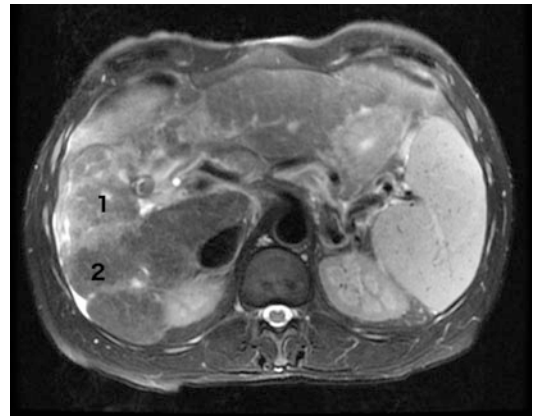


Fig. 12.1 A T2w fast spin echo image of the liver demonstrating geographic variation of fibrosis. Clearly biopsies obtained at location *1* and *2* are likely to differ considerably in their grading undermining the concept of a whole liver fibrosis stage based on histopathology

interobserver and intraobserver variation exists for liver fibrosis grading even by experienced histopathologists [6, 55].

12.4.2 CT Colonography and Optical Colonoscopy

Establishing the sensitivity of optical colonoscopy is an area that is difficult, largely because of the absence of an alternative, easily available reference standard. A per-polyp miss rate of 6 % for large adenomas (>10 mm) has been suggested by comparing back-to-back colonoscopies [56]. Given that this study used colonoscopy as its own reference standard, the miss rate of 6 % is likely to be an underestimate. For example, a polyp hidden behind a fold would be more likely to be missed by both colonoscopists (Fig. 12.2). This limitation may explain why when CT colonography is used as the reference standard, optical colonoscopy has a much higher miss rate of 12 % for large adenomas (10 mm or more in size) [51].

All colonoscopies are not equal however, for example, there is considerable variation in the polyp detection rate between colonoscopists [59] and even for the same colonoscopist at different times of day [66]. The upshot of this is that studies assessing the positive predictive value of CTC will be entirely dependent on the quality of local colonoscopy. Even in large academic centres, a significant proportion of so-called CTC false positives missed at initial colonoscopy will be true positive findings on repeat targeted colonoscopy [52].

12.4.3 CT Perfusion Metrics

As discussed earlier CT perfusion presents particular problems as none of the reference standards available in vivo can actually directly measure perfusion. Added to this the analysis models used for the biomarker itself vary, rely on different assumptions and algorithms (Fig. 12.3), may have different metrics and have been shown to generate differing results based on the same raw acquisition data [74]. In this situation it is clear that not only are the CT perfusion biomarkers generated by CT multiple but

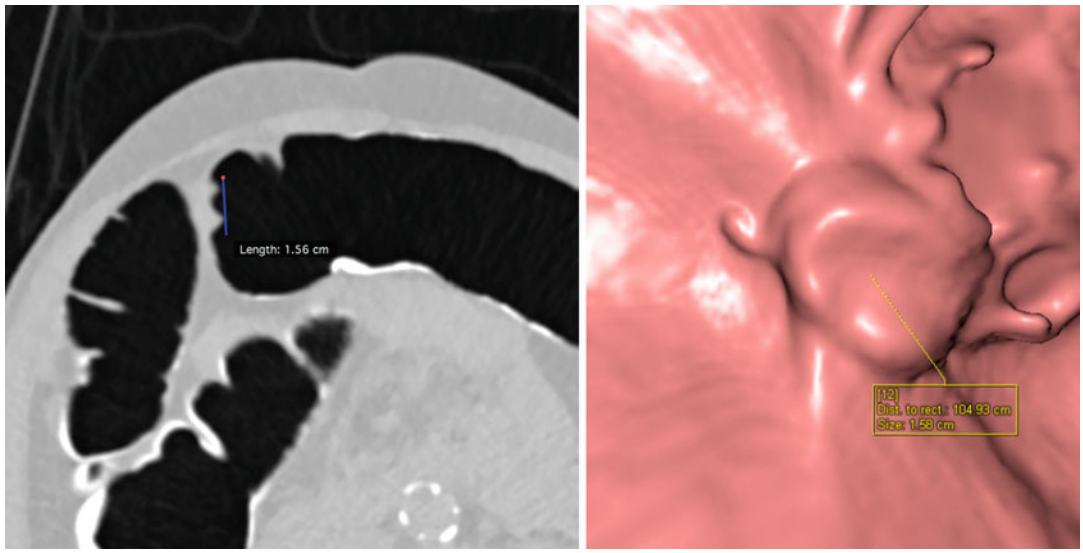


Fig. 12.2 A 1.5 cm polyp cancer at CT colonography on 2D (*left*) and 3D magnified (*right*). It is difficult to establish without repeat colonoscopy whether this polyp is actually a false negative at colonoscopy and true positive at CT colonography – or the reverse. It is typically very difficult to establish that a new test performs better than the reference test used and it has been argued that CT colonography should be the reference standard by which to validate optical colonoscopy

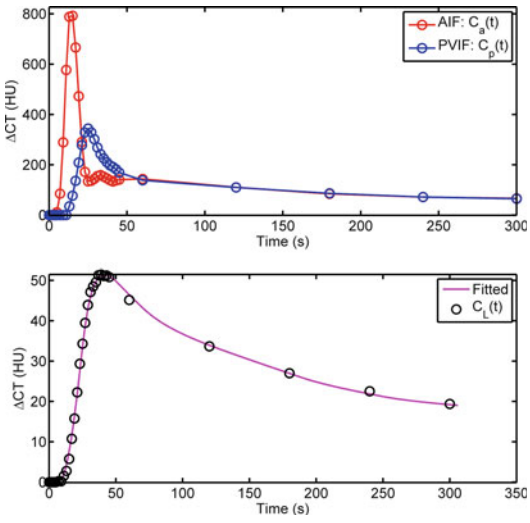


Fig. 12.3 CT time-density curves (*upper graph*) from a dynamic liver study demonstrating the measured arterial signal (*red*) and the portal venous signal (*blue*). In the lower graph, a multi-parameter model has been used to fit a curve to the simultaneously acquired liver parenchyma signal measurements. Several different models have been used with varying background assumptions leading to variable results. In some applications image registration to counter physiological motion adds another variable factor into the analysis. The lack of a widely accepted robust reference standard for validating tissue perfusion biomarkers has probably contributed to the profusion of different types of model and analysis software, in turn limiting clinical application (Courtesy of Dr A. Gill)

they have substantial intrinsic variability adding to the difficulty of validation.

PET-based radiotracer approaches offer possibly the best in vivo option for validating CT perfusion as a biomarker. However, these techniques also have several potential confounding factors, such as attenuation correction, low spatial resolution and correction for blood pool signal, and the results of the two techniques are difficult to directly compare. The modelling approaches taken for CT and PET perfusion have evolved separately and the metrics generated do not directly compare [1, 39]. In practice, the models developed for PET also differ between radiotracers. Although ^{15}O-labelled water is an ideal tracer for flow quantification (as its uptake is linearly related to flow), a single-compartment model is used, whereas a 3-compartment model has been developed for blood flow quantification with ^{13}N-ammonia and a two-compartment model is used for ^{82}Rb [2].

12.5 Managing Imperfect Standards

Given that the majority of reference standards are imperfect, it is important that the limitations are recognised when validation is being considered. An important preliminary step is to ensure good reliability and repeatability of the biomarker itself. Then attempts should be made to improve the reliability and repeatability of the reference standard or develop a new more robust standard [20]. Typically, most reference standards are influenced by some factors which do not affect the biomarker so reliability and repeatability may be improved by identifying these factors and carefully controlling for them. This might be achieved by using carefully specified conditions for the validation process or by excluding those reference measurements where the additional factors clearly have an influence. If the standard remains "flawed" at this stage, it may still be possible to quantify any bias effect and account for this using a statistical model approach. Statistical strategies for combining information from two "imperfect" standards have been devised for the validation [42]. Alternatives might include the use of imaging "phantoms" where a carefully defined and known test object can be used as an ex vivo substitute for an in vivo measurement. This approach has been used, for example, to initially validate MR measurements of fat fraction, hepatic iron concentration and tissue stiffness; however ex vivo references usually introduce their own limitations, which may influence the results. Finally, where a suitable reference standard cannot be found to validate the biomarker, then it has been argued that it is still acceptable to provide specific qualification of a biomarker based on its performance against clinical endpoints.

12.5.1 Liver MRE and Histopathology

As discussed in the previous section, there are many reasons why histopathology is an imperfect measure of liver stiffness against which to validate MR elastography. The effects of some of these factors can be managed by controlling the relevant biomarker and reference standard conditions to make them both as likely as possible to reflect the desired underlying process. As an example for MRE, excluding livers from validation studies where there is known iron accumulation (Fig. 12.4) or regions of focal confluent fibrosis (on imaging) would reduce sampling variation. Avoiding cases with other obvious biomarker-confounding factors (Fig. 12.5) such as acute inflammation (e.g. acute hepatitis) or where elevated vascular or biliary tract pressure changes are likely (e.g. right heart failure, bile duct obstruction) could also improve the validation process. Using pathologist consensus and a single "fibrosis" grading scale may improve reliability and repeatability of the standard. Given the subjective and non-linear nature of the established grading scales for fibrosis, an alternative approach to improve the reference standard that has been tried is semiautomated quantitative collagen estimation [16, 46] within a selected microscope field of view. This is achieved using specific collagen stains such as sirius red and automated microscopy and image analysis methods to quantify the proportion of stain present as a percentage of the whole field of view [21]. Several authors have considered direct mechanical stiffness measurements in post-mortem livers, but clearly the lack of vascular perfusion and alteration of tissue mechanical properties with temperature make this approach unlikely to be comparable to in vivo measurements. A method for mechanically assessing the liver stiffness at open surgery has been developed [44], but as the authors discuss

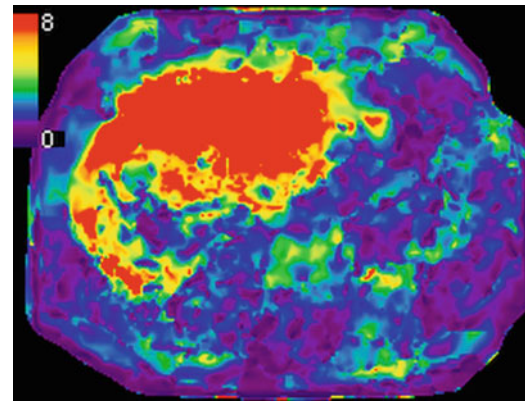


Fig. 12.5 An abnormally stiff liver (>8kPa) at MRE indicating advanced fibrosis. However, it is important to recognise that increased liver stiffness may occur acutely owing to hepatitis, hepatic vein occlusion and biliary obstruction. Validation studies should recognise and exclude such confounding aetiologies for both the biomarker and the reference standard

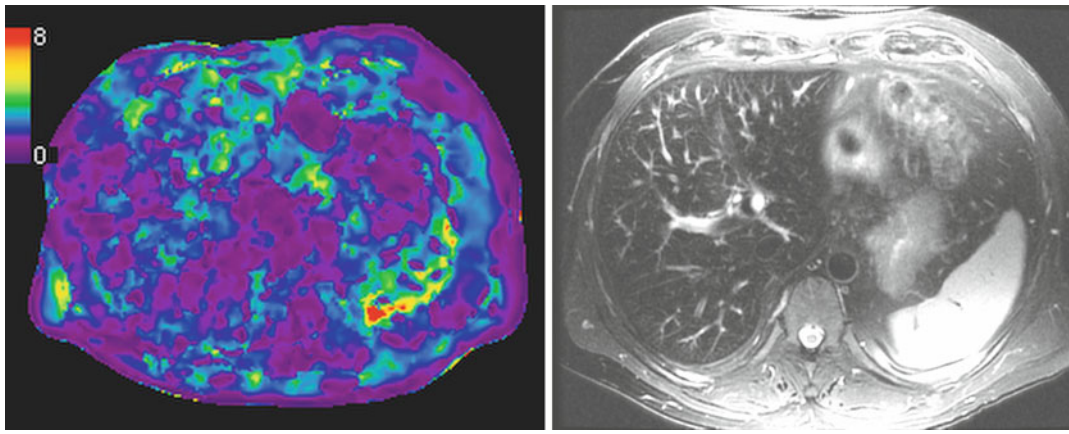


Fig. 12.4 An elastogram (*left*) giving heterogeneous values of 0–2kPa over the liver. This could be interpreted as normal, but in fact the MRE measurement has failed completely as liver signal is abnormally reduced on the T2w imaging (*right*) owing to accumulation of iron in the liver. This has reduced the signal from liver making shear wave detection impossible

in some detail, their results do not correlate well with other approaches including MR methods. This is partly owing to their tube aspiration method being influenced by the constraining liver capsule. It is also not yet fully understood how liver stiffness is influenced by the "closed" abdomen and intra-abdominal pressure changes quite apart from possible influences of anaesthetic drugs on hepatic haemodynamics. Finally statistical approaches to utilising more than one imperfect reference standard have been applied to achieve liver fibrosis validation with mixed results [53, 54].

12.5.2 CT Colonography and Optical Colonoscopy

Training is of crucial importance in maintaining a high adenoma detection rate, and improving training can have a positive effect [31]. In the UK, the Joint Advisory Group (JAG) on GI endoscopy sets standards for competence in GI endoscopy and quality assures endoscopy units, training and services. As an example of how JAG monitors standards, research using the JAG Endoscopy Training System (JETS) database has provided evidence that the current requirement for trainee colonoscopists to complete 200 procedures prior to assessment may be too few [70]. In the future, this research is likely to drive up the minimum requirements for independent practice.

Various technical improvements in colonoscopy have been used in attempts to drive up the adenoma detection rate. Chromoendoscopy describes the use of dye spraying during colonoscopy and is a way of improving sensitivity for diminutive polyps [33]. Technique changes, for example, changing the patient's position during the examination [33], as well as the use of accessories such as the "Endocuff" [35], can also improve the adenoma detection rate.

In addition to improving the "gold standard", the new biomarker can be validated against itself over time. For example, a polyp seen at CTC in the same place but larger over time is likely to be a genuine finding even if not seen at colonoscopy. This can be used as justification for repeating invasive tests such as colonoscopy [54].

12.5.3 CT Perfusion Metrics

CT perfusion biomarkers are largely based on pharmacokinetic models, and because there are several different models, different analysis software and potentially different methods of data acquisition, it is important to control for these. Most research groups have developed a specific technique often using a specific CT system to ensure acceptable repeatability and reliability of the biomarker metrics. It is of interest that the simplest least processed metric – the integrated area under the time-concentration curve (IAUCC) – is often the metric with highest correlation in many research studies rather than the model-based calculated parameters such as Ktrans, distribution volume, etc.

Having established a repeatable CT perfusion biomarker then, the problem of validation arises as the best techniques are not practical or ethical in humans in vivo. Animal-based validation has been performed successfully using microsphere techniques giving some confidence in the technique as discussed earlier, but in man, imperfect standards such as ^{15}O or ^{82}Rb tracer methods must be used. As these typically use different tracers (e.g. freely diffusible $H_2{}^{15}O$ vs extracellular iodine-based CT tracers), models and different metrics, direct comparison for validation is very difficult, and unsurprisingly results often vary substantially [22, 23]. In addition other factors may influence the reference standard results, for example, radioactive tracer dose variation, cardiac output, arterial blood sampling variation and filtering and post-processing algorithms. Instead of absolute metrics, those based on pre- and post-intervention ratio changes of the biomarker can be compared with the same changes in a reference standard, and this may prove to be the best in vivo validation that is available. As a result of these challenges and practical issues, the validation of CT perfusion in man has been severely limited. However, based on the evidence from animal work and clinical studies, and despite the relatively poor validation, there are many advocating [33] that the technique can still be successfully "qualified" for use as a tumour biomarker.

12.6 Biomarker Qualification

When a biomarker has been validated against a reference standard, it may still require qualification in order to confirm its relationship with all important clinical endpoints. This has become the accepted process for a biomarker to be approved by regulatory authorities such as the FDA or the EMA for use as a surrogate endpoint in clinical drug trials. Where validation of a biomarker proves impossible, for example, ethically, or is limited by the problems of the inadequate or impractical reference standard, then a case can still be made for biomarker qualification.

Qualification involves linking the biomarker directly to biological and clinical endpoints, and there is an inherent risk in this step as the plausible relationship between the biomarker and the underlying biology is extended towards disease outcomes and any assumed relationship may in fact be more associative than causal. The risk of confounding and bias increases and the biological linkage may no longer be valid – as in the CAST trial discussed initially. Qualification requires demonstrating that biomarker changes truly reflect clinical endpoints. This often requires carefully controlled conditions, for example, by limiting biomarker use to specific disease cohorts, specific clinical endpoints and under a specific range of conditions that reduces the risk of other factors influencing or confounding the results. This may involve constraining the biomarker acquisition to a specific imaging system manufacturer and analysis software. If prospective trials can demonstrate the value of the biomarker under these specific conditions, then it can be qualified for use in trials – but only under the same "qualifying" conditions.

12.6.1 Liver MRE

MRE has been extensively validated using histopathologic fibrosis stage as a gold standard, but awaits qualification for its use in predicting clinical endpoints such as development of HCC, liver decompensation or death. Given that the histopathologic fibrosis stage has only relatively recently been qualified against these clinical endpoints [14], it is perhaps not surprising that MRE awaits this step. There is some initial evidence that MRE-measured liver stiffness may be a risk factor for developing HCC in patients with chronic liver disease referred for MRI following an abnormal ultrasound [44]. Although this does not qualify the technique for the prediction of HCC development, it is at least encouraging. Further support can be drawn from qualification of similar biomarkers, in this case the demonstration that ultrasound elastography is predictive of clinical endpoints [50, 63], but this does not obviate the need for separate qualification of MRE. Several studies have also demonstrated that liver MRE could in future be qualified as a predictor of portal hypertension and variceal haemorrhage [59, 62, 67]. Given the limitations of histopathological staging, it is probably appropriate that qualification of hepatic MRE is explored until a better reference standard for validation is developed.

12.6.2 CT Colonography

CTC has been extensively validated in patients undergoing both CTC and optical colonoscopy (using optical colonoscopy as the gold standard) [3, 27, 30, 38]. Data to support qualification against hard clinical endpoints are emerging but as yet are incomplete. In patients undergoing CTC to clear the colon proximal to an obstructing tumour, there is evidence using the resection specimen as the gold standard that CTC is 100 % sensitive for malignancy but only 70 % sensitive for lesions $\geq$ 6 mm [51].

Qualification of CTC for its use more widely in symptomatic patients or as a screening tool is lacking. The endoscopic removal of colonic adenomas has been shown to reduce deaths from colorectal cancer [60]. The effect of carrying out CTC (and subsequent endoscopic polypectomy) on cancer-specific or all-cause mortality in symptomatic or asymptomatic patients has yet to be reported – this is likely a reflection of the relative novelty of the technique and the time it would take for meaningful results to emerge (i.e. at least

5 years). As there are ongoing trials evaluating CTC in screening, it may be that this is the next context for its qualification [11].

12.6.3 CT Perfusion Metrics

Owing to the difficulties of validating CT perfusion metrics in vivo in humans and the complexity of the various models and metrics employed, there has been a tendency to qualify CT perfusion metrics against clinical endpoints. Often these studies have involved relatively small cohorts and several authors [19] consider that further studies using much large cohorts are needed to properly qualify CT perfusion metrics for clinical use. In order to try and standardise as well as reduce the complexity and variability of CT perfusion metrics, several groups have also issued guidelines on the technique and analysis [41].

Many of these relatively small studies have demonstrated that CT perfusion metrics that indicate "high" levels of tumour perfusion at baseline correlate with a relatively good response to therapy and overall outcomes, for example, in breast cancer [37]. The survival and response of liver metastases to radioembolisation has been correlated with baseline CT perfusion metrics [43]. Perfusion metrics have also been used to predict survival in gliomas [73] and outcomes in head and neck tumours [5]. Other qualification studies include stroke outcomes which have been correlated with CT perfusion [52] and the risk of hyperperfusion syndrome after carotid stenting [75].

12.7 Pearls to Help Avoid Pitfalls

This chapter should make clear that full validation of imaging biomarkers against a robust reference standard is often extremely difficult to achieve. The list of key questions in Table 12.1 should be considered when validation of an imaging biomarker is being planned. It is important to understand that even following validation of a biomarker it is often impossible to demonstrate unequivocally that the underlying biological processes (and their reference standards) used for validation are causal in respect of the desired clinical endpoints when a biomarker is utilised in a specific disease setting. The inevitable desire to use convenient non-invasive imaging biomarkers as a surrogate endpoint in trials should not override the need for strong supporting evidence both for validation and qualification of the biomarker wherever possible. Many imaging biomarkers lack evidence in both respects but are being promoted as surrogates for use in clinical trials. The imaging community needs to be aware of the pitfalls not just inherent in the use of inadequate or misleading reference standards but perhaps more importantly of the need to focus carefully on undertaking carefully designed prospective qualification trials in ideally large populations. The paucity of these trials explains why, despite substantial interest in this area, only a small number of imaging biomarkers have survived the regulatory process of qualification as surrogate endpoints for drug trials or clinical use.

Table 12.1 Questions to ask before attempting imaging biomarker validation

Have I chosen the most appropriate biomarker?
Have I chosen the best available reference standard?
Is the reference standard practical/ethical to use for validation?
How will I validate the biomarker with the reference standard?
Is there a plausible biological link between biomarker, reference standard and clinical outcomes?
What factors influence the biomarker and reference standard that may confound validation?
Can these factors be adequately controlled for during validation and clinical use?
What is the sensitivity and specificity of the reference standard for the underlying process?
What is the reliability and repeatability of the biomarker and the reference standard?
Can the standard be improved before using it for validation?
Can I quantify and take account of any bias effects in the reference standard?
Can I use more than one reference standard for validation?
Can the biomarker be qualified for specific relevant clinical outcomes?

References

1. Aboagye EO, Gilbert FJ, Fleming IN, Experimental Cancer Medicine Centre Imaging Network Group, et al. Recommendations for measurement of tumour vascularity with positron emission tomography in early phase clinical trials. Eur Radiol. 2012;22:1465–78.
2. Al-Mallah MH, Sitek A, Moore SC, et al. Assessment of myocardial perfusion and function with PET and PET/CT. J Nucl Cardiol. 2010;17:498–513.
3. Atkin W, Dadswell E, Wooldrage K, et al. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. Lancet. 2013;381:1194–202.
4. Bathon JM, Martin RW, Fleischmann RM, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. N Engl J Med. 2001;343:1586–93.
5. Bisdas S, Rumboldt Z, Surlan-Popovic K, et al. Perfusion CT in squamous cell carcinoma of the upper aerodigestive tract: long-term predictive value of baseline perfusion CT measurements. AJNR Am J Neuroradiol. 2010;31:576–81.
6. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med. 2001;344:495–500.
7. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med. 1989;321:406–12.
8. Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol. 2007;25:1753–9.
9. Chu DZ, Giacco G, Martin RG, et al. The significance of synchronous carcinoma and polyps in the colon and rectum. Cancer. 1986;57:445–50.
10. Clark JC, Collan Y, Eide TJ, et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. Int J Cancer. 1985;36:179–86.
11. de Haan MC, Pickhardt PJ, Stoker J. CT colonography: accuracy, acceptance, safety and position in organised population screening. Gut. 2015;64:342–50.
12. Dighe S, Blake H, Jeyadevan N, et al. Perfusion CT vascular parameters do not correlate with immunohistochemically derived microvessel density count in colorectal tumors. Radiology. 2013;268:400–10.
13. Eide TJ. Remnants of adenomas in colorectal carcinomas. Cancer. 1983;51:1866–72.
14. Everhart JE, Wright EC, Goodman ZD, et al. Prognostic value of Ishak fibrosis stage: findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. Hepatology. 2010;51:585–94.
15. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67.
16. Friedenberg MA, Miller L, Chung CY, et al. Simplified method of hepatic fibrosis quantification: design of a new morphometric analysis application. Liver Int. 2005;25:1156–61.
17. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology. 2008;134:960–74.
18. Gillespie PE, Chambers TJ, Chan KW, et al. Colonic adenomas – a colonoscopy survey. Gut. 1979;20:240–5.
19. Goh V, Ng QS, Miles K. Computed tomography perfusion imaging for therapeutic assessment: has it come of age as a biomarker in oncology? Invest Radiol. 2012;47:2–4.
20. Gold R, Reichman M, Greenberg E, et al. Developing a new reference standard. Acad Radiol. 2010;17:1079–82.
21. Goodman ZD, Becker Jr RL, Pockros PJ, et al. Progression of fibrosis in advanced chronic hepatitis C: evaluation by morphometric image analysis. Hepatology. 2007;45:886–94.
22. Grüner JM, Paamand R, Kosteljanetz M, et al. Brain perfusion CT compared with 15O-H2O PET in patients with primary brain tumours. Eur J Nucl Med Mol Imaging. 2012;39:1691–701.
23. Grüner JM, Paamand R, Højgaard L, et al. Brain perfusion CT compared with 15O-H2O-PET in healthy subjects. EJNMMI Res. 2011;1:28.
24. Haldorsen IS, Grüner R, Husby JA, et al. Dynamic contrast-enhanced MRI in endometrial carcinoma identifies patients at increased risk of recurrence. Eur Radiol. 2013;23:2916–25.
25. Hamilton Bailey's demonstration of physical signs in clinical surgery. Chapter 35. 19th Edition. Editors: Lumley JSP, D'Cruz AK, Hoballah JJ et al. 2016. CRC Press.
26. Halligan S, Wooldrage K, Dadswell E, et al. Computed tomographic colonography versus barium enema for diagnosis of colorectal cancer or large polyps in symptomatic patients (SIGGAR): a multicentre randomised trial. Lancet. 2013;381:1185–93.
27. Heresback D, Djabbari M, Riou F, et al. Accuracy of computed tomographic colonography in a nationwide multicenter trial, and its relation to radiologist expertise. Gut. 2011;60:658–65.
28. Hlatky L, Hahnfeldt P, Folkman J. Clinical application of antiangiogenic therapy: microvessel density, what it does and doesn't tell us. J Natl Cancer Inst. 2002;94:883–93.
29. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. J Hepatol. 1995;22:696–9.
30. Jensch S, de Vries AH, Peringa J, et al. CT colonography with limited bowel preparation: performance characteristics in an increased-risk population. Radiology. 2008;247:122–32.
31. Kaminski MF, Anderson J, Valori R, et al. Leadership training to improve adenoma detection rate in screening colonoscopy: a randomized trial. Gut. 2016;65(4):616–24.
32. Kim AH, Kamaya A, Willmann JK. CT Perfusion of the Liver: Principles and Applications in Oncology Radiology. 2014;272:322-344.

33. Köksal AŞ, Kalkan IH, Torun S, et al. A simple method to improve adenoma detection rate during colonoscopy: altering patient position. Can J Gastroenterol. 2013;27:509–12.
34. Lapalus MG, Helbert T, Napoleon B, et al. Does chromoendoscopy with structure enhancement improve the colonoscopic adenoma detection rate? Endoscopy. 2006;38:444–8.
35. Lenze F, Beyna T, Lenz P, et al. Endocuff-assisted colonoscopy: a new accessory to improve adenoma detection rate? technical aspects and first clinical experiences. Endoscopy. 2014;46:610–4.
36. Lemoine S, Papillard M, Belloi A, et al. Renal perfusion: noninvasive measurement with multidetector CT versus fluorescent microspheres in a pig model. Radiology. 2011;260:414–20.
37. Li SP, Makris A, Gogbashian A, et al. Predicting response to neoadjuvant chemotherapy in primary breast cancer using volumetric helical perfusion computed tomography: a preliminary study. Eur Radiol. 2012;22:1871–80.
38. Liedenbaum MH, de Vries AH, van Rijn AF, et al. CT colonography with limited bowel preparation for the detection of colorectal neoplasia in an FOBT positive screening population. Abdom Imaging. 2010;35:661–8.
39. Materne R, Van Beers BE, Smith AM, et al. Non-invasive quantification of liver perfusion with dynamic computed tomography and a dual-input one-compartmental model. Clin Sci (Lond). 2000;99:517–25.
40. Mazza E, Nava A, Hahnloser D, et al. The mechanical response of human liver and its relation to histology: An in vivo study. Medical Image Analysis. 2007;11:663–672.
41. Miles KA, Lee TY, Goh V, Experimental Cancer Medicine Centre Imaging Network Group, et al. Current status and guidelines for the assessment of tumour vascular support with dynamic contrast-enhanced computed tomography. Eur Radiol. 2012;22:1430–41.
42. Miller E, Huppert A, Novikov I, et al. Estimation of infection prevalence and sensitivity in a stratified two-stage sampling design employing highly specific diagnostic tests when there is no gold standard. Stat Med. 2015;34(25);3349-61.
43. Morsbach F, Pfammatter T, Reiner CS, et al. Computed tomographic perfusion imaging for the prediction of response and survival to transarterial radioembolization of liver metastases. Invest Radiol. 2013;48:787–94.
44. Motosugi U, Ichikawa T, Koshiishi T, et al. Liver stiffness measured by magnetic resonance elastography as a risk factor for hepatocellular carcinoma: a preliminary case-control study. Eur Radiol. 2013;23:156–62.
45. Nassif A, Jia J, Keiser M, et al. Visualization of hepatic uptake transporter function in healthy subjects by using gadoxetic acid-enhanced MR imaging. Radiology. 2012;264:741–50.
46. Nava A, Mazza E, Furrer M, et al. In vivo mechanical characterization of human liver. Med Image Anal. 2008;12:203–16.
47. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med. 2007;356:2457–71.
48. O'Brien MJ, Keating NM, Elderiny S, et al. An assessment of digital image analysis to measure fibrosis in liver biopsy specimens of patients with chronic hepatitis C. Am J Clin Pathol. 2000;114:712–8.
49. Okubo H, Mogami M, Ozaki Y, et al. Liver function test by gadolinium-ethoxybenzyl-diethylenetriaminepentaacetic acid-enhanced magnetic resonance imaging with consideration of intrahepatic regional differences. Hepatogastroenterology. 2013;60:1547–51.
50. Pang JX, Zimmer S, Niu S, et al. Liver stiffness by transient elastography predicts liver-related complications and mortality in patients with chronic liver disease. PLoS One. 2014;9:e95776.
51. Park SH, Lee JH, Lee SS, et al. CT colonography for detection and characterisation of synchronous proximal colonic lesions in patients with stenosing colorectal cancer. Gut. 2012;61:1716–22.
52. Parsons MW, Pepper EM, Chan V, et al. Perfusion computed tomography: prediction of final infarct extent and stroke outcome. Ann Neurol. 2005;58(5):672–9.
53. Pickhardt PJ, Nugent PA, Mysliwiec PA, Choi JR, Schindler WR. Location of adenomas missed by optical colonoscopy. Ann Intern Med. 2004;141(5):352–9.
54. Pooler BD, Kim DH, Weiss JM, et al. Colorectal polyps missed with optical colonoscopy despite previous detection and localization with CT colonography. Radiology. 2016;278(2):422–9.
55. Poynard T, Ingiliz P, Elkrief L, et al. Concordance in a world without a gold standard: a new non-invasive methodology for improving accuracy of fibrosis markers. PLoS One. 2008;3:1–8.
56. Poynard T, De Ledinghen V, Zarski JP, et al. Relative performances of FibroTest, Fibroscan, and biopsy for the assessment of the stage of liver fibrosis in patients with chronic hepatitis C: a step toward the truth in the absence of a gold standard. J Hepatol. 2012;56:541–8.
57. Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol. 2002;97:2614–8.
58. Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology. 1997;112:24–8.
59. Ronot M, Lambert S, Elkrief L, et al. Assessment of portal hypertension and high-risk oesophageal varices with liver and spleen three-dimensional multifrequency MR elastography in liver cirrhosis. Eur Radiol. 2014;24:1394–402.
60. Selby JV, Friedman GD, Quesenberry Jr CP, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med. 1992;326:653–7.

61. Shaukat A, Oancea C, Bond JH, et al. Variation in detection of adenomas and polyps by colonoscopy and change over time with a performance improvement program. Clin Gastroenterol Hepatol. 2009;7:1335–40.
62. Shin SU, Lee JM, Yu MH, et al. Prediction of esophageal varices in patients with cirrhosis: usefulness of three-dimensional MR elastography with echo-planar imaging technique. Radiology. 2014;272:143–53.
63. Singh S, Venkatesh SK, Wang Z, et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: a systematic review and meta-analysis of individual participant data. Clin Gastroenterol Hepatol. 2015;13:440–51.
64. Singh S, Fujii LL, Murad MH, et al. Liver stiffness is associated with risk of decompensation, liver cancer, and death in patients with chronic liver diseases: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2013;11:1573–84.
65. So A, Hsieh J, Li JY, et al. Quantitative myocardial perfusion measurement using CT perfusion: a validation study in a porcine model of reperfused acute myocardial infarction. Int J Cardiovasc Imaging. 2012;28:1237–48.
66. Stryker SJ, Wolff BG, Culp CE, et al. Natural history of untreated colonic polyps. Gastroenterology. 1987;93:1009–13.
67. Sun HY, Lee JM, Han JK, et al. Usefulness of MR elastography for predicting esophageal varices in cirrhotic patients. J Magn Reson Imaging. 2014;39:559–66.
68. Teng TY, Khor SN, Kailasam M, et al. Morning colonoscopies are associated with improved adenoma detection rates. Surg Endosc. 2016;30(5):1796–803.
69. Verloh N, Haimerl M, Zeman F, et al. Assessing liver function by liver enhancement during the hepatobiliary phase with Gd-EOB-DTPA-enhanced MRI at 3 Tesla. Eur Radiol. 2014;24:1013–9.
70. Waikar SS, Betensky RA, Emerson SC, et al. Imperfect gold standards for kidney injury biomarker evaluation. J Am Soc Nephrol. 2012;23:13–21.
71. Walter SD, MacAskill P, Lord SJ, et al. Effect of dependent errors in the assessment of diagnostic or screening test accuracy when the reference standard is imperfect. Stat Med. 2012;31(11–12):1129–38.
72. Ward ST, Mohammed MA, Walt R, et al. An analysis of the learning curve to achieve competency at colonoscopy using the JETS database. Gut. 2014;63:1746–54.
73. Yeung TP, Wang Y, He W, Project of Emilia-Romagna Region on Neuro-Oncology Study Group, et al. Survival prediction in high-grade gliomas using CT perfusion imaging. J Neurooncol. 2015;123:93–102.
74. Yoon JH, Lee JM, Paek M, et al. Quantitative assessment of hepatic function: modified look-locker inversion recovery (MOLLI) sequence for T1 mapping on Gd-EOB-DTPA-enhanced liver MR imaging. Eur Radiol. 2016;26(6):1775–82.
75. Yoshie T, Ueda T, Takada T, et al. Prediction of cerebral hyperperfusion syndrome after carotid artery stenting by CT perfusion imaging with acetazolamide challenge. Neuroradiology. 2016;58(3):253–9.
76. Zussman BM, Boghosian G, Gorniak RJ, et al. The relative effect of vendor variability in CT perfusion results: a method comparison study. AJR Am J Roentgenol. 2011;197:468–73.

13 Imaging Biobanks, Big Data, and Population-Based Imaging Biomarkers

Annalisa Mantarro, Paola Scalise, and Emanuele Neri

13.1 Introduction

The modern medicine is constantly looking for the link between genetics, etiology, clinical manifestations, and treatment of diseases. In this regard, the anatomical and physiological interindividual differences and similarities, the predisposition to disease development, and its responsiveness to treatment depend widely from detailed information enclosed in the genoma [1].

The advent of biomedical research promoted the creation of an increasing number of facilities for long-term storage and retrieval of human cell and tissue samples. These biorepositories are known with the term of "biobanks," which represent organized collections of biological samples (usually of human origin), centrally stored for one or more research purposes [2]. Human biobanks include biological material of healthy subjects and/or patients with specific pathologies (disease oriented), of which the most frequent are cancer related. However, there is some confusion about the meaning of this term: some definitions are general, including all facility types for biological sample collection, while others are specific, comprehending strictly human sample collections [3]. Therefore, a clear definition of the term is an important step toward fostering the collaboration among researchers, allowing easy access to potential sample sources [4].

The history of biobanks starts with the pathology collections on the eighteenth–nineteenth century. During the second half of the twentieth century, the biomedical research was promoted in the United States and then in Europe with numerous collections of human samples for research purposes. In the recent decade, several international initiatives have emerged in order to promote and coordinate all existing and new biobanks and to develop standardized protocols and metrics [5]. The main goal of these initiatives is the implementation of infrastructural projects, aimed to improve the biomedical research by encouraging generic interoperability [6, 7].

The Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) was one of the first European Research Infrastructure projects funded by the European Commission in January 2011.

Actually, BBMRI is the largest organization of biobanks and biospecimen collections worldwide, including a 53-member consortium with over 280 associated organizations from over 30 countries.

A. Mantarro, MD • P. Scalise, MD • E. Neri, MD (✉)
Diagnostic and Interventional Radiology – Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
e-mail: annalisamantarro@gmail.com; scalisepl@gmail.com; emanuele.neri@med.unipi.it; emanueleneri1@gmail.com; e.neri@ao-pisa.toscana.it

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_13

BBMRI is implemented under the European Research Infrastructure Consortium (ERIC) legal entity and its headquarters, located in Graz (Austria), and is responsible for the coordination of national activities in all participating countries [8].

Nevertheless, all data deriving from radiological imaging were not included in such biobanks; only recently, several projects have been started up for creating large repositories of image data, called "imaging biobanks" [9]. In this context, the registration of all imaging biobanks is essential, as well as the definition of structured approach for imaging data storage and retrieval. The latest goal is the research of a connection between the imaging and tissue biobanks, providing a deep association between the phenotype and genotype, by means of possible imaging biomarkers [2].

13.2 Radiomics and Personalized Care

The recent advent of high-throughput techniques for molecular analysis, including genomics, transcriptomics, proteomics, metabolomics, and imaging techniques, has allowed the storage of a large collection of data for identifying biomarkers used in the disease stratification, prediction, and early diagnosis of diseases [10, 11].

In this regard, the "radiomics" can be defined as the science that deals with the high-throughput extraction, storage, and analysis of a large amount of quantitative imaging features (*imaging biomarker*) in order to create accessible databases from radiological images and to reveal quantitative predictive or prognostic associations between images and medical outcomes [12, 13]. The modern and multiparametric imaging, characterized by digital and quantifiable informations, provides a set of biomarkers of the same patient that allow us to quantify the information. These biomarkers may refer to the organ function or neoplastic mass characteristics, and they are expressed by a number.

Examples of biomarkers are the diameter, volume, computed tomography (CT) density measurement, magnetic resonance (MR) signal intensity, standard uptake value (SUV) in positron emission tomography (PET) imaging, contrast enhancement (valuated in MR, CT, or ultrasound examination), perfusion parameters (i.e., blood flow, blood volume, mean transit time, and permeability), tissue elasticity in elastosonography, tissue pattern (texture analysis), morphological pattern, and much more. Moreover, beyond radiology, other types of images can be collected, for example, from endoscopy, microscopy, and surgery, providing measurable personalized data. Each of these biomarkers is patient specific and will be stored, analyzed, and correlated as part of a cluster of biomarkers of that patient [14].

The main focus of imaging biobanks is the "personalized medicine," where the treatment is increasingly tailored on the basis of specific characteristics of the patient and their disease [15]. Quantitative medical imaging, with the identification of imaging biomarkers, represents a crucial part of personalized medicine providing selection criteria and follow-up strategies, tailored to the patient's needs [16]. All these imaging informations should be considered as the phenotypic expression of a patient and can be correlated to the genotype. In this setting, the radiogenomics, which is the extension of radiomics, aims to identify a link between genotype and phenotype imaging [17].

13.3 Imaging Biobanks: Current Status

The imaging biobanks are wide data collection including medical images and their correlated imaging biomarkers. The content of these biobanks, linked to that of biorepositories, should be available in a shared workflow among all researchers. A European network of imaging biobanks could significantly enhance the validation of new imaging biomarkers that could be potentially used as prognostic and predictive descriptors in the clinical practice.

In March 2014 the European Society of Radiology (ESR) instituted a dedicated working group (ESR WG on imaging biobanks) aimed at

monitoring all existing imaging biobanks in Europe, promoting the federation and communication among them in a white paper [2]. Furthermore, the ESR Working Group promoted the realization of imaging biobanks and techniques for the analysis and processing of imaging biomarkers, stimulating the integration of existing image data repositories and also the link between the imaging biobanks and traditional biobanks, as well as encouraging the researcher cooperation for the standardization, validation and benchmarking of all data stored. The development of imaging biobanks is focused on imaging data collection and sharing for clinical research programs (i.e., clinical trials). By the definition and validation of new biomarkers, the imaging biobanks meet the need for storage, diffusion, and comparison of disease-specific data [14]. In this context, the international research collaboration promotes the comparing of imaging tools, protocols, data, and expertise, in order to establish common acquisition protocols and to ensure high image quality. These data collections could be based on the regional/national screening programs or clinical trials (i.e., performed for colorectal cancer, breast cancer, or lung cancer). The oncologic imaging represents the most suitable field for the discovery and validation of new biomarkers from multiple imaging modalities, since the oncologic patients are frequently monitored for staging and follow-up of treatment response. Nowadays, there is a significant need of detailed and accurate biomarkers, in order to reduce cancer morbidity and mortality, promoting the progression of the traditional "one size fits all" strategy toward a new "personalized" cancer therapy [18].

13.4 Imaging Data Standardization

Accordingly to the dissemination and implementation of imaging biobank, the imaging collection and storage standardization are needed. The development of data standards promotes the communication among all the biobanks, using a standardized format, in order to integrate and share suitable informations for all researchers, as well as to provide legal regulation in the institutions [19]. Nowadays, this cooperation among all researchers about imaging biobanks is very poor. On these bases, the main focus of imaging biobanks is the endorsement of high-quality standard levels, yielding harmonized datasets for biomarker extraction thus reducing the intervariability [20]. All imaging researchers should cooperate to improve and standardize the image acquisition protocols and archiving, the softwares for data analysis and processing, and further methodologies for imaging biobanks.

The current standardization efforts promote the spreading of new techniques for medical image acquisition, visualization, storage, and sharing. In this setting, the Picture Archiving and Communication System (PACS) is a medical imaging technology, which provides the storage and access of digital imaging datasets deriving from multiple modalities through a network connection. The universal format for PACS image storage and transfer is the Digital Imaging and Communications in Medicine (DICOM) that encourages the interoperability between various systems of the healthcare institutions [21]. Moreover, this collaboration between healthcare enterprise professionals and industries resulted in the Integrating the Healthcare Enterprise (IHE), an international initiative, which defines how existing standards should be used (integration profile) [22]. The Cross-Enterprise Document Sharing (XDS) integration delineates the guidelines for sharing documents among all healthcare institutions, promoting the connection between the imaging centers and imaging biobanks [23].

Finally, the radiology report communicates all informations to the patient and referring physicians. For these purposes, it should be uniform, comprehensive, and easily understood. In 2008, the Radiological Society of North America (RSNA) promoted the dissemination of "structured" report templates, consisting of clinical data, coded terminology (e.g., SNOMED, RadLex), technical parameters, measurements, annotations, and key images [24]. Furthermore, the IHE profile (Management of Radiology Report Templates) defines the appropriate use of

templates, resulting in a wider diffusion of best practices and an improvement of radiology communication quality [25].

13.5 Ethical Issues

The recent biobank development gives rise to some issues about the big data management, in order to guarantee the respect of donors' privacy. In this context, the ethical aspects as well as legal and security issues represent crucial steps in biobank building and diffusion [2]. The ethical framework comprehends various steps, such as informed consent, donors' privacy, data protection, and sharing across the biobanks, thus sustaining of public trust [26–28].

- *Informed consent*
  In the first instance, each participant must be informed about how to store, manage, and share own data and samples. On this ground, after the understanding of research purposes, the donor will be asked to sign a very clear consent, legally proving his voluntary authorization to data treatment in an anonymous form. No participant should be exposed to any research risk without their consent. However, there is still no consensus about the type of consent: some authors promoted a standardized consent form, ensuring comparability among all biobanks; while, according to others, the consent should respect geographical, social, and religious diversity [29, 30].
- *Data protection*
  The data protection remains a very challenging legal task in biobanking. Among the European Union (EU) member countries, cross border data sharing and protection has been coordinated by the EU Data Protection Directive, leaving some margins for management by the EU member states [5]. The identity protection of research participants must be respected in the biobank framework [31]. The best way to ensure donors' privacy is the data collection in anonymous form; however, this process destroyed the correlation between the genomic and phenotypic informations. For these reasons, many authors refused permanent anonymization and supported the coding of information as the most appropriate way for data protection. In this setting, the Ethics Review Boards ensure the identity protection of all participants, according to the ethical and legal frameworks and national legislation [29]. The international collaboration in data exchanging should be promoted, in order to minimize the risks for donors [32, 33].

Conclusion

The imaging biobanks represent virtual and accessible databases, focused on the discovery and validation of imaging quantitative biomarkers in order to guarantee a "personalized medicine," which is increasingly tailored on the specific characteristics of the patients and their diseases.

These imaging biomarkers are employed in the early disease diagnosis, follow-up, and response to treatment, creating a link between the phenotype and genotype data (traditional biobanks).

The worldwide dissemination of the imaging biobanks is already ongoing; particularly in Europe, the ESR Working Group established the definition, management, legal and ethical issues, and federation among them.

References

1. Venter JC, et al. The sequence of the human genome. Science. 2001;292(5523):1838.
2. European Society of Radiology (ESR). ESR position paper on imaging biobanks. Insights Imaging. 2015;6(4):403–10.
3. Hewitt R, Watson P. Defining biobank. Biopreserv Biobank. 2013;11(5):309–15.
4. Shaw DM, et al. What is a biobank? differing definitions among biobank stakeholders. Clin Genet. 2014;85(3):223–7.
5. Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) Biobanks and the Public. Governing Biomedical Research Resources in Europe. 2013.
6. van Ommen GJ, et al. BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based expert centres. Eur J Hum Genet. 2015;23(7):893–900.

7. Riegman PH, et al. The organization of European Cancer Institute Pathobiology Working Group and its support of european biobanking infrastructures for translational cancer research. Cancer Epidemiol Biomarkers Prev. 2010;19(4):923–6.
8. Biobanking and BioMolecular Resources Infrastructure. Available at http://bbmri-eric.eu/, Lastly Accessed 6 Dec 2015.
9. Woodbridge M, et al. MRIdb: medical image management for biobank research. J Digit Imaging. 2013;26(5):886–90.
10. Galli J, et al. The biobanking analysis resource catalogue (BARCdb): a new research tool for the analysis of biobank samples. Nucleic Acids Res. 2015;43(Database issue):D1158–62.
11. Lambin P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441–6.
12. Kumar V, et al. QIN "radiomics: the process and the challenges". Magn Reson Imaging. 2012;30(9):1234–48.
13. Bourgier C, et al. Radiomics: definition and clinical development. Cancer Radiother. 2015;19(6–7):532–7.
14. Hsu W, et al. Biomedical imaging informatics in the era of precision medicine: progress, challenges, and opportunities. J Am Med Inform Assoc. 2013;20(6):1010–3.
15. Aerts HJ, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
16. Hewitt RE. Biobanking: the foundation of personalized medicine. Curr Opin Oncol. 2011;23(1):112–9.
17. Goyen M. Radiogenomic imaging-linking diagnostic imaging and molecular diagnostics. World J Radiol. 2014;6(8):519–22.
18. Mordente A, et al. Cancer biomarkers discovery and validation: state of the art, problems and future perspectives. Adv Exp Med Biol. 2015;867:9–26.
19. Quinlan PR, et al. A data standard for sourcing fit-for-purpose biological samples in an integrated virtual network of biobanks. Biopreserv Biobank. 2014;12(3):184–91.
20. Peakman T, Elliott P. Current standards for the storage of human samples in biobanks. Genome Med. 2010;2(10):72.
21. Huang HK. Medical imaging, PACS, and imaging informatics: retrospective. Radiol Phys Technol. 2014;7(1):5–24.
22. Viana-Ferreira C, et al. A framework for integration of heterogeneous medical imaging networks. Open Med Inform J. 2014;8:20–32.
23. Fatehi M, et al. Data standards in tele-radiology. Acta Inform Med. 2015;23(3):165–8.
24. Bosmans JM, et al. Structured reporting: a fusion reactor hungry for fuel. Insights Imaging. 2015;6:129–32.
25. IHE Radiology Technical Committee. IHE Radiology Technical Framework V13.0. Supplement. Management of Radiology Report Templates (MRRT). IHE International; 2014. Available at http://www.ihe.net/uploadedFiles/Documents/Radiology/IHE_RAD_Suppl_MRRT.pdf.
26. Hansson MG. The need to downregulate: a minimal ethical framework for biobank research. Methods Mol Biol. 2011;675:39–59.
27. D'Abramo F, et al. Research participants' perceptions and views on consent for biobank research: a review of empirical data and ethical analysis. BMC Med Ethics. 2015;16:60.
28. Pullman D, et al. Personal privacy, public benefits, and biobanks: a conjoint analysis of policy priorities and public perceptions. Genet Med. 2012;14(2):229–35.
29. Budimir D, et al. Ethical aspects of human biobanks: a systematic review. Croat Med J. 2011;52(3):262–79.
30. Artizzu F. The informed consent aftermath of the genetic revolution. An Italian example of implementation. Med Health Care Philos. 2008;11(2):181–90.
31. Schulte in den Baumen T, et al. Data protection in biobanks – a european challenge for the long-term sustainability of biobanking. Rev Derecho Genoma Hum. 2009;(31):13–25.
32. Zika E, et al. Sample, data use and protection in biobanking in Europe: legal issues. Pharmacogenomics. 2008;9(6):773–81.
33. Mee B, et al. Development and progress of Ireland's biobank network: ethical, legal, and social implications (ELSI), standardized documentation, sample and data release, and international perspective. Biopreserv Biobank. 2013;11(1):3–11.

A Proposed Imaging Biomarkers Analysis Platform Architecture for Integration in Clinics

14

Angel Alberich-Bayarri, Enrique Ruiz Martínez, Rafael Hernández Navarro, José Tomás Cucarella, and Fabio García Castro

14.1 Introduction

Imaging biomarkers play a fundamental role in the new era of precision medicine and are driving the current change of paradigm in radiology, from the traditional qualitative inspection of the images by the naked eye to an improvement in the diagnosis and treatment follow-up from a quantitative perspective.

Nevertheless, despite the growing tendency in research and development in the so-called era of radiomics, a winner solution for fully integrated quantitative radiology workflows has still not been achieved in clinical routine. In fact, although many different research groups worldwide have been implementing new image processing algorithms and have discovered new potential biomarkers that can be extracted from images, there exists a gap between these research results and their applications in routine radiology, since a paradigm change towards quantitative reporting in radiology has still not occurred. The main limitations are related to the technical knowledge required by image processing tools, the lack of specific software solutions, the integration of applications within the hospital information systems, and the funding required for the incorporation of these solutions or even biomedical engineers into radiology departments [4]. More efforts toward integration into clinical routines and radiological workflows need to be addressed to expand the application of imaging biomarkers in clinical validation studies and clinical trials not only for the diagnosis but also for the assessment of therapy response.

International societies have realized about this challenge, and several alliances within the most important radiological societies (Radiological Society of North America, RSNA; European Society of Radiology, ESR) have been created (see Chap. 1 on international alliances to support imaging biomarkers development). Some of these initiatives are focused on the standardization of the imaging biomarkers development workflow, which methodologically can be structured in image acquisition, image preparation, image analysis, and measurements [1].

Due to data privacy and security aspects, hospital networks and systems are highly standardized and regulated; therefore, any new technology has to be universal, decentralized, adapted to

A. Alberich-Bayarri (✉)
Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain

QUIBIM SL, Valencia, Spain
e-mail: alberich_ang@gva.es

E.R. Martínez • J.T. Cucarella
Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain

R.H. Navarro • F.G. Castro
QUIBIM SL, Valencia, Spain

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_14

communication standards, and highly efficient to avoid network saturation. An ideal imaging biomarkers platform should work in the back end, almost transparent to the radiological workflow, with well-defined automated analysis pipelines, and generate a final structured report containing the most relevant results. For minimal interactivity, the platform interface should be available over the network from any hospital computer or workstation, avoiding installation requirements, handling data in a synchronized way, maintaining coherency and consistency, and managing medical imaging studies through user control policies while enabling direct integration with current hospital information systems [4].

Up to now, most manufacturers of imaging equipment have traditionally developed post-processing workstations that require the radiologist to leave the picture archiving and communications system (PACS) user interface and launch specific post-processing tools that allow the analysis of basic imaging biomarkers with a certain interaction. Frequently, these analysis tools are only available in specific workstations of the department and are vendor specific, and therefore their use interrupts the radiology reporting workflow, even sometimes physically, forcing the radiologists to displace from their workplace.

Nowadays, there is an emerging research field in developing new image analysis algorithms and biomarkers and integrating them in analysis platforms. Among other solutions, probably the most popular are the ImageJ platform [6] that allows for the installation of new image analysis pipelines and the XNAT platform that, besides medical imaging projects management, allows the integration of neuroimaging analysis pipelines [5]. Most of these platforms, however, lack from easy integration in clinical routine, requiring advanced skills in informatics and standards connectivity either for its installation or use.

There is also a growing number of medical imaging companies that decide to allocate their platforms in the cloud oriented to provide worldwide service. However, most of these solutions are focused in providing PACS services, that is, storage, visualization, and basic post-processing, but are not centered in handling with imaging biomarker analysis algorithms and the exploitation of quantitative information. This field has however a strong regulatory dependency, since the data protection legislation is highly heterogeneous across countries and cloud providers have a spread distribution of their servers.

Some of the existing imaging biomarkers analysis solutions have achieved regulatory clearance (e.g., Food and Drug Administration – FDA or CE Marking), and some others have been developed mainly for research purposes. Imaging biomarkers analysis tools are considered a medical device class IIa (Europe) and a class II (FDA), since they are a software that is used for measuring certain biological properties that can be used for diagnostic purposes. Typically, the gap between research and clinical application of imaging biomarkers can be justified also by the difficulty in establishing clinical validation projects for imaging biomarkers with a large number of patients. Therefore, an architecture to allow the integration of biomarkers, not only clinically validated but also in research mode, would allow the fast-track certification of algorithms and increase the number of solutions that can be used in clinics.

In the present chapter, an architecture and the components to implement an imaging biomarkers platform fully integrated in a hospital radiological environment is presented.

14.2 Current Solutions

14.2.1 Image Processing Workstations

Although image processing workstations are being used in many hospitals today, they have several drawbacks when compared to emerging solutions like server–client or cloud-based platforms. These workstations generally consist of a computer with high performance capabilities with a stand-alone installation of the processing software solutions. The tools on these workstations usually contain a patient list and a worklist in order to manage incoming studies. The main computer of these solutions

hosts a Digital Imaging and Communications in Medicine (DICOM) node that is connected to the hospital PACS. Radiologists send the studies that require a certain post-processing to the node of the workstation. After the image transmission, the radiologists can move to the workstation, perform the specific processing, capture results, and send them back to the PACS as DICOM images.

As it can be appreciated, this workflow clearly disturbs the radiology process, forcing the radiologist to perform the image analysis in a different computer to the one used for conventional visualization and reporting. Furthermore, the assessment of quantitative imaging biomarkers results is not oriented to allow for future scientific exploitation, since these solutions do not include a database to track and extract conclusions from all the quantitative data. Last but not least, the vendor neutrality of these solutions is not guaranteed, and it is frequent to have different workstations in radiology departments that have different technology providers (e.g., different providers for MR and CT).

14.2.2 Server–Client Platforms

As an evolving step, workstations are giving way to server–client solutions, where a central server is installed in the hospital data-processing center (DPC) and a desktop client application is installed in every computer of the department to be used by the radiologists. The images to be analyzed are sent to the DICOM node of the server, and then the user can perform different analysis from the desktop application. The most important advantage of this solution is that it is not forcing to stop the routine reporting process, since the application can be opened from the same PACS computer of the radiologist.

14.2.3 Service-Oriented Solutions

Some companies nowadays have been focusing in providing a service rather than a software platform for imaging biomarkers analysis. By this approach, the customer sends the data to the company through standard transfer procedures, either physical (CD, DVD, or hard disk) or by secure file transfer protocol (SFTP).

14.2.4 Modality-Embedded Solutions

The development of medical imaging devices has been evolving in parallel to the advance of image processing solutions and imaging biomarkers algorithms. However, current MR, CT, PET, and other modalities still lack from advanced image processing solutions embedded in their operative systems to directly provide quantitative imaging data in their output. This can be considered a consequence of the different business units that hardware and software areas constitute in most medical imaging multinational companies. This strategy has the origin in the workstations market, where specific departments were responsible in bringing new image processing software to workstations installed in radiology departments exclusively dedicated to image analysis. Today, however, several algorithms could be directly integrated in the modalities in order to extract quantitative information as the images are acquired, and technicians should be the professionals responsible on interacting with the software when required, for example, for regions of interest (ROI) delineation.

14.2.5 Cloud-Based Solutions

There is an emerging market in the combination of cloud technologies with medical imaging platforms. This market is predicted to grow at a compounded annual rate of 27% through 2018 [3]. Although a strong progress has been made in the introduction of this technology, there is still a limitation for the legal aspects of image sharing and storage in servers allocated in foreign countries. Images can be stored in the cloud with no anonymization under secure https protocol if a responsible in health data treatment exists in the research group or company working with medical images. This employment profile is audited frequently by

the data protection agency, and it has a high economic cost. A best approach consists of the encryption of personal data from the images following Health Insurance Portability and Accountability Act (HIPAA) recommendations [2]. For that purpose, a hash key introduced by the user is used for encrypting the personal information contained in images metadata like patient name, patient ID, and study ID, among others. Face images are also considered as traceable personal data; therefore, in brain studies the face removal should be mandatory to avoid identification of patients in case of a 3D skin surface reconstruction.

14.3 The Requirements

The experience of the authors in imaging biomarkers integration in clinics has been synthesized in a model that guarantees a successful integration of a quantitative paradigm in the radiological workflows:

14.3.1 Modular

The architecture must be organized in different blocks and layers able to work as connected entities. As an example of modules, we can find DICOM node, report generator, biomarker algorithm, database, and back-end and front-end user interface.

14.3.2 Integrated

The solution must be adapted to current healthcare information systems, including the capability of DICOM communication with PACS and XML data management, and able to understand HL7 messaging from electronics health records (EHR) and health information system (HIS).

14.3.3 Scalable

The platform must be rapidly sizeable in terms of the number of users and computational demand at a minimum economic cost. For that, elastic architectures must be achieved, being able to wake up new servers when an increase in the analysis demand exists.

14.3.4 Pipelines

The platform must be pipeline oriented, differentiating image preparation, image analysis and results measurement, and extraction steps. The pipeline architecture allows for the fast identification of potential errors during the analysis of imaging biomarkers.

14.3.5 Data Mining

The platform must allow also for data exploitation for scientific purposes and be able to cover the need of current millions of Excel spreadsheets managed by researchers in radiology and medical imaging. As an example, a radiologist today is not able to rapidly get the patients of the last 2 years with a K^{trans} value higher than 150 mL/min/100 mL using current PACS and workstations, since hospital databases are not managing this kind of quantitative measurements from medical images. Ideally, the platform should include also a statistics package with the capability of rapidly obtaining graphs, diagrams, and histograms.

14.3.6 Web Based

A web-based interface instead of a desktop application is the best option nowadays either for client-service or cloud architectures, since this means that the platform will be accessible from any place of a hospital by simply using a web browser or from any place with Internet access. Ideally, a zero-footprint DICOM viewer would allow for the visualization of the images and basic tools like regions of interest (ROI) delineation without the need to download a viewer.

14.3.7 Vendor Agnostic

The platform must be able to process the studies with independence of the manufacturer of the equipment used to perform the examination to the patient. Although DICOM is a standard, several relevant information to be used in different analyses are still stored in private tags. For that, the DICOM conformance statements of the most relevant manufacturers must be used to make the platform compatible and vendor agnostic.

14.3.8 Marketplace Strategy

The high modularity of the solution would allow for a plug-in-oriented architecture, creating an internal standard for integrating biomarkers, allowing for other researchers to adapt their algorithms to the platform and expose them in the form of applications (apps) that the user can select and use. In the case of commercialization, an agreement similar to what is used in other marketplaces such as Apple's App Store or Android's Google Play might exist.

14.3.9 Structured Reporting Generation

Last but not least, the platform must provide a quantitative output in the form of structured reports, ideally in one page, with the most relevant information for the user, allowing for a fast visualization and applicability to complement the radiological report or the clinical trial.

In Fig. 14.1, a diagram of the features covered by current architectures available in the market is provided.

14.4 A Proposed Architecture

The ideal solution would consist on a platform covering the features mentioned in the previous section and being able to be installed not only in the cloud but also in local facilities at hospitals. Furthermore, it should be oriented not only to medical doctors as users but also to any researcher in medical imaging quantification and also to clinical research organizations (CROs) managing clinical trials.

For these purposes, a modular architecture must be one of the most important requirements in order to take advantage of as much modules as possible for either cloud or local solutions (see Fig. 14.2).

For this structure, a database was engineered taking into account all entities and relationships present in the imaging biomarkers platform. The database considers the imaging studies, the processes of imaging biomarkers analysis launched, their status (not started, in process, error, finished), the results, the institution and equipment

	Workstations	Server-client platform	Services	Modality embedded	Cloud
Modular	✖	✔	✖	✖	✔
Integrated	✖	✔	✖	✔	✔
Scalable	✖	–	✖	✖	✔
Pipeline oriented	–	✔	✖	✔	✔
Data mining	✖	✔	–	✖	✔
Web-based	✖	✔	✖	✖	✔
Vendor agnostic	✖	–	✔	✖	✔
Marketplace strategy	✖	–	✖	✖	✔
Structured reports	✔	–	✔	✖	✔

✔ Most solutions accomplish
– Variable depending on the solution
✖ Not frequently accomplished

Fig. 14.1 Main features of the different imaging biomarkers platform architectures existing nowadays

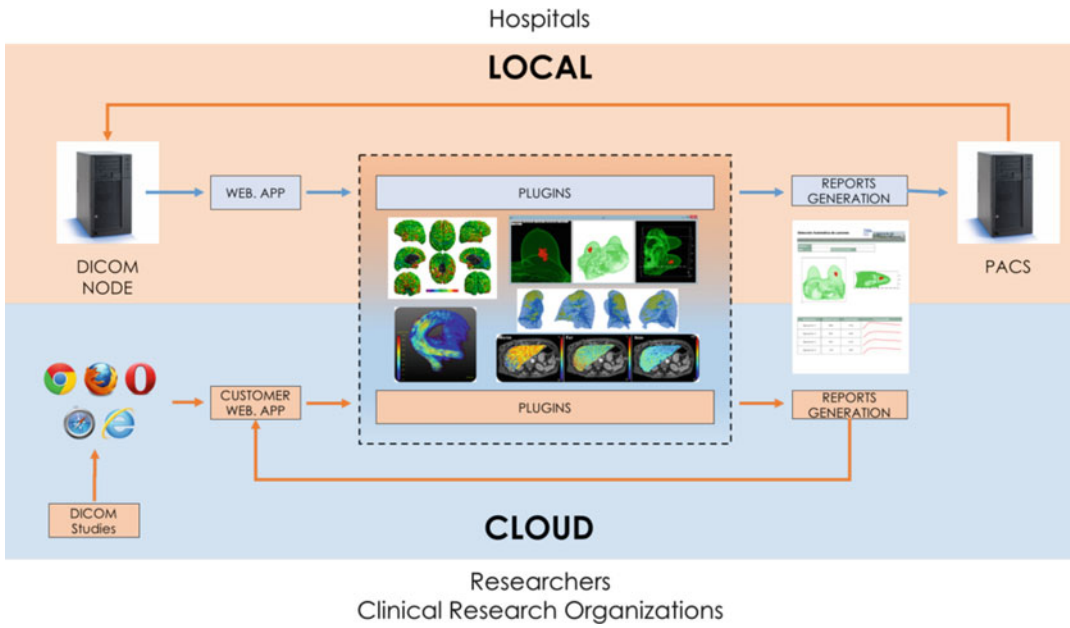


Fig. 14.2 Imaging biomarkers analysis local and cloud platform sharing the algorithms for image analysis for both architectures

of the users, and the reference values for healthy populations. The database structure can be appreciated in Fig. 14.3.

Regarding the local solution to work within the radiological workflow, a detailed diagram of the local solution can be appreciated in Fig. 14.4.

A DICOM node is used to continuously "listen" for any incoming data. It can be configured in the systems of the network by its IP number, AE title, and port number. The node allows for the reception of imaging studies either from the PACS or directly from the equipment. If appropriate rules are set by the PACS provider, it is possible to auto-push studies from the PACS to the imaging biomarkers' DICOM node as they are received, specially for those analysis pipelines that are fully automatic (i.e., brain volumetry, multiple sclerosis lesion segmentation, among others). The node performs two main tasks: stores the imaging data in a standardized way in the local server for imaging biomarkers analysis and introduces a new registry in the platform database. The DICOM node is also responsible for the notifications to the users managing the platform.

The web application is the core of the solution; all the logics in the application is checking for the entry of new studies by queries to the database and deleting the source images of those studies stored for more than 1 month. The application is responsible for launching processes of imaging biomarkers analysis. For that, the imaging biomarker solutions have been implemented by following an internally standardized format, which consists in specifying the route with the source images, the route where the results must be stored, the process ID, and some biomarker configuration settings that the user has selected in the application before pushing the button for the analysis. The application is also checking for the existence of errors in any of the running processes either in Python, MATLAB, or in the system. The interfaces of the web application are built based on the HTML5, CSS3 and JavaScript standards, allowing a rich and friendly usability. For this propose, and considering the characteristics of the product and the user needs, AngularJS has been selected as the most suitable framework. These interfaces create a client-side environment that allow intuitive interactions and fast commu-

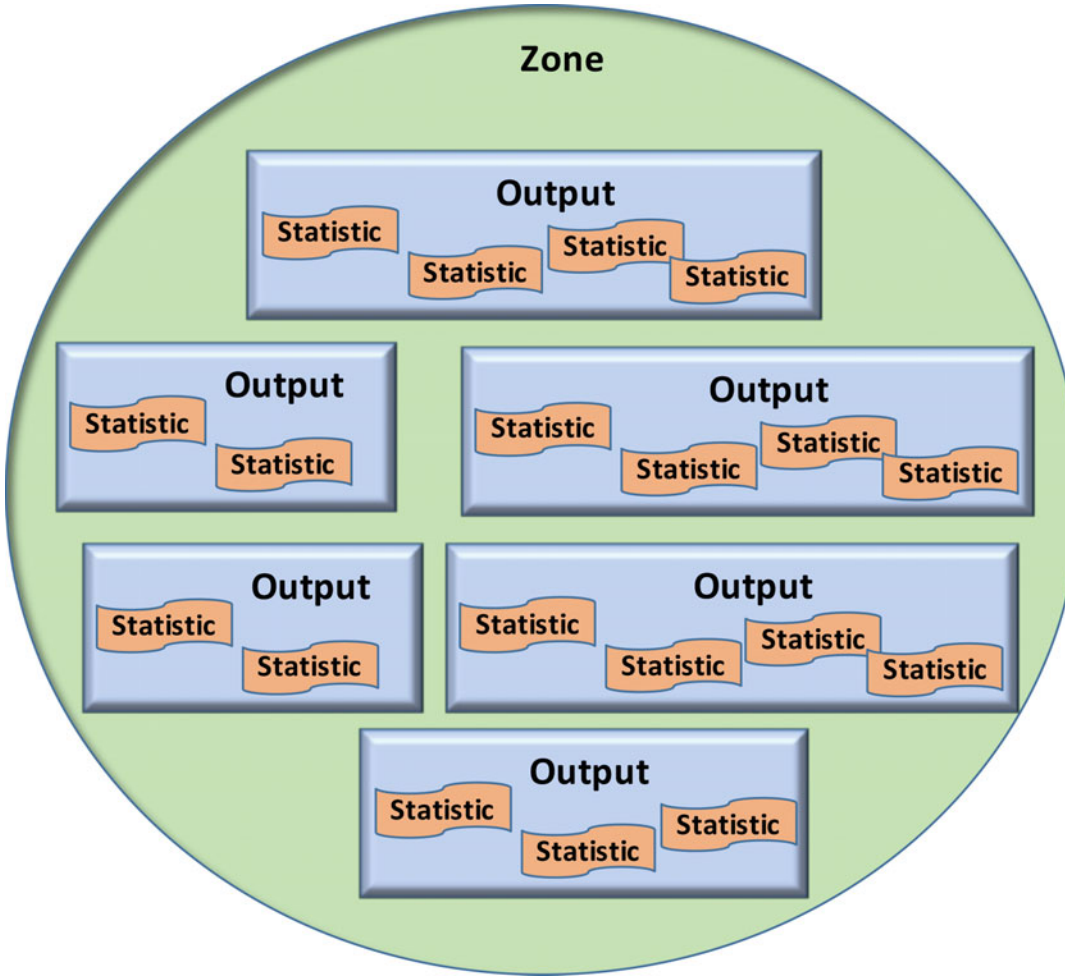


Fig. 14.3 Design of the results hierarchy for the imaging biomarkers analysis platform

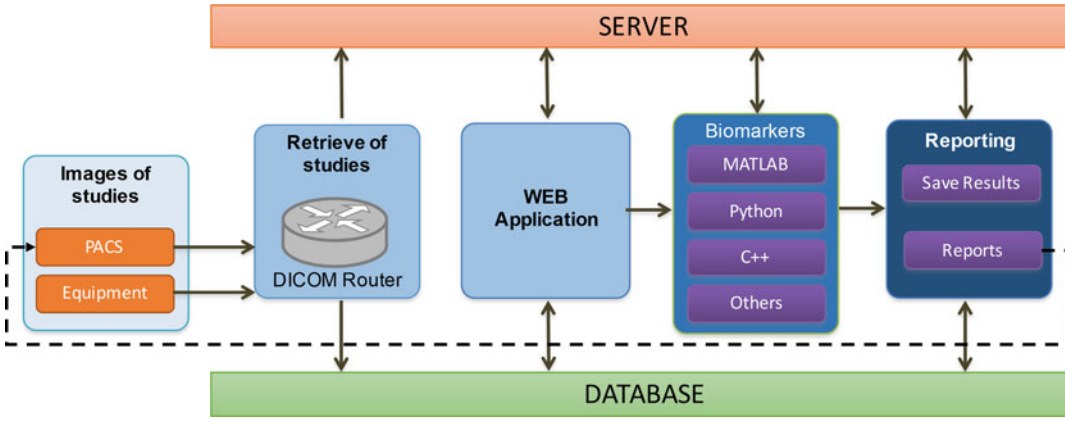


Fig. 14.4 Detailed architecture of the local platform for imaging biomarkers analysis

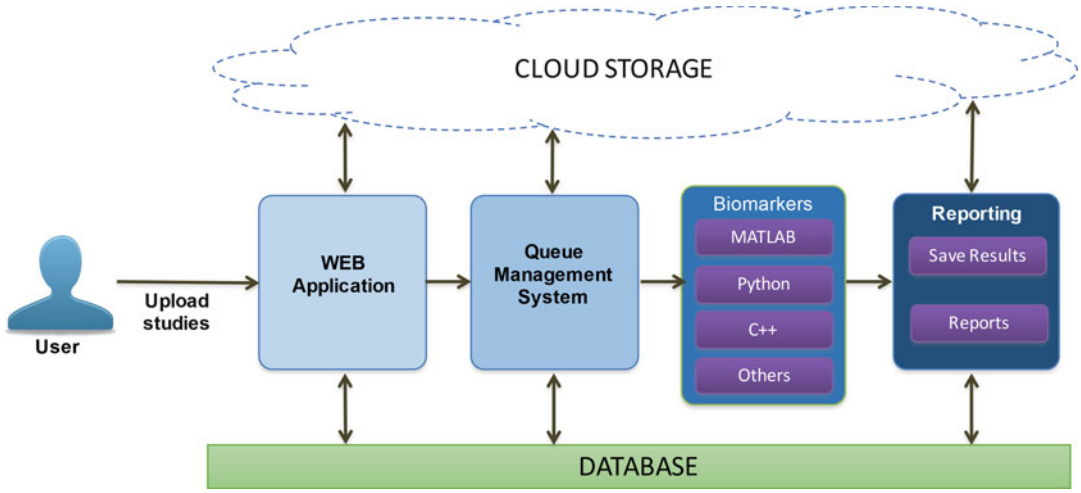


Fig. 14.5 Detailed architecture of the cloud platform for imaging biomarkers analysis

nications with a server-side build using Java. The communication between server (backEnd) and client (frontEnd) is made via API-REST (Representational State Transfer), so a SPA (Single Page Application) can be built allowing to show many different views with a small amount of data browsing through the web via structured JSON objects. This methodology implies that other Web applications are able to use API-REST to connect with the database.

Using these technologies, the client's browser can show the list of studies, details of each study, a graphical representation of DICOMs images via an embedded viewer, web reports of the biomarkers analyses, tabulated values, statistics graph charts of the results for value comparison or research and an increasing number of views to satisfy the user needs.

Once an imaging biomarker analysis pipeline is finished, a Java program is launched. This program has three main functions. First, it stores the results of the process in the database using a XML file created by each biomarker. Second, it generates a report using Jasper Reports technology and a Java support library. This report uses a XML file as its input source, which is generated processing the biomarker. A previously designed template (JRXML) is later filled in with this XML file. Finally, it sends the report to the PACS and any relevant images generated by the post processing. For structured reports implementation, please refer to chapter 12 of this book.

Regarding the cloud solution, a similar architecture was designed, where the web application, biomarkers, and structure reports do not suffer significant variations, but the DICOM node is replaced by a study anonymization and an upload tool, and the local server storage is substituted by the cloud storage. A detail on the architecture can be appreciated in Fig. 14.5.

Although from the general perspective the architecture does not suffer from structural changes, from the processes perspective, a different philosophy exists in the cloud performance. The experience of the authors is focused in Microsoft Azure Cloud environment, where the storage repository does not contain a folder style structure … Also, the imaging biomarkers processes have to enter in queues, since the provider is responsible for the workload activity in the different machines…

A detailed diagram of the cloud solution can be appreciated in Fig. 14.6.

Conclusions

The authors have presented in this chapter the main solutions to bridge the gap between imaging biomarkers analysis research advances and the clinical application. The impact of the lack of current integration

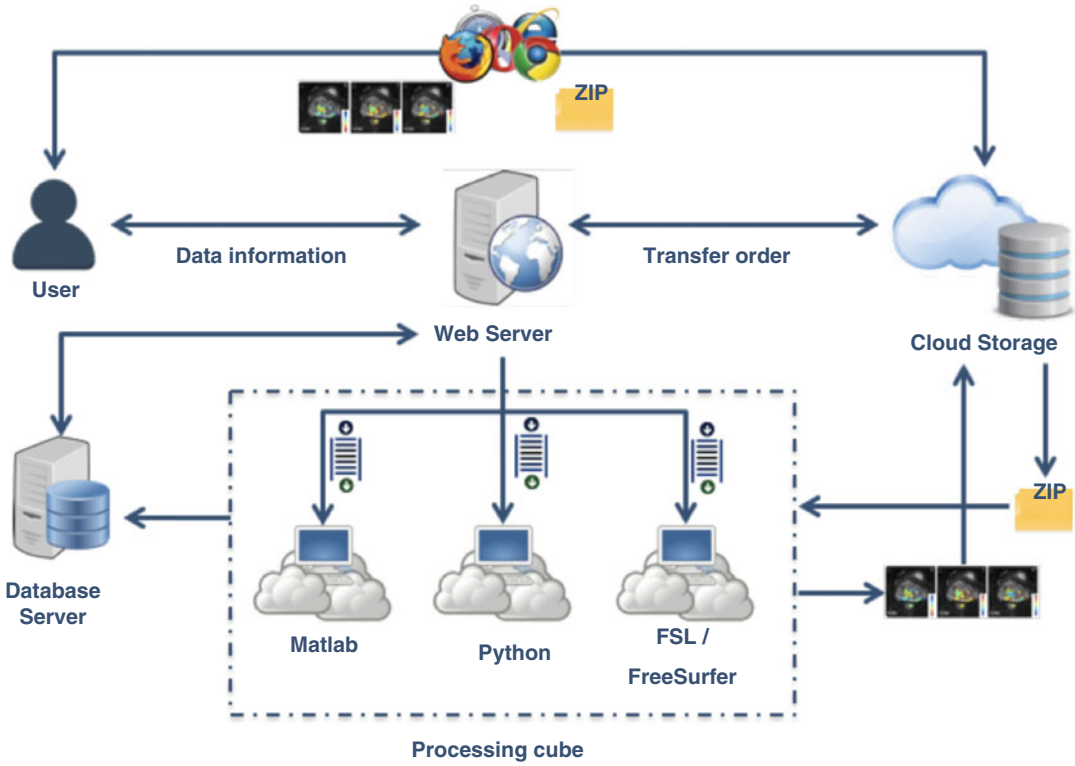


Fig. 14.6 Architecture for the integration of an imaging biomarkers platform with a cloud architecture and image analysis algorithms running in MATLAB (Mathworks Inc, Natick, MA, USA), Python, or third-party image analysis tools like FSL (FMRIB Software Library) or FreeSurfer (Athinoula A. Martinos Center for Biomedical Imaging, Harvard-MIT, Boston)

between systems can be minimized by implementing imaging biomarkers analysis platforms, either in local networks or in the cloud similar to the herein exposed.

References

1. European Society of Radiology (ESR). ESR statement on the stepwise development of imaging biomarkers. Insights Imaging. 2013;4(2):147–52. doi:10.1007/s13244-013-0220-5. Epub 2013 Feb 9.
2. HHS – United States Department of Health and Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. November 2012.
3. Malvey D, Slovensky DJ. mHealth: transforming healthcare. New York: Springer; 2014. ISBN: 978-1-4899-7456-3.
4. Mañas Garcia A, Sanz Requena R, Alberich-Bayarri A, García-Martí G, Aparici Tortajada L, Ruiz Martínez E, Martí-Bonmatí L. Imaging biomarkers processing platform. Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on, vol., no., pp. 416–9, 1–4 June 2014 doi: 10.1109/BHI.2014.6864391.
5. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics. 2007;5:11–34.
6. Rasband WS, Image J. U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997–2015.

15 Use Case I: Imaging Biomarkers in Neurological Disease. Focus on Multiple Sclerosis

Diana M. Sima, Dirk Loeckx, Dirk Smeets, Saurabh Jain, Paul M. Parizel, and Wim Van Hecke

15.1 Introduction

Imaging is widely used for diagnosis and monitoring of neurological diseases. CT scans are routinely acquired in emergency units in patients with traumatic injuries or stroke. PET imaging has gained a strong foothold in oncology. MRI has become the standard of practice for the diagnosis, follow-up and management of numerous neurological and psychiatric conditions. All of these imaging techniques have in common that, in clinical practice, the images need to be interpreted visually by trained specialists, who are responsible for initial diagnosis or for interpretation of follow-up examinations.

Within the scientific literature there is increasing emphasis on the use of quantitative medical imaging biomarkers, i.e. relevant numerical values that can be extracted from 2D or 3D medical image data sets, using advanced image processing techniques. Many imaging biomarkers, such as volumetric assessment of brain structures, have been shown to have excellent sensitivity and specificity for diagnosis or prognosis of various neurological diseases.

In this chapter, we shall focus on the development of relevant MR imaging biomarkers for patients with multiple sclerosis (MS). However, several of the techniques described below can be generalised to other neurological conditions.

15.2 Imaging Biomarkers Relevant to MS

15.2.1 Background

Multiple sclerosis (MS) is a chronic, inflammatory demyelinating disease of the central nervous system (CNS) which is hallmarked by white matter lesions [9]. Since 2001, MRI has been formally incorporated in the diagnostic workup of patients with a clinical suspicion of MS [38]. Recently, MRI has become an important tool for assessing the extent of brain damage, which is used in the monitoring of disease progression and therapeutic efficacy. In current clinical practice, these assessments are based on visual inspection of MR images by expert neurologists and neuroradiologists, who evaluate the presence and distribution of focal white matter lesions. A huge body of research has been devoted to white matter lesions, since they are considered to be a hallmark of the disease (even though abnormalities

D.M. Sima (✉) • D. Loeckx • D. Smeets
S. Jain • W. Van Hecke
icometrix, Leuven, Belgium
e-mail: diana.sima@icometrix.com

P.M. Parizel
Department of Radiology, Antwerp University Hospital, University of Antwer, Wilrijkstraat 10, 2650, Antwerp, Belgium

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_15

also occur in the grey matter). Lesions (also known as "plaques") can be visualised with several MRI sequences:

- *T1-weighted MR images*: chronic stage lesions with axonal destruction and irreversible damage appear as dark spots ("black holes"), compared to the surrounding white matter (WM) tissue intensities (see Fig. 15.1).
- *Gadolinium-enhanced T1-weighted MR images*: "active" lesions taking up contrast material and indicating inflammation and breakdown of the blood–brain barrier; the presence of enhancing lesions indicates ongoing disease activity, since only new lesions (under 6 weeks old) enhance (see Fig. 15.2).
- *T2-weighted MR images*, *fluid attenuated inversion recovery* (*FLAIR*) *MR images and proton density* (*PD*) *MR images*: on these imaging sequences with a long repetition time (TR), lesions appear as hyperintense spots compared to the surrounding brain parenchyma (see Fig. 15.1).

The "lesion load", defined as the total volume of lesions in the brain, is one of the most important biomarkers in MS. Often, a distinction is made between T2 lesions (i.e. lesions that appear hyperintense on T2-weighted or FLAIR images), T1 lesions (i.e. lesions that appear hypointense on T1-weighted images, the so-called black holes), as well as contrast-enhancing lesions.

In addition to the lesion load, brain volumetry [25] and, more specifically, cerebral atrophy [5] and, in particular, grey matter (GM) atrophy [23] are currently considered to be essential biomarkers, since they are correlated with the speed of disease progression (see Fig. 15.3). Thus, apart from the detection of lesions, quantification of brain volumes and atrophy rates is increasingly important in the management of patients with MS.

15.2.2 Natural Course of the Disease

From a clinical point of view, MS starts with a first attack or a "clinically isolated syndrome" (CIS) suggestive of MS. CIS patients with a

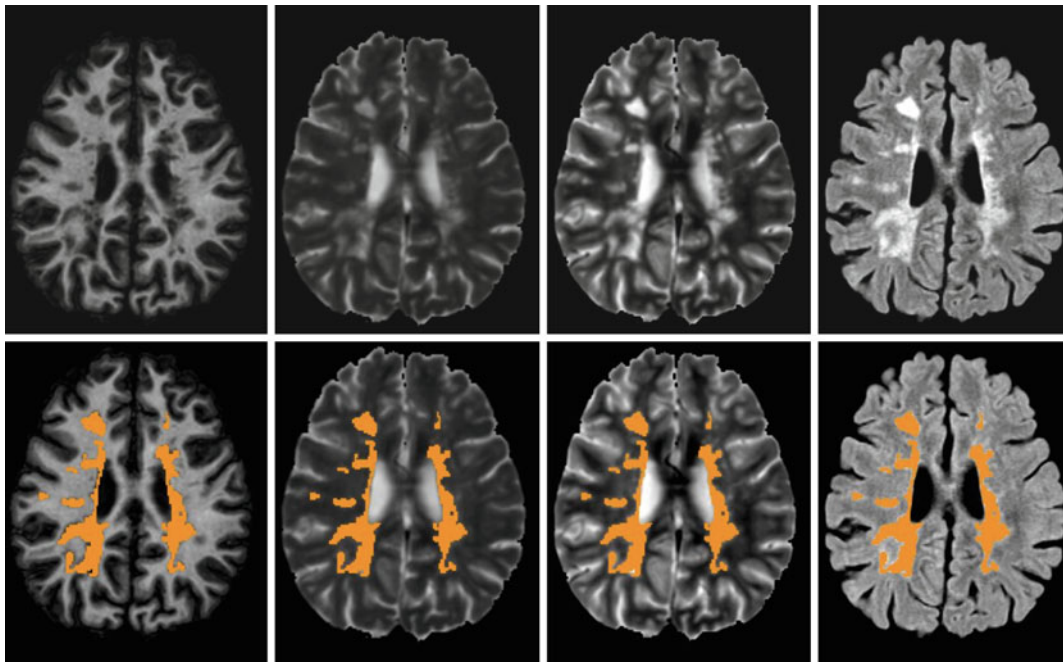


Fig. 15.1 *Top*, from left to right: T1-weighted, T2-weighted, proton density and FLAIR images, obtained in an MS patient. *Bottom*: same images overlaid with expert manual delineation of MS lesions (Data from the "MS longitudinal lesion segmentation challenge", ISBI 2015; training subject 02, time point 01.)

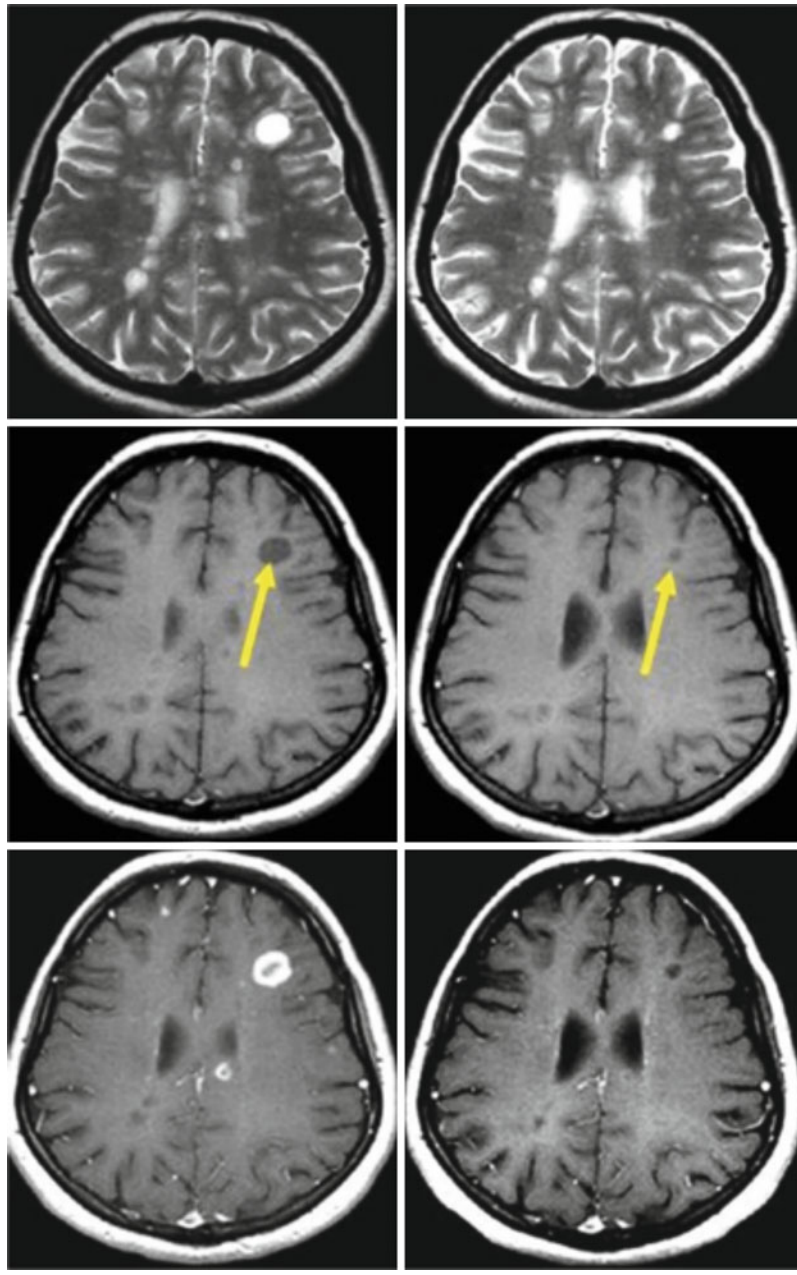


Fig. 15.2 Contrast enhancement in MS (Figure 5 from [45]; original caption: *Serial magnetic resonance imaging* (*MRI*) *scans obtained in a patient with relapsing-remitting multiple sclerosis. T2-weighted* (*upper row*), *unenhanced T1-weighted* (*middle row*) *and contrast-enhanced T1-weighted* (*lower row*) *MRI scans obtained at baseline* (*left*) *and 1 year later* (*right*). *Observe the active 'black hole' in the subcortical white matter of the left frontal lobe* (*arrow*), *which shows a ring-enhancement pattern of contrast uptake. After 1 year*, *the lesion decreased in size* (*arrow*), *but remained hypointense on T1-weighted images*, *indicating an irreversible black hole).* (Adapted from Reprinted by Permission from SAGE Publications, Ltd.: *Ther Adv Neurol Disord.* 6(5):298–310, copyright 2013

normal cerebral MRI at presentation have only a 5 % risk of subsequent clinical attack and thus of progression to clinically definite MS in the next 1–5 years [3]. Conversely, patients with cerebral lesions on MRI have a considerably higher risk, although the risk remains below 50 % when the total lesion volume does not exceed 1.2 ml [3].

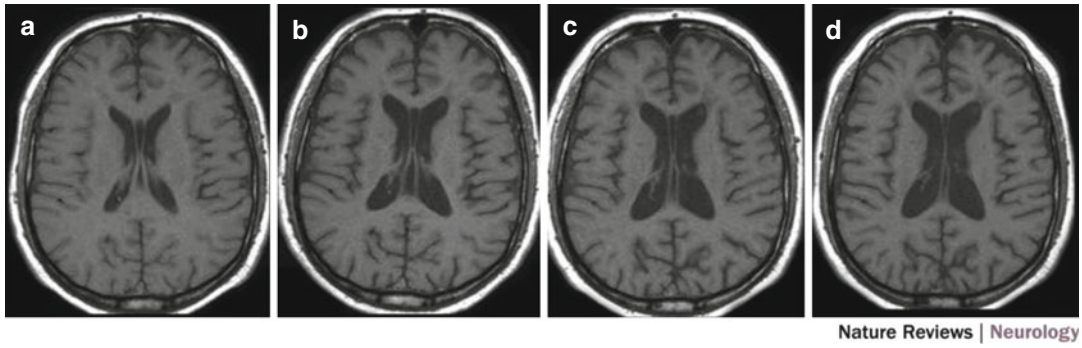


Fig. 15.3 Brain atrophy in an MS patient (Figure 2 from [60] with original caption: (***a***) *Baseline scan.* (***b–d***) *Regular scans over a 6-year follow-up period. Disease progression can be seen in the form of the increasing size of ventricular and subarachnoid spaces. These changes reflect brain volume loss over time, indicating progressive neurodegeneration*). (Adapted from Reprinted by permission from Macmillan Publishers Ltd: *Nature Reviews Neurology* 11, 597–606, copyright 2015

When the disease evolves, it may take one of several forms: (1) relapsing-remitting MS (acute attacks are followed by remission periods), (2) primary progressive MS (steady worsening of neurologic functioning without any distinct relapses) and (3) secondary progressive MS (can follow relapsing-remitting MS and is characterised by a sustained build-up of disability, independent of any relapses). MR studies have confirmed the occurrence of lesions and the development of brain atrophy in all the MS types [14, 24, 29, 40].

Clinical evolution of MS is characterised by both motor and cognitive degradation [7, 41]. Pathological changes in the normal appearing white and grey matter are better correlated with progressive cognitive deficits than with visual, sensory and motor symptoms [4, 24, 51]. Brain atrophy, defined as the decrease in brain volume over time, is recognised as a typical consequence of MS [5]. The rate of brain volume loss in patients with MS exceeds the rate observed in healthy controls up to a factor of 2–8, that is, 0.5–1 % per year in MS patients versus 0.1–0.3 % per year in age-matched healthy controls [24, 48]. Formerly, it was believed that MS was defined pathologically as an inflammatory process confined to the white matter (WM). Nowadays, we know that in addition to white matter lesions, MS is also characterised by grey matter lesions and atrophy [15]. Moreover, MRI-based volumetric data have shown that grey matter atrophy better correlates with physical and cognitive disability than WM atrophy and T1 and T2 lesions [23, 24, 34].

Investigators have examined, and then confirmed, the unwritten rule that five new lesions (compared to the baseline MRI scan of the MS patient) are correlated with a higher risk of subsequent relapses [36]. A 10-year follow-up study on patients with relapsing-remitting MS confirmed the long-term clinical relevance of brain lesion evolution by showing that an accelerated clinical disability is particularly well correlated with the increase in T1 lesions [26]. Annualised lesion volume growth of 0.25 ± 0.5 mL ($+6.7 \pm 8.7$ %) for T2-weighted lesions and $+0.20 \pm 0.31$ mL ($+11.5 \pm 12.3$ %) for T1-weighted lesions was established over a period of 10 years [26]. In a 20-year follow-up of 107 MS patients, lesion growth was 0.80 mL/year in those who were relapsing-remitting but was 2.89 mL/year in secondary progression [18]. In another follow-up study comparing progressive and nonprogressive MS patients over 10 years, it was found that GM atrophy is a good candidate as a disease progression biomarker [30]. In addition, brain atrophy and lesion load have been shown to be correlated with long-term disability in a multicentre 10-year follow-up study [39]. Table 15.1 provides a summary of these investigations.

Table 15.1 Evidence for the relevance of lesion and atrophy biomarkers in MS

Biomarker	Findings	Study population	Reference
T2 lesions	Annualised change: +0.25 ± 0.5 mL	10-year follow-up RRMS	[26]
T1 lesions	Annualised change: +0.20 ± 0.31 mL	Idem	Idem
T2 lesions	Annualised change: +0.80 mL/year	20-year follow-up RRMS	[18]
T2 lesions	Correlated with disability progression	10-year follow-up	[39]
Total brain atrophy	Correlated with disability progression	Idem	Idem
Total brain atrophy	CIS = −0.40 % ± 0.47 %, RR = −0.49 % ± 0.65 %, SP = −0.64 % ± 0.68 %, PP = −0.56 % ± 0.55 %	>1-year follow-up 963 subjects	[12]
GM atrophy	Significant differences between progressive and nonprogressive	10-year follow-up	[30]

15.2.3 Treatment

MS researchers throughout the world acknowledge that, in addition to its well-established diagnostic value, MRI has an essential role in monitoring disease progression and therapeutic efficacy [2, 20]. Recent treatments, especially for the early stages of MS, focus on modifying the natural course of the disease and do not merely treat symptoms. Some of the more aggressive disease-modifying therapies can have serious side effects and are therefore not prescribed as first-line treatments. Criteria for switching from one treatment to another are still under active investigation. MRI-based monitoring of therapeutic effects becomes more and more essential in clinical trials and also in clinical practice. However, MRI-derived metrics are not yet standardised and still under development [2, 20, 27, 44].

In patients developing three or more active MRI lesions, in addition to a clinically active disease (relapses and disability progression), a change in treatment strategy is recommended [44]. During the course of disease-modifying therapy, new or enlarging lesions should be monitored on scans every 6 months to assess for change [43]. The presence of one or more Gd-enhancing lesions on a 6- or 12-month follow-up scan, or two or more new or enlarging lesions on a 12-month follow-up scan, should prompt consideration of therapy change [43].

Many lessons have been learned from clinical trials, for instance, that using lesion count and brain atrophy as endpoints might be more efficient than the Expanded Disability Status Scale (EDSS) [34]. Placebo-controlled trials in secondary progressive MS patients would require 32 subjects per arm to detect a 50 % treatment effect at 80 % power over 2 years, if MRI measures of brain atrophy (using the SIENA method [56]) were used as outcome measures [1]. Using EDSS as outcome measure, placebo-controlled trials would require about 150 patients per study arm to demonstrate statistically significant therapeutic effects for a study of 2–3 years [47].

The debate is ongoing whether whole-brain atrophy should be used as the gold standard for effective treatment of MS after the first year of treatment [48] or not [16]. A confounding factor is that whole-brain atrophy after 1 year of treatment might be an inaccurate parameter, due to the occurrence of pseudo-atrophy: the early reduction of brain volume as a result of a decreased inflammatory profile, rather than of the underlying disease [11]. Measuring GM atrophy, instead of whole-brain atrophy, is potentially more useful since pseudo-atrophy appears to affect white matter more than grey matter [48] and may persist for more than 2 years after treatment initiation.

Some clinical trials have shown that brain volume loss (or GM loss) is a good predictor for the

natural course of the disease [11]. However, when disease-modifying treatments (DMTs) are used, conflicting results have been observed in various clinical trials. Differences in the mechanism of action of the drugs, the patient populations, the quality of the MRI scans and the software packages used for the analysis could explain the occasional discrepancy between these results.

15.3 Acquisition Requirements

Widespread application of MR imaging biomarkers is hampered by issues such as non-standardised imaging protocols, imaging artefacts, lack of normative data and manual segmentations to interpret values in clinical practice. In order to mitigate such issues and promote MR imaging in MS clinical practice, the MAGNIMS study group published guidelines for the use of MRI in MS diagnosis [46], as well as recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal MS studies [59, 60]. The group recommends that "images should be acquired using 3D pulse sequences, with near-isotropic spatial resolution and multiple image contrasts to allow more comprehensive analyses of lesion load and atrophy, across time points. Image artifacts need special attention given their effects on image analysis results" [59]. Image artefacts interfering with MRI readings include radiofrequency (RF) intensity non-uniformity, phase-encode ghosting, signal wrap and geometric distortion due to gradient non-uniformity and B_0 inhomogeneity.

Investigators of the Canadian MRI Consensus Group further specify that "a standardized MRI protocol is important during patient follow-up" [2]. The recommended brain MRI sequences are 3D FLAIR (or axial + sagittal FLAIR), post-gadolinium T1, axial T2 and/or PD, obtained with a minimum MRI field strength of 1.5 T and a slice thickness of 1 mm for T1 and ≤3 mm for FLAIR with no gap; total head coverage should include the entire brain and brainstem.

MSCare, the (US) Consortium of Multiple Sclerosis Centers [8], proposes a standardised MRI protocol that they regularly update (Table 15.2).

Table 15.2 MSCare guidelines for standardised MRI protocol

Standardized brain MRI protocol (diagnosis and routine follow-up of MS)	
Field strength	Scans should be of good quality, with adequate signal-noise ratio (SNR) and resolution (in-sections, pixel resolution of ≤ 1mm × 1mm)
Scan prescription Coverage Section thickness and gap	Use the subcallosal plane to prescribe or reformat axial oblique sections Whole brain coverage ≤3 mm, no gap (for 2D acquisition or 3D reconstruction)
Core sequences	Anatomic 3D inversion recovery-prepared T1 gradient echo (e.g. 1.0–1.5 mm thickness) Gadolinium single dose 0.1 mmol/kg given over 30 seconds[a] 3D sagittal T2-weighted FLAIR[b] (e.g. 1.0 to 1.5 mm thickness) 3D T2-weighted[b] (e.g. 1.0 to 1.5 mm thickness) 2D axial DWI (≤5 mm slices, no gap) 3D FLASH (non IR[c] prep) post-gadolinium[b] (e.g. 1.0 to 1.5 mm thickness) 3D series would be typically reconstructed to 3mm thickness for display and subsequent comparison for lesion counts
Optional sequences	Axial proton attenuation Pre-or post-gadolinium axial T1 spin-echo (for chronic black holes) Susceptibility weighted imaging (SW1) for identification of central vein within T2 lesions

Table 2 in [8]. Reprinted by permission from AMERICAN SOCIETY OF NEURORADIOLOGY: AJNR Am J Neuroradiol 37(3):394–401; copyright 2016

[a]Minimum 5-minute delay before obtaining post-gadolinium T1. The 3D sagittal FLAIR may be acquired immeaiately after contrast injection before the 3D FLASH series

[b]If unable to perform a 3D acquisition, then perform a 2D axial and sagittal FLAIR, axial fast spin-echo proton attentuation/T2, and axial post-gadolinium T1-weighted spin-echo at ≤3mm slice thickness

[c]Inversion recovery

These requirements are sufficient not only for visual assessment but also for automated image analysis software, since most packages performing MRI-based brain segmentation, atrophy computations or lesion segmentation work either on single MR images or on a subset of multiparametric images, simultaneously taken into account.

15.4 Analysis Methods

15.4.1 Cross-Sectional Biomarkers

15.4.1.1 Brain Volume Computations

The volume of the whole brain, or volumes of brain structures, can be easily computed through brain segmentation techniques. Brain segmentation implies that the whole brain, its constituent tissue types or individual brain substructures can be identified based on MRI(s). A typical first step is "brain extraction" or "skull stripping", a preprocessing step that ensures that only brain tissue is transmitted to the segmentation pathway. Various brain extraction methods, such as the brain extraction tool (BET) [55], brain surface extractor (BSE) [50], ROBEX [28], etc., are available. Approaches are diverse, including morphological, geometrical, image processing and modelling operations (hole filling, surface modelling, edge detection, intensity thresholding, atlas matching, deformable models, patch-based labelling, etc.). Moreover, the results of individual brain extraction methods can be enhanced by applying hybrid techniques, thus combining results from several individual methods.

After the first step of brain extraction, the process of brain segmentation can be started. This is typically based on a probabilistic modelling of voxel intensities, exploiting the fact that different tissue types have different MR image characteristics. Recent literature provides an excellent overview of brain segmentation methods [10]. Well-known and validated examples include FAST [61], SIENAX [56] and FreeSurfer [17]. Gaussian mixture models are popular; image intensities for each tissue type are modelled as a (sum of) Gaussian components. This modelling is usually performed using expectation–maximisation (EM), a well-known iterative parameter estimation algorithm. Spatial priors, serving as starting values and also as spatial constraints, can be obtained from appropriate brain atlases available in the literature [42]. The EM framework can be extended to intrinsically model some of the common distortions present in MR images, such as spatial inhomogeneity of image intensities known as bias field. Otherwise, such correction should be performed in preprocessing, e.g. using methods such as N3 or N4ITK [53, 58]. The EM results are probabilistic, i.e. each voxel is assigned probability of belonging to each of the classes of interest (WM, GM, CSF, etc.). These maps can be thresholded to obtain hard segmentations. Volumes in millilitres for each class can be computed either based on the hard or the fuzzy segmentation, by simply multiplying the sum of the tissue segmentation over all voxels by the voxel volume.

15.4.1.2 Lesion Detection and Volume Estimation

Some automatic lesion segmentation methods belong to the family of supervised classification methods, for which a representative training dataset, including expert segmentation, is required in order to build a model that can be used on new patients for lesion segmentation. Depending on the features extracted from images (local gradient intensity, mean intensity, spatial information, etc.) and on the type of classifier (k-nearest neighbours, artificial neural networks, Bayesian learning, support vector machines, etc.), many variants have been proposed ([22, 32, 33, 57]; see also García-Lorenzo et al. [21] and Mortazavi et al. [37] for overviews of algorithms and software solutions). Although excellent results can be obtained with supervised classification on the training dataset, these methods have two disadvantages. The first difficulty lies in building a training dataset that encompasses MS lesions of all possible shapes and intensities and is heterogeneously distributed in the WM. The second nontrivial problem lies in preprocessing a new image (acquired on a different scanner than

the one used for the training dataset), such that it matches the characteristics of the training dataset, e.g. by intensity normalisation. In other words, supervised methods perform well only when the new image to be segmented is well represented in the training dataset.

Another family of methods is based on unsupervised classification and does not require training images. These methods are usually based on stochastic modelling of voxel intensity distribution. They perform brain segmentation into GM, WM and CSF (with or without lesion detection) and often rely on post-processing approaches in order to segment lesions (e.g. lesion growing or pruning). The assumptions that are made in order to segment lesions have a great impact on the results. For instance, LST [49] and MSmetrix [31] detect FLAIR-hyperintense outliers, which are further promoted as lesions according to their spatial probability of being in the WM, where the WM segmentation is basically derived from T1-weighted image segmentation. LesionTOADS [52], on the other hand, employs a sophisticated mechanism of combining information from different MR sequences (T1-weighted, T2, PD or FLAIR) in order to simultaneously segment lesions and brain structures, while distance maps from the boundaries of structures such as CSF are used to confine the segmented lesions to typical locations.

15.4.2 Longitudinal Biomarkers

In contrast to cross-sectional approaches, longitudinal methods take into account two (or more) MRI scans of the same subject from different time points to calculate brain volume changes or atrophy. Typical preprocessing steps prior to longitudinal atrophy computations include [13] extraction of the intracranial cavity mask at baseline, correction of intensity inhomogeneities, rigid registration of follow-up scans on the baseline scan and differential bias field correction to correct for differences in intensity inhomogeneity artefacts.

Longitudinal methods for brain atrophy typically try to match two MRI scans using registration techniques and directly extract small changes in brain volume from this process. Approaches include brain edge motion analysis, voxel-based statistical analysis for voxel-based morphometry, statistical parametric mapping and local Jacobian determinant analysis after nonlinear matching between coregistered images [6, 19, 54, 56].

In what concerns lesions, many methods focus on segmenting MS lesions at a single time point, and there is not yet a single approach, according to the review of Lladó et al. [35], that can emerge as a standard in clinical practice for the analysis of lesion evolution over time.

15.5 How to Transmit the Information to the Clinician

Consistent with current clinical practice, MS patients are referred for an MRI examination by their neurologist. Good communication between neurologist and (neuro)radiologist is of paramount importance. According to recent recommendations [2], the radiologist should report back to the neurologist, qualitatively if not quantitatively, over the lesion status and the atrophy of MS patients, covering the following points:

- Comparison with previous scan(s)
- Evidence of new disease activity
- Number of new lesions (T2/T1)
- Lesion size
- Overall assessment, including the presence (definite/probable) and extent (number of new/enlarging lesions or gadolinium-enhancing lesions) of disease activity, change in T2 lesion volume and evidence of brain atrophy

Taking into account these recommendations, it is obvious that MRI biomarkers are already considered an important factor for making therapeutic decisions. Unfortunately, most MRI reports are written in prose and do not make use of the full potential embedded within the MRI datasets. Fortunately, communication regarding MRI findings between the (neuro)radiologist and the

neurologist can be improved with automatically computed, quantitative values for the relevant imaging biomarkers. To this end, the (neuro)radiologist should have easy access to approved tools for calculating these biomarkers.

In addition to having access to a structured radiology report, which includes quantitative data, the neurologist would benefit from having easy access, not only raw MRI scans, but also annotated image data sets. For example, the neurologist could examine overlays of tissue segmentations compared to previous MRI scans, overlays of individual brain structures or colour-coded overlays of lesions (new, enlarging or gadolinium-enhancing lesions).

For an adequate follow-up of patients with MS, it is essential to present the evolution of the imaging biomarkers in a relevant context. For instance, all available time points from a single patient should be used to plot the trend of each biomarker over time (see Fig. 15.4). These data could then be correlated with possible changes in treatment, or other events, over the same timeline. Furthermore, when following the evolution of changes in an individual patient, comparisons could be made of biomarker values against relevant populations (e.g. healthy controls, MS patients that respond well to therapy, etc.). Obviously, relevant confounding factors (such as age and sex) should be taken into account.

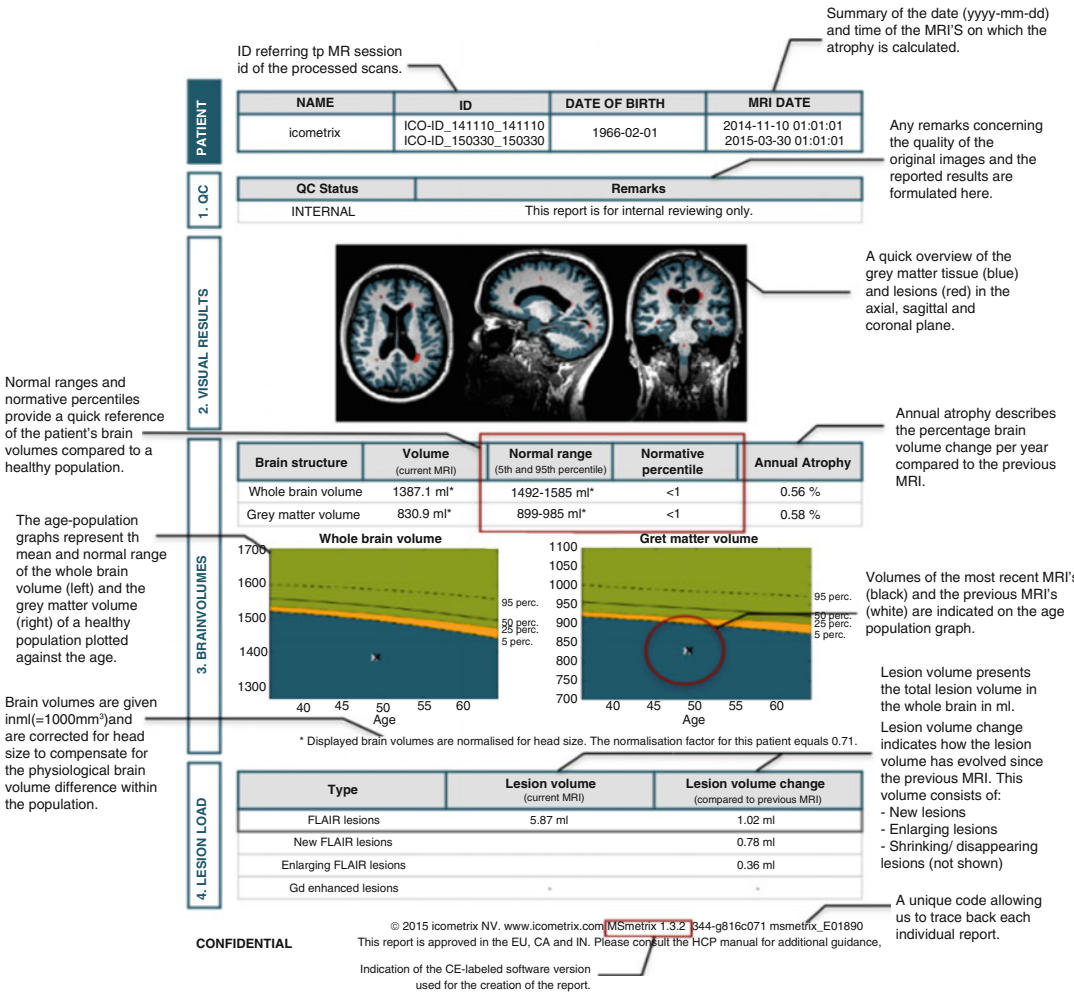


Fig. 15.4 Example of a typical MS imaging biomarker report (Image courtesy of ico*metrix*, Belgium)

The development of imaging biomarkers has led to a significant improvement in the diagnosis, management and follow-up of patients with MS. Standardisation of MRI acquisition protocols and improvement of quantitative reporting tools will provide better understanding of the natural history of MS and allow accurate treatment monitoring, for the greater benefit of patients.

References

1. Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, Molyneux P, Polman CH, Pozzilli C, Thompson AJ, Wagner K, Yousry TA, Miller DH. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. Neurology. 2009;72(7):595–601.
2. Arnold DL, Li D, Hohol M, Chakraborty S, Chankowsky J, Alikhani K, Duquette P, Bhan V, Montanera W, Rabinovitch H, Morrish W, Vandorpe R, Guilbert F, Traboulsee A, Kremenchutzky M. Evolving role of MRI in optimizing the treatment of multiple sclerosis: Canadian consensus recommendations. Mult Scler J Exp Transl Clin. 2015;1:1–9.
3. Bakshi R, Minagar A, Jaisani Z, Wolinsky JS. Imaging of multiple sclerosis: role in neurotherapeutics. NeuroRx J Am Soc Exp NeuroTher. 2005;2:277–303.
4. Benedict RHB, Zivadinov R. Risk factors for and management of cognitive dysfunction in multiple sclerosis. Nature reviews. Neurology. 2011;7(6):332–42.
5. Bermel R, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. Lancet Neurol. 2006;5(2):158–70.
6. Boyes RG, Rueckert D, Aljabar P, Whitwell J, Schott JM, Hill DLG, Fox NC. Cerebral atrophy measurements using Jacobian integration: comparison with the boundary shift integral. Neuroimage. 2006;32:159–69.
7. Calabrese M, Rinaldi F, Grossi P, Gallo P. Cortical pathology and cognitive impairment in multiple sclerosis. Expert Rev Neurother. 2011;11(3):425–32.
8. A. Traboulsee, J.H. Simon, L. Stone, E. Fisher, D.E. Jones, A. Malhotra,S.D. Newsome, J. Oh, D.S. Reich, N. Richert, K. Rammohan, O. Khan,E.-W. Radue, C. Ford, J. Halper, and D. Li. Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. AJNR Am J Neuroradiol. 2016;37(3):394–401.
9. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372:1502–17.
10. Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. Comput Math Methods Med. 2015;(450341):23. doi:10.1155/2015/450341.
11. De Stefano N, Airas L, Grigoriadis N, Mattle HP, O'Riordan J, Oreja-Guevara C, Sellebjerg F, Stankoff B, Walczak A, Wiendl H, Kieseier BC. Clinical relevance of brain volume measures in multiple sclerosis. CNS Drugs. 2014;28(2):147–56.
12. De Stefano N, Giorgio A, Battaglini M, Rovaris M, Sormani MP, Barkhof F, Korteweg T, Enzinger C, Fazekas F, Calabrese M, Dinacci D, Tedeschi G, Gass A, Montalban X, Rovira A, Thompson A, Comi G, Miller DH, Filippi M. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. Neurology. 2010;74(23):1868–76.
13. Durand-Dubief F, Belaroussid B, Armspache JP, Dufoura M, Roggeronea S, Vukusica S, Hannounb S, Sappey-Marinierb D, Confavreuxa C, Cotton F. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. AJNR Am J Neuroradiol. 2012;33:1918–24.
14. Filippi M, Rocca M. MR imaging of gray matter involvement in multiple sclerosis: implications for understanding disease pathophysiology and monitoring treatment efficacy. AJNR Am J Neuroradiol. 2010;31(7):1171–7.
15. Filippi M, Rocca M. MRI and cognition in multiple sclerosis. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2010;31 Suppl 2:S231–4.
16. Filippi M, Rocca M. Preventing brain atrophy should be the gold standard of effective therapy in MS (after the first year of treatment): No. Mult Scler (Houndmills, Basingstoke, England). 2013;19(8):1005–6.
17. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774–81. doi:10.1016/j.neuroimage.2012.01.021.
18. Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R, Thompson AJ, Miller DH. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain. 2008;131(3):808–17.
19. Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans Med Imaging. 1997;16(5):623–9.
20. Freedman MS, Selchen D, Arnold DL, Prat A, Banwell B, Yeung M, Morgenthau D, Lapierre Y, On Behalf Of The Canadian Multiple Sclerosis Working Group. Treatment optimization in MS: Canadian MS working group updated recommendations. Can J Neurol Sci Le Journal Canadien Des Sciences Neurologiques. 2013;40:307–23.
21. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal. 2013;17:1–18.
22. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, Ayache N. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage. 2011;57:378–90.
23. Geurts JJG, Calabrese M, Fisher E, Rudick RA. Measurement and clinical effect of grey matter

pathology in multiple sclerosis. Lancet Neurol. 2012;11(12):1082–92.
24. Giorgio A, De Stefano N. Cognition in multiple sclerosis: relevance of lesions, brain atrophy and proton MR spectroscopy. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2010;31 Suppl 2:S245–8.
25. Giorgio A, De Stefano N. Clinical use of brain volumetry. J Magn Reson Imaging. 2013;37(1):1–14.
26. Giorgio A, Stromillo ML, Bartolozzi ML, Rossi F, Battaglini M, De Leucio A, Guidi L, Maritato P, Portaccio E, Sormani MP, Amato MP, De Stefano N. Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis. Mult Scler. 2014;20(2):214–9.
27. Hyland M, Rudick RA. Challenges to clinical trials in multiple sclerosis: outcome measures in the era of disease-modifying drugs. Curr Opin Neurol. 2011;24(3):255–61.
28. Iglesias JE, Liu CY, Thompson PM, Tu ZW. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans Med Imaging. 2011;30:1617–34.
29. Inglese M, Grossman RI, Filippi M. Magnetic resonance imaging monitoring of multiple sclerosis lesion evolution. J Neuroimaging Off J Am Soc Neuroimaging. 2005;15(4 Suppl):22S–9.
30. Jacobsen C, Hagemeier J, Myhr K-M, Nyland H, Lode K, Bergsland N, Ramasamy DP, Dalaker TO, Larsen JP, Farbu E, Zivadinov R. Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. J Neurol Neurosurg Psychiatry. 2014;85(10):1109–15.
31. Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, De Mey J, Barkhof F, Steenwijk MD, Daams M, Maes F, Van Huffel S, Vrenken H, Smeets D. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. Neuroimage Clin. 2015;8:367–75. doi:10.1016/j.nicl.2015.05.003.
32. Khayati R, Vafadust M, Towhidkhah F, Nabavi M. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. Comput Biol Med. 2008;38:379–90.
33. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. Acad Radiol. 2008;15:300–13.
34. Lavery AM, Verhey LH, Waldman AT. Outcome measures in relapsing-remitting multiple sclerosis: capturing disability and disease progression in clinical trials. Mult Scler Int. 2014;2014:262350.
35. Lladó X, Ganiler O, Oliver A, Martí R, Freixenet J, Valls L, Vilanova JC, Ramió-Torrentà L, Rovira A. Automated detection of multiple sclerosis lesions in serial brain MRI. Neuroradiology. 2012;54(8):787–807.
36. Morgan CJ, Ranjan A, Aban IB, Cutter GR. The magnetic resonance imaging "rule of five": predicting the occurrence of relapse. Mult Scler (Houndmills, Basingstoke, England). 2013;19(13):1760–4.
37. Mortazavi D, Kouzani AZ, Soltanian-Zadeh H. Segmentation of multiple sclerosis lesions in MR images: a review. Neuroradiology. 2012;54(4):299–320.
38. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, Fujihara K, Havrdova E, Hutchinson M, Kappos L, Lublin FD, Montalban X, O'Connor P, Sandberg-Wollheim M, Thompson AJ, Waubant E, Weinshenker B, Wolinsky JS. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69:292–302.
39. Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP, Enzinger C, Ropele S, Alonso J, Sastre-Garriga J, Rovira A, Montalban X, Bodini B, Ciccarelli O, Khaleeli Z, Chard DT, Matthews L, Palace J, Giorgio A, De Stefano N, Eisele P, Gass A, Polman CH, Uitdehaag BM, Messina MJ, Comi G, Filippi M, Barkhof F, Vrenken H, MAGNIMS Study Group. Brain atrophy and lesion load predict long term disability in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2013;84(10):1082–91.
40. Radü EW, Bendfeldt K, Mueller-Lenke N, Magon S, Sprenger T. Brain atrophy: an in-vivo measure of disease activity in multiple sclerosis. Swiss Med Wkly. 2013;143(November):w13887.
41. Rao SM, Martin AL, Huelin R, Wissinger E, Khankhel Z, Kim E, Fahrbach K. Correlations between MRI and information processing speed in MS: a meta-analysis. Mult Scler Int. 2014;2014:975803.
42. Richards JE, Sanchez C, Phillips-Meek M, Xie W. A database of age-appropriate average MRI templates. Neuroimage. 2016;124(Pt B):1254–9. doi:10.1016/j.neuroimage.2015.04.055.
43. Riley C, Azevedo C, Bailey M, Pelletier D. Clinical applications of imaging disease burden in multiple sclerosis: MRI and advanced imaging techniques. Expert Rev Neurother. 2012;12(3):323–33.
44. Rocca M, Anzalone N, Falini A, Filippi M. Contribution of magnetic resonance imaging to the diagnosis and monitoring of multiple sclerosis. Radiol Med. 2013;118(2):251–64.
45. Rovira À, Auger C, Alonso J. Magnetic resonance monitoring of lesion evolution in multiple sclerosis. Ther Adv Neurol Disord. 2013;6(5):298–310.
46. Rovira À, Wattjes MP, Tintoré M, Tur C, Yousry TA, Sormani MP, De Stefano N, Filippi M, Auger C, Rocca MA, Barkhof F, Fazekas F, Kappos L, Polman C, Miller D, Montalban X, MAGNIMS study group. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process. Nat Rev Neurol. 2015;11(8):471–82. doi:10.1038/nrneurol.2015.106.
47. Rudick R, Weinshenker B, Cutter G. Therapeutic considerations: rating scales. In: Cook SD, editors. Handbook of multiple sclerosis. 3rd ed. ISBN 9780824741846 – CAT# DKE276. Series: neurological disease and therapy. CRC Press; New York – Basel. 2001.

48. Rudick R, Fisher E. Preventing brain atrophy should be the gold standard of effective therapy in MS (after the first year of treatment): Yes. Mult Scler (Houndmills, Basingstoke, England). 2013;19(8):1003–4.
49. Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C, Hemmer B, Mühlau M. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage. 2010;59(4):3774–83.
50. Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahy RM. Magnetic resonance image tissue classification using a partial volume model. Neuroimage. 2001;13:856–76.
51. Shi J, Baxter LC, Kuniyoshi SM. Pathologic and imaging correlates of cognitive deficits in multiple sclerosis: changing the paradigm of diagnosis and prognosis. Cogn Behav Neurol Off J Soc Behav Cogn Neurol. 2014;27:1–7.
52. Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. Neuroimage. 2010;49(2):1524–35.
53. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998;17(1):87–97.
54. Smeets D, Ribbens A, Sima DM, Cambron M, Horakova D, Jain S, Van Vlierberghe E, Terzopoulos V, Maertens A, Van Binst AM, Vaneckova M, Krasensky J, Uher T, Seidl Z, De Keyser J, Nagels G, De Mey J, Havrdova E, Van Hecke W. Reliable measurements of brain atrophy in individual patients with Multiple Sclerosis. Hum. Brain Mapp. 2016, 00: 1–12.e00518. doi:10.1002/brb3.518.
55. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. 2002;17:143–55.
56. Smith SM, Zhang YY, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage. 2002;17:479–89.
57. Steenwijk MD, Pouwels PJ, Daams M, van Dalen JW, Caan MW, Richard E, Barkhof F, Vrenken H. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). Neuroimage Clin. 2013;4(3):462–9. doi:10.1016/j.nicl.2013.10.003.
58. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: Improved N3 Bias Correction. IEEE Trans Med Imaging. 2010;29(6):1310–20. doi:10.1109/TMI.2010.2046908.
59. Vrenken H, Jenkinson M, Horsfield M, Battaglini M, van Schijndel RA, Rostrup E, Geurts JJ, Fisher E, Zijdenbos A, Ashburner J, Miller DH, Filippi M, Fazekas F, Rovaris M, Rovira A, Barkhof F, De Stefano N, MAGNIMS Study Group. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. J Neurol. 2013;260(10):2458–71. doi:10.1007/s00415-012-6762-5.
60. Wattjes MP, Rovira À, Miller D, Yousry TA, Sormani MP, De Stefano N, Tintoré M, Auger C, Tur C, Filippi M, Rocca MA, Fazekas F, Kappos L, Polman C, Barkhof F, Montalban X, on behalf of the MAGNIMS study group. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—establishing disease prognosis and monitoring patients. Nat Rev Neurol. 2015;11:597–606.
61. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging. 2001;20(1):45–57.

Use Case II: Imaging Biomarkers and New Trends for Integrated Glioblastoma Management

16

Elies Fuster-Garcia, Juan Miguel García-Gómez, Elena De Angelis, Arthur Sraum, Arthur Molnar, Sabine Van Huffel, and Georgios Stamatakos

Abbreviations

ADC	Apparent diffusion coefficient
CHTH	Chemotherapy
DCE	Dynamic contrast-enhanced MRI
DSC	Dynamic susceptibility contrast
DSS	Decision support system
DWI	Diffusion-weighted imaging
EHR	Electronic health record
GB	Glioblastoma
GUI	Graphical user interface
Kep	Contrast extraction coefficient
Ktrans	Volume transfer coefficient
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MRSI	Magnetic resonance spectroscopy imaging
NGS	Next-generation sequencing
PET	Positron emission tomography
PWI	Perfusion-weighted imaging
RCBV	Relative cerebral blood volume
RT	Radiotherapy
TMZ	Temozolomide
UX	User experience
WHO	World Health Organization

E. Fuster-Garcia (✉) • J.M. García-Gómez
Instituto Universitario de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas, Universitat Politècnica de València, Camí de Vera s/n. 46022 València, España
e-mail: elfusgar@itaca.upv.es; juanmig@ibime.upv.es

E. De Angelis
Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

A. Sraum
Xcommerce BVBA, Namsesteenweg 248, Heverlee 3001, Belgium

A. Molnar
INFO WORLD SRL, Strada Intrarea Glucozei 37-39, Bucuresti Sector 2, 023828, Romania

S. Van Huffel
Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, Box 2446, 3001 Heverlee, Belgium

G. Stamatakos
In Silico Oncology and In Silico Medicine Group, Laboratory of Microwaves and Fiber Optics, Institute of Communication and Computer Systems, School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou, Zografos, GR15780, Athens, Greece

16.1 Introduction

Glioblastoma (GB) implies a devastating prognosis with an average survival of 14–16 months using the current standard of care treatment [1]. GB is the most frequent malignant tumour originating from the brain parenchyma, and it is characterised by a marked intratumoural heterogeneity, proneness to infiltrate throughout the brain parenchyma, robust angiogenesis and necrosis as well as intense resistance to apoptosis and genomic instability [2].

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_16

Up till now, treatment and follow-up of GB remains one of the most challenging tasks in clinical oncology. The critical points in GB management are related to neurosurgical and radiotherapy (RT) planning and early-response-to-therapy assessment. These points link with (1) maximum safe resection of the tumour; (2) local RT dose value, distribution and technique; (3) histopathology diagnosis in terms of GB molecular characterisation; and (4) duration of adjuvant chemotherapy (CHTH) and best treatment during follow-up.

Recently, important advances have been made in the multiscale (molecular-cellular-tissue-patient) study of GB through the identification of parallel and dynamic tumour markers by techniques such as next-generation sequencing (NGS), immunohistochemistry characterisation, radiogenomics, multi-parametric images and circulating biomarkers from liquid biopsies. These have led to the definition of different molecular subtypes of GB, with prognostic and predictive-of-response implications [3], although this molecular classification is not actually extended in the clinical practice.

Additionally, the number of imaging modalities and associated imaging biomarkers available for the assessment of patients is considerably high and probably will grow in the following years. These include perfusion-weighted imaging (PWI), diffusion-weighted imaging (DWI), magnetic resonance spectroscopy imaging (MRSI) and positron emission tomography (PET). Although the added value of medical imaging in GB diagnostic, prognostic and treatment assessment is unmistakable, it has been demonstrated that no single modality in itself is specific enough to reveal the early response to treatment of GB tumours due to their heterogeneity and rapid evolution [4].

In this setting, decision-making requires the joint analysis of complex data acquired throughout the treatment and follow-up process, including molecular biomarkers, imaging biomarkers and clinical data. Moreover, a comprehensive analysis of the data acquired from the patient requires taking into account the three main dimensions of GB data: multilevel dimension, from voxel to population-based subtypes; multiscale dimension, from molecular to tissue scale; and temporal dimension, from single to longitudinal studies.

To support the analysis of these complex data, in recent years, significant advances have been made in the development of automated medical image analysis tools for brain tumours. These tools are able to generate automated segmentations of the different GB-related tissues (i.e. oedema, enhancing tumour, necrosis), hypoxia maps and other useful nosological images. The last decade has also witnessed increased research efforts in the field of multiscale cancer modelling including the development of in silico (i.e. on the computer) oncology models able to simulate different therapy outcomes based on the individual patient information.

The purpose of this chapter is not only to introduce the role of imaging biomarkers in the GB management but also to identify and introduce the new trends that will contribute to the successful inclusion of these biomarkers in an integrative multiscale analysis. To do so, this chapter will focus on (1) the description of the standard clinical workflow based on accepted clinical guidelines, (2) the identification of the main open questions in GB management, (3) the role of imaging biomarkers in GB management and (4) the introduction of the new trends in GB management.

Moreover, this chapter introduces an approach of how these new trends could be integrated in the complex scenario of multidisciplinary teams enabling the access and analysis of multiscale and multilevel data. This approach is based on a modular clinical decision support system (DSS) architecture for GB management to easily include and actualise analytic modules. Moreover, an overview of the integration strategy based on user experience (UX) is described to ensure the acceptability of the DSS by the multidisciplinary clinical community.

Potential clinical benefits of incorporating this knowledge in the tumour board meetings include advances in surgery and RT planning, adjuvant treatment selection, assessment of response, early recurrence detection and selection of subse-

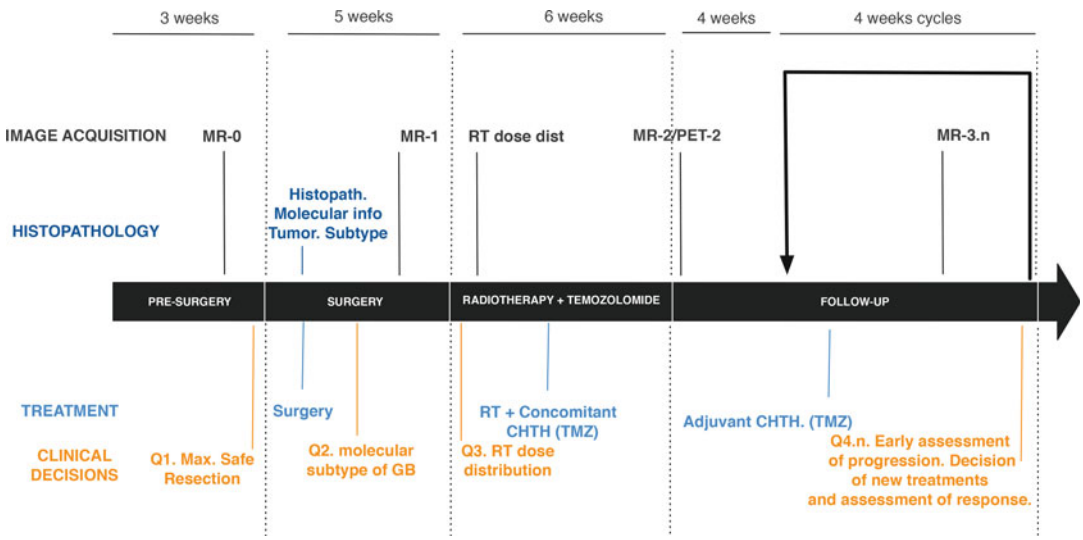


Fig. 16.1 Temporal diagram of the treatment and follow-up of GB patients including (1) the available clinical information at each stage, (2) the treatments (in *blue*) and (3) the main clinical decisions (in *orange*). RT dose dist. mean the information about the radiotherapy dose distribution. Q4.n and MR3.n mean the successive decisions and image acquisitions done during the follow-up, respectively

quent therapies. Moreover, this integrated approach will contribute to a better characterisation of GB subgroups, identification of new circulating biomarkers and identification of new targets for the treatment of patients with GB.

16.2 The Standard Clinical Workflow

Primary treatment after clinical or radiological evidences suggesting existence of GB consists on the maximum safe tumour resection based on the neurosurgical feasibility study. The extension of the tumour resection should be confirmed by postoperative magnetic resonance imaging (MRI) scan within 72 h after surgery, with and without contrast [5]. In case the resection is not recommended, a stereotactic or open biopsy or subtotal resection should be performed to establish the diagnosis. As soon as the pathology is available, the tumour expert panel or tumour board consultation is recommended.

After surgical intervention, the standard of care for newly diagnosed GB consists of adjuvant chemo-radiation therapy. In particular, surgery should be followed by RT and concurrent temozolomide (TMZ) CHTH and followed by six cycles of adjuvant TMZ. In the case of significant improvement on therapy, the inclusion of additional cycles of TMZ could be considered.

After the completion of RT, the follow-up of patients will consist on serial MRI scans. These MRI scans will be done in the second and sixth weeks (after RT), then every 2–4 months for 2–3 years and then less frequently [5]. The use of complementary imaging modalities such as MRSI, PWI or PET can be considered to facilitate the differentiation between pseudoprogression and radiation-induced necrosis.

In the case of local recurrence, a second resection is encouraged whenever it is possible. Following re-resection, or if the local recurrence is unresectable, poor prognosis patients should undergo best supportive care without further active treatment [5]. In case of diffuse or multiple recurring lesions, the options include surgery to reduce mass effect, the administration of systemic CHTH and best supportive care for poor prognosis patients.

The temporal diagram of the treatment and follow-up of GB patients is presented in Fig. 16.1.

16.3 Main Questions in Glioblastoma Management

Based on the above-mentioned standard clinical workflow, we could identify the following main steps in standard treatment for GB:

- Presurgery: to generate a first diagnosis based on medical imaging information
- Surgery: to remove the maximum safe area suspected to be affected by the tumour and to analyse the resected tissue to generate a more accurate diagnosis
- Concomitant RT with CHTH (based on TMZ): to irradiate the tumour tissue and to avoid the fast propagation of the tumour cells
- Follow-up including CHTH as adjuvant treatment to avoid the fast propagation of the tumour cells

In each of these four steps, important clinical decisions have to be addressed in order to select the most adequate treatment for each individual patient. Among them, the key decisions in GB diagnosis, therapy and follow-up are presented in the following subsections:

16.3.1 Presurgery Decision

What is the precise extension of the tumour that determines the maximum area that can be safely resected? Surgery of GB is by definition incomplete given the diffuse infiltrative nature of the tumour and the inability to remove it entirely without causing too much harm to the healthy brain. A major challenge in therapy of GB is the selection of the area for maximum safe resection of the tumour in order to reduce the degree and time to tumour recurrence while at the same time affecting the patient functionality as less as possible [6].

16.3.2 Post-surgery Decision

What is the molecular subtype of GB? What are the implications of GB subtyping in patient prognosis, treatment and follow-up? In recent years, analysis of genomics, transcriptomics and proteomics have identified subtypes of GB with prognostic implications and different responses to treatment. After surgery, it is possible to characterise the molecular subtype of GB using high-throughput arrays and immunohistochemistry techniques. Moreover, liquid biopsy may provide a wide set of biomarkers related to diagnosis, prognosis and treatment response. These biomarkers can circumvent problems of tumour heterogeneity and can be obtained to monitor tumour changes over time. It is now fully clear that different genetic subtypes of GB exist, associated with differences in molecular pathways involved and in biological behaviour. Therefore, clinical questions related to the prognosis and treatment response will be analysed in the context of knowledge of the molecular and genetic underpinnings.

16.3.3 Pre-radiotherapy Decision

What are the best RT dose value, distribution and technique for a specific patient? Currently the RT dose is estimated homogeneously based on anatomical images from PET or magnetic resonance (MR) scanners. The challenge in the use of RT is to reduce the margins beyond the conventional clinical target volume to the minimum in order to have optimised planning target volumes in accordance with the ICRU 62 definitions [7]. A reduction of RT treatment region uncertainty and a better estimation of the RT dose distribution based on the integration of functional information extracted from the images with dose painting could allow reduction of the radiation applied to brain functional areas where necessary and increase of radiation where possible, thereby improving the quality of life of the patients and their survival times.

16.3.4 Follow-Up Decisions

Is the treatment working properly? What should be the duration of adjuvant CHTH? What is the best treatment management during follow-up? Accurate interpretation of MRI scans in terms of

the biological evolution of the tumour is an important issue for measuring treatment response both in the setting of clinical trials and in routine clinical care. However, the evaluation of the disease progression still remains a difficult task in the face of treatment modality. Pseudoprogression of tumour versus true progression has become a confusing issue after treatment with TMZ and RT. Radiation injury (radionecrosis) is a potential late complication of RT, especially focal high-dose RT, and can easily be confused with tumour progression. Differentiating the two entities is problematic and often requires long-term follow-up with standard MRI, clinical assessment and use of corticosteroids [1]. By contrast, pseudo-responses may occur after angiogenesis-targeted therapies, as a consequence of changes in vascular permeability. In this sense, an early and accurate assessment of treatment response will improve the decision on maintenance or discontinuation of adjuvant CHTH, as well as the election and timing of subsequent treatments, including second-line CHTH and new local therapies.

16.4 Imaging Biomarkers in Glioblastoma Management

The development of imaging biomarkers is providing new insights into tumour behaviour that were not available from conventional medical imaging. Imaging biomarkers have demonstrated to be relevant for the assessment of tumour grading and response to therapy, without any spatial or temporal constraints. These imaging biomarkers are based on imaging modalities such as PWI, DWI, MRSI and PET.

The inclusion of PWI biomarkers characterising the presence and properties of angiogenesis, vasculogenesis and tumour vascular heterogeneity might improve tumour grading, prognosis and follow-up evaluation [8–10]. The complex modelling of dynamic susceptibility contrast (DSC) MRI sequences has also allowed for the quantification of tumour permeability and angiogenesis processes, through pharmacokinetic models of the lesion.

DWI may allow the cellularity of tumours to be graded noninvasively; because cells constitute a relative barrier to water diffusion, compared with extracerebral space, tumours that are more cellular are expected to show less of an increase in apparent diffusion coefficient (ADC) than tumours that are less cellular [15]. Diffusion tensor imaging and diffusion kurtosis imaging are used to describe diffusion 3D variability by means of mean diffusivity, fractional anisotropy and mean kurtosis. Several studies suggest that diffusion tensor imaging allows not only to observe high cellularity regions but also to evaluate tumour invasion into the surrounding tissue [16]. Studies of patients with brain tumours have shown that increases in water diffusion generally indicate positive response to therapy [15].

MRSI provides information regarding the concentration of specific metabolites throughout the brain, which has proven to be relevant in brain tumour diagnosis and prognosis. Thus, increased lipid levels are found in high-grade gliomas, indicating the presence of necrosis, which is a hallmark of GB [17]. Choline has been related to cell membrane density and is recognised as a marker of cell proliferation [17]. Statistically significant higher metabolite ratios of choline/creatine and choline/NAA have been reported in high-grade gliomas compared to low-grade gliomas [18]. Elevated choline levels have been found in peritumoural oedema surrounding GB, suggesting tumour invasion. After treatment, MRSI has also shown potential to differentiate tumour recurrence from radiation necrosis [19].

Several PET tracers have shown their added value when it comes to the diagnosis, prognosis and treatment monitoring of brain tumours. 18 F-FDG, which is a marker of glucose metabolism, has shown correlation with tumour grade and survival rate in gliomas [20]. Increased amino acid PET tracer uptake has been related to angiogenesis and increased cell metabolism within gliomas, resulting in a higher 11C-MET uptake in high-grade than in low-grade gliomas [21]. Labelled nucleotides such as 18 F-FLT are indicators of cellular proliferation, promoting 18 F-FLT kinetic analyses to assess early treatment response [22]. 18 F-FMISO is a hypoxia marker, showing increased uptake in high-grade

but not in low-grade gliomas. Tumour progression and survival after RT have been related to 18 F-FMISO uptake levels [23].

Hypoxia plays a central role in tumour development, angiogenesis, growth and resistance to treatment. Hypoxia measurements have been shown to correlate with the probability of metastatic spread, tumour recurrence, resistance to CHTH and radiation, invasion and decreased patient survival. Only a few imaging techniques have potential for in vivo assessment of hypoxia in humans, particularly for repeated, sequential measurements [24]. These methods use either PET tracers or MRI techniques sensitive to variations in local oxygen changes such as blood oxygenation level-dependent MRI (BOLD-MRI) or dynamic contrast-enhanced MRI (DCE-MRI)). An additional approach to map regional hypoxia is through the use of 3D MRSI and the quantification of lactate to N-acetyl-aspartate ratio with long echo times.

Although the added value of PWI, DWI, MRSI and PET is unmistakable, it has become clear that no single modality in itself is specific enough to show the early response to treatment of GB tumours due to their heterogeneity and evolution speed [4]. Hence, some groups have studied the complementary information provided by different modalities and techniques. Laimon et al. described the complementarity regarding tumour progression and response of dynamic [18 F] fluorothymidine (F-18 FLT) PET, sodium (23Na) MRI and 3-T morphological MRI biomarkers.

Manual segmentation is still the gold standard for brain tumours in clinical practice; however it implies a time-consuming and user-dependent bias, prone to errors and with questionable reproducibility. Significant progresses have been made in automated brain tumour segmentation based on machine learning [25–27]. Brain Tumour Segmentation (BRATS) Challenge on MICCAI Conference revealed that machine learning performs well in the whole tumour segmentation compared to manual segmentation. However, supervised learning requires an expensive, time-consuming and biased task to retrieve a sufficiently large set of labelled samples from which to learn discriminant functions for the posterior segmentation [26]. Moreover, spatio-temporal changes in clinical environment such as new MR machines, protocols or centres may distort the data and hence could affect the performance of the supervised models [28]. Unsupervised learning tackles these limitations in a more straightforward manner, as it directly learns the patient specific data to build an intra-patient segmentation model which is independent from the differences among patients [29].

16.5 New Trends for Integrated GB Management

Current clinical practices in GB management need to evolve to improve the poor results obtained to date in the treatment of this complex disease. To do so, the following promising approaches need to be particularly taken into account.

16.5.1 GB Molecular Subtypes

In the last decade, genomic analyses, transcriptomics and proteomics have identified different GB subtypes and molecular pathways with implications for prognosis and treatment response. For optimal management of patients, more precise classification of gliomas is needed, and molecular markers hold great promises in this respect.

The proneural subtype of GB, which is associated with better prognosis, is characterised by the expression of the histological markers OLIG2, DLL3, PDGFRA, IDH1 mutation (isocitrate dehydrogenase1), the absence of chromosomal gains or losses, the loss of TP53 heterozygosity and the normality of EGFR (epidermal growth factor receptor) as well as PTEN (phosphatase and tension homolog). The mesenchymal subtype, which corresponds to tumours of worse prognosis with strong angiogenic and inflammatory features, is characterised by the expression of mesenchymal markers such as YKL-40, PECAM1 (CD31), VEGF and its receptors one and two, gain of chromosome 7, loss of PTEN, normal or extended EGFR and MET, 17q11.2

deletion as well as high expression of genes of the TNF superfamily and NF-kB (nuclear factor kappa-light-chain-enhancer of activated B cells) signalling pathway. Other less consensual subtypes are proliferative and classic, which share loss of PTEN and frequent EGFR amplification. Proliferative subtype is characterised by histological markers, such as TOP2A and PCNA (proliferative cell nuclear antigen), and loss of chromosome 10. The classic subtype harbours frequent amplification of chromosome 7, loss of chromosome 10, amplification of EGFR gene and absence of alterations in TP53, NF1, PDGFRA or IDH1 [3, 30, 31]. The neural subtype is characterised by neural markers. In retrospective studies, it was observed that classic and mesenchymal tumours benefited from combined treatment of RT plus TMZ, while in the proneural, TMZ did not seem to provide therapeutic benefits [3]. However, prospective studies are necessary to confirm these findings.

On the other hand, recent studies suggest that antiangiogenic therapy could be beneficial in the proneural subtype and possibly in the proliferative subtype, but not in the mesenchymal [32]. All in all, it is currently clear that GB constitutes a 'mixed bag' of tumours and that the diagnosis of particular molecular subtypes (especially those characterised by mutations in IDH1/IDH2, H3F3A, BRAF, EGFR variant III (EGFRvIII) or EGFR amplification) will be of relevance in daily clinical practice soon.

Given the complexity and costs of these studies, promising approaches endeavour to develop easy-to-use panels of workable tests in the clinical context able to provide a more precise classification of gliomas, such as the analysis of messenger RNA (mRNA) by Colman and coworkers [8], which identifies nine genes with prognostic value. Another study using immunohistochemistry methods used only three markers for the classification in proneural-like and classical-like subtypes (p53, PDGFRA and EGFR) [33]. In addition, the expression of specific proteins, such as OLIG2, DLL3, TOP2A, CD44, VEGF and FOXG1, has been validated as a feature of GB molecular subtypes. Notably, activation of the signalling pathways pErk1,2/pMAPK and pAKT has also shown its prognostic value in GB [9]. More recently, an innovative, minimal IHC-based scheme for GB subclass assignment was proposed in terms of positive staining for IDH1R132H for proneural, high-EGFR expression for the classical subtype and a combined high expression of PTEN, VIM and/or YKL40 for the mesenchymal subtype [10].

16.5.2 Key Enabling Molecular Biomarkers in the Clinical Practice

Tumour-derived molecular biomarkers include proteins, nucleic acids and tumour-derived extracellular vesicles. These molecular biomarkers are mainly identified in plasma, serum, blood platelets, urine and/or cerebrospinal fluid. These molecular biomarkers provide valuable information of the mechanisms associated with cancer hallmarks such as cell proliferation, tumour progression, invasion, cell cycle, angiogenesis and apoptosis. Recently, circulating tumour cells have also been identified in the blood of glioma patients. Circulating molecules, vesicles, 'tumour-educated' platelets and cells may be useful as easily accessible diagnostic, prognostic and/or predictive biomarkers to guide the patient management.

There is an increasing interest in identifying the protein profile of each GB subtype from peripheral blood samples, in addition to the immunohistochemical analysis of a small set of proteins for GB stratification and the activation of key signalling pathways to identify potential therapeutic targets. A different hypothesis suggests that the data obtained will allow the correlation of a simple immunohistological classification pattern with cell-free circulating proteins that may be used to predict prognosis as well as therapeutic response. If this hypothesis is confirmed, it is expected that a wide perspective will be opened for identifying new and more effective therapeutic targets. In this sense, advanced approaches aim to incorporate an accurate selection of molecular biomarkers (e.g. IHC, NGS, methylation status and chromosomal copy number aber-

rations) including those obtained from liquid biopsy in the clinical management of GB (e.g. cell-free circulating proteins and RNA sequencing of 'tumour-educated' blood platelets).

Thereby, these approaches may help to circumvent problems related to tumour heterogeneity and sampling error at the time of diagnosis. If the success of these methodologies is confirmed, it is expected that a wide perspective will be opened for identifying more informative biomarkers with diagnostic, prognostic, predictive and/or monitoring value and innovative, more promising therapeutic targets.

16.5.3 Advanced Multiscale Data Modelling in GB: In Silico Oncology Models

The main approaches for multiscale mathematical modelling of cancer have as a common starting point the fact that cancer is a genetic disease and that its evolution is related, since the very early stage to mutations that give acquired abilities in few or even single cells [34, 35]. The observation that the biological system under consideration has multiscale features has resulted in the development of mathematical models that essentially couple different models operating at different scales and that are able to cope with genomics, proteomics, cell-cell interaction, cell-environment interaction, release, diffusion and absorption of chemical factors. Specifically, the modelling of cancer dynamics at the lowest scale, namely, molecular and cellular scale, focuses on the critical changes within the cell that characterise cancer growth. These changes (i.e. self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, evading immune system attack and tissue invasion and metastasis) incorporate some aspects of genetic mutation, gene expression and evolutionary selection, leading to malignant progression. In various cases, this evolution is induced by external or concomitant actions (as an example, the effect of therapies) [36]. At the tissue scale, macroscopic models of gliomas focus on the heterogeneous and anisotropic characteristics of the brain-deducing models that are able to describe the growth of tumour masses and the diffusion of metastases in such an environment [37, 38].

There are two major cancer-modelling schools that may be identified: predominantly continuous and predominantly discrete models. Predominantly continuous models rely primarily on differential equations to describe processes such as diffusion of molecules, changes in tumour cell density and invasion of tumour cells into the surrounding tissue. Even the continuous mathematical models, which make use of partial ordinary differentiation equations and appropriate boundary conditions, have to undergo discretisation through the application of methods such as finite difference time domain or finite element techniques in order to practically deal with the high geometrical complexity of the biomechanical problem.

A tumour growth modelling approach based solely on the continuous and/or finitised form of the diffusion's reaction equation has a limited potential to efficiently address the complexities of the treatment response phenomena in the multiscale context. The latter include inter alia the existence and dynamics of different proliferation potential cell categories (stem cells, limited mitotic potential cells, differentiated cells), different cell-cycle phases (G1, S, G2, M), different radiosensitivity and chemosensitivity profiles, different times spent within each cell-cycle phase, etc.

Discrete modelling has gained significant momentum lately; it considers several discrete states in which cells may be found and possible transitions between them, governed by decision calculators, such as cytokinetic diagrams and agent-based techniques. Due to the hypercomplexity of cancer-related topics, each modelling approach is intrinsically able to successfully address only some of the aspects of this multifaceted problem. By combining the continuous and the discrete mathematical approaches, more comprehensive hybrid models addressing both glioma invasion and response to complex treatment modalities could emerge.

The ultimate goal of clinically oriented cancer simulation models is their eventual translation into clinical practice, which entails (a) thorough sensitivity analyses, in order to both comprehend and validate their behaviour, and at the same time gain further insight into the simulated mechanisms, in a more quantitative way, and (b) an adaptation and validation process based on real clinical data [39]. On the global level, the first large-scale, clinical trial-driven and clinically adaptable and testable oncosimulators have been developed by the In Silico Oncology and In Silico Medicine Group (ISO and ISMG) of the Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA), in the context of the ACGT FP6 EU project (http://acgt.ercim.eu/) for nephroblastoma and breast cancer within the framework of the SIOP 2001/GPOH (http://www.siop-online.org/) and the neoadjuvant trial of principle (TOP) clinical trials, respectively.

16.6 Including Key Enabling Technologies in Clinical Practice

After reviewing the new trends in GB management, we are ready to present an approach to cover the gap between the technologies (i.e. existing molecular GB subtyping techniques, multiscale-multilevel predictive models and biomarker images) and their integration in the clinical workflow for the management of GB patients.

Our approach consists on the development of a clinical DSS to support multidisciplinary tumour boards in the therapy planning and early treatment response assessment of GB, that is, a health information technology system designed to assist the different actors involved in the GB management with main clinical decision-making tasks. The proposed DSS will use heterogeneous data to support and personalise treatment and follow-up for GB patients (see Fig. 16.2). The main functionalities to achieve this goal are:

1. Accessible and structured data: To be able to access the heterogeneous data in a transparent and secure way by developing interoperability and security layers.
2. Generation of knowledge: The novel image analysis tools will generate segmentations of the GB extension, hallmark and nosological images and hypoxia mappings. The findings obtained from imaging data together with molecular and clinical information will feed the multiscale-multilevel predictive models to obtain predictions of the evolution of the tumour depending on the simulated treatment. Finally an automatic characterisation of the GB molecular subtype will be done.
3. Support to clinical decision: Once the knowledge that addresses the clinical questions is generated, it will be used to support the clinical decisions. The DSS will adapt the presentation of their outputs to the clinical workflow. Two main scenarios have been considered: the first one is the scenario where the clinician wants to access the DSS findings using the hospital electronic health record (EHR) viewer. The DSS will include visualisation templates for EHR viewers tailored to each user profile. The second one is the tumour board multidisciplinary scenario. In this scenario, the DSS will facilitate a multidisciplinary collaborative interface including the latest visual and interactive technologies to improve user experience and acceptability.

A large body of evidence over many years suggests that DSS can be helpful in improving both clinical outcomes and adherence to evidence-based guidelines. However, to this day, clinical decision support systems are not widely used outside of a small number of sites, the main reasons being (1) the relative difficulty of integrating such systems into clinical workflows and computer systems, (2) the acceptability by the final users (user experience) and (3) the capability of keeping DSS up to date [40].

In the case of GB, the fulfilment of these requirements is even more challenging due to the wide variety of multidisciplinary users that will interact with the DSS; the need for acquiring, integrating and processing a wide variety of complex clinical information (ranging from molecu-

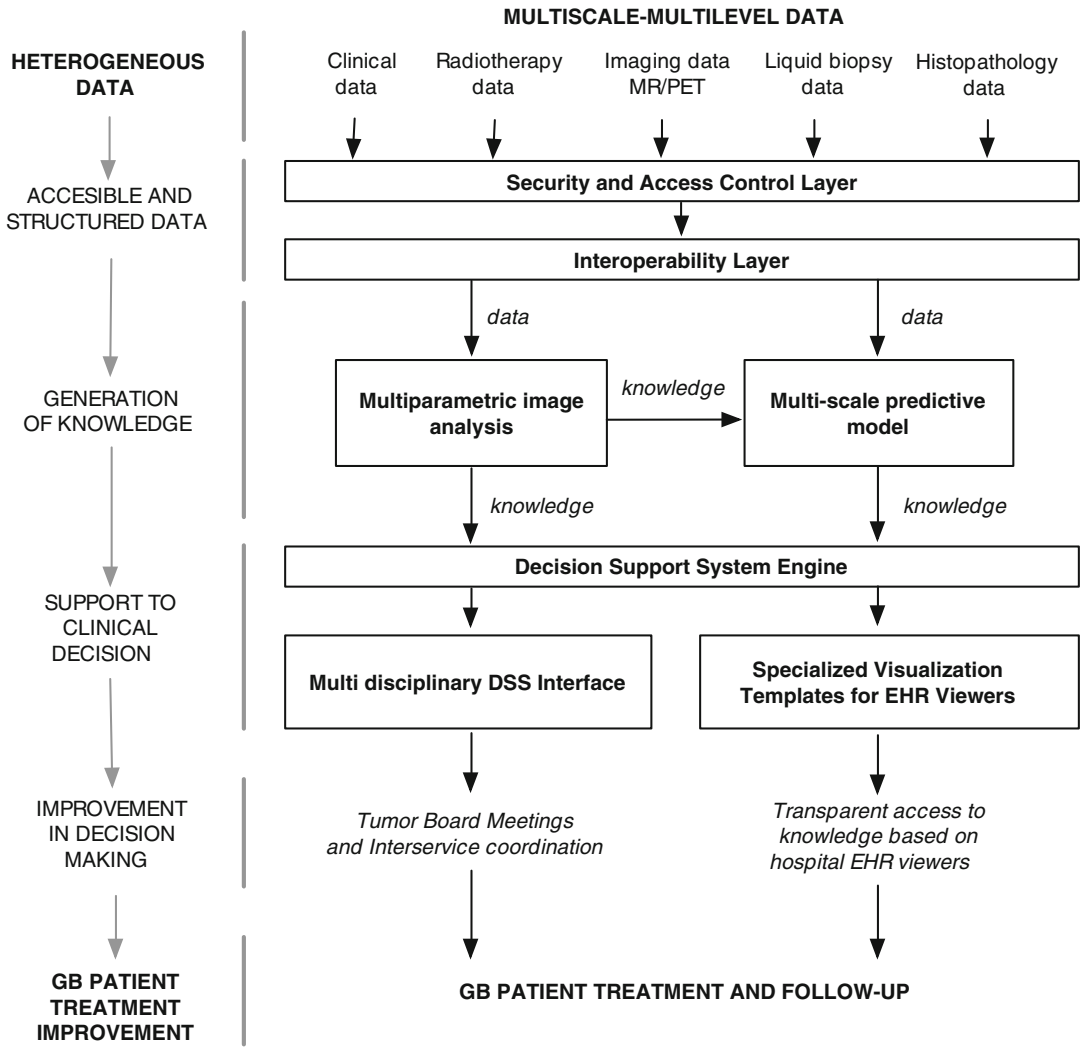


Fig. 16.2 Conceptual diagram of an advanced multiscale data modelling for GB management

lar data to multi-parametric stacks of images); and the need for covering the whole management of GB patients including surgery and RT planning and CHTH assessment.

A tentative schema of how the proposed DSS could be integrated in a clinical scenario is presented in Fig. 16.3. In this figure, we can see (1) how the DSS is integrated with the hospital clinical information systems, (2) how the results of the DSS are presented to the actors involved in the GB patient management by using specialised EHR-based visualisation templates and dedicated multidisciplinary DSS interface for tumour board meetings and (3) the structure of the DSS architecture.

In the following subsection, we will detail our approach to overcome the above section barriers:

16.6.1 Integrating the DSS into Clinical Workflows and Computer Systems

In order to facilitate the adoption of a DSS for GB management, it is critical to implement the mechanisms to accomplish the semantic interop-

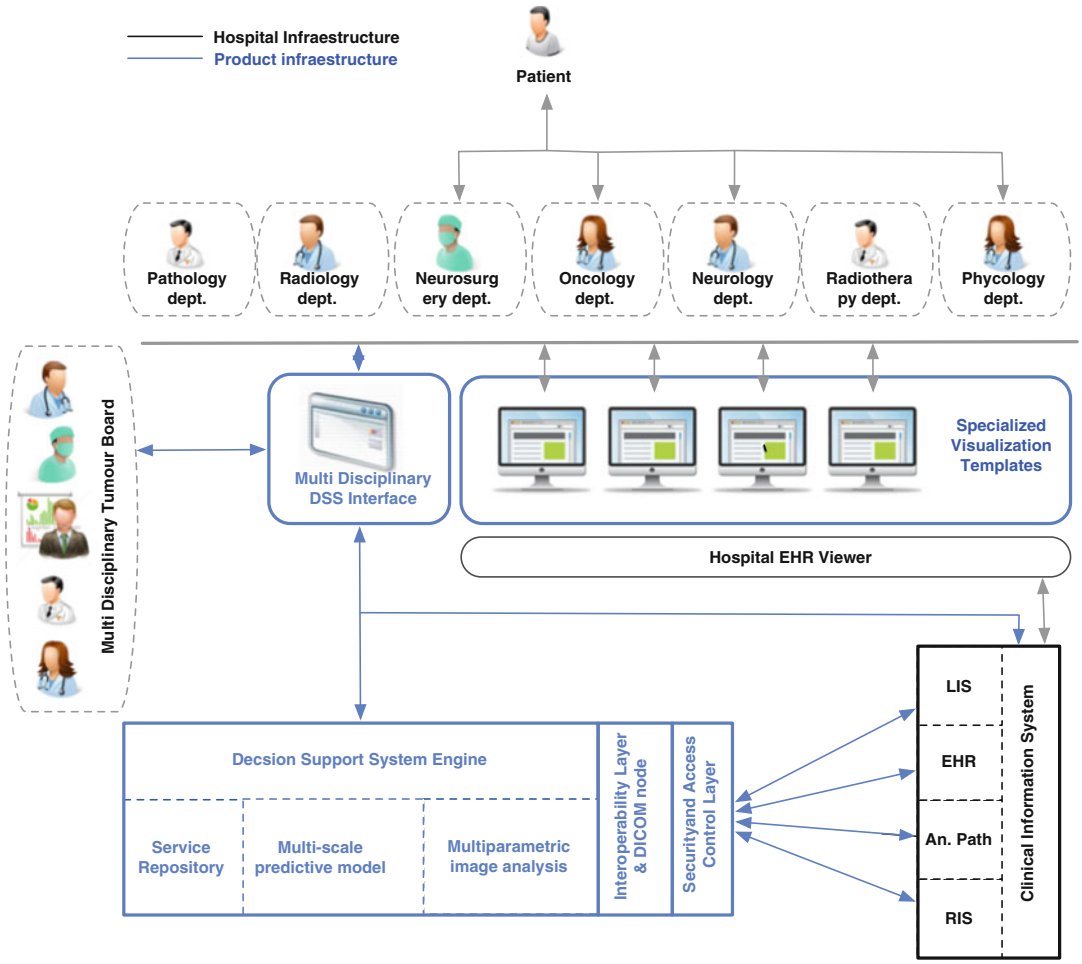


Fig. 16.3 Schema of an advanced DSS integrated in a clinical scenario. The DSS modules and connections are represented in *blue*, while the hospital infrastructures are represented in *black*

erability and complete integration with the existing hospital information systems. Following the interoperability standards and IHE profiles, we will ensure the integration and communication with different IT products already established in the IT infrastructure of the hospitals, enabling data capture from existing RIS/pathology/LIS/EHR systems. From our previous experience, this facilitates the adoption of the system at end-user level by presenting an already familiar user interface, reducing the requirements for manual data recording as well as the elimination of errors in the management of complex data by means of automation and integration at both hardware and software boundaries.

Moreover, it is important to make the DSS results accessible to clinicians at the moment when decisions are taken. To do this, we propose a double strategy consisting on the development of (1) highly visual user interfaces tailored for each of the multiple hospital areas involved in the tumour treatment and (2) an interactive interface and automatic reports for the multidisciplinary meetings.

As a result, the DSS will be fully integrated in the hospital workflows by means of the IHE profiles, which will ensure the sharing of the complete information, as well as the control and coordination of all the medical services involved in the GB management.

16.6.2 The Acceptability by the End Users

One of the major reasons why so many DSS are not used in clinical practice is that they lack positive user experience. Thus, the design of the graphical user interface (GUI) is of paramount importance in any healthcare tool and DSS development since it is the steering wheel. One could say that the GUI is actually as important as the accuracy of the algorithm and it should be intuitive and user-friendly and fulfil the needs of the user(s).

In the medical field, we are witnessing the situation that with more and more specialised examinations, tests and monitoring, physicians are faced with a significant amount of different but related pieces of information on each patient. Moreover, medical work is collaborative. Thus, for a patient with a brain tumour, several medical specialties and competences (i.e. neurologists, neurosurgeons, radiologists, oncologists, radiotherapists, (neuro)pathologists and clinical psychologists) must gather in several multidisciplinary team meetings (i.e. tumour board meetings), to present their findings and collaboratively discuss the diagnosis, treatment and follow-up. Although these tumour board meetings are part of the healthcare work process, there is room for improvement, particularly by tools that would facilitate the multidisciplinary meetings' flow and the access to multiscale information and would support the medical decision.

Identifying how technology can improve specialist interactions and enhance awareness at multidisciplinary meetings delivers real benefits [41]. In this chapter, we aim to design a GUI that will improve patient information visualisation and the interaction in tumour board meetings.

There have been a number of research studies for the improvement of displaying medical information when dealing with multiscale information. In [42], it is shown that using advanced visualisation techniques helps clinicians in improving their work process. Moreover, in [43], it is demonstrated that improved patient information visualisation is given by showing details prominently and presenting overviews.

Due to the extraordinary boost of technology in the field of human-machine devices, special attention was given in the last years to human-machine interaction and therefore to the user interface design. A number of diverse methodologies outlining techniques for human-computer interaction design have emerged in the last years. Among them UX methodology is the most popular as it is the best approach for evaluating how the user perceives the system before, during and after interacting with it. UX methodology is very intuitive and user-friendly and therefore highly recommended in clinics as several studies suggest that user involvement is crucial to a successful design and implementation of a successful tool for healthcare (see [44]). To avoid a top-down approach of designing a GUI for healthcare, close collaboration is needed between designers, implementers and the end users.

Acknowledgements This work was partially supported by project TIN2013-43457-R: Caracterización de firmas biológicas de glioblastomas mediante modelos no-supervisados de predicción estructurada basados en biomarcadores de imagen, co-funded by the Ministerio de Economía y Competitividad of Spain; the project CON-2014-001 Unsupervised glioblastoma tumour component segmentation based on perfusion multi-parametric MRI and spatial/temporal constraints, co-funded by the Global Investigator Initiated Research Committee (GIIRC) research programme by BRACCO, the Flemish Government FWO project G.0869.12 N and the project CURIAM-FDFT: Solución computacional del modelo mul-tinivel in vivo de la dinámica de la angiogénesis para la detección temprana de respuesta a tratamiento en glioblastomas primarios, co-funded by the ITACA Institute, UPV. Additionally, E. Fuster-Garcia acknowledges the financial support from the programme PAID- 10–14: Ayudas para la Contratación de Doctores para el Acceso al SECTI founded by the Universitat Politècnica de València.

References

1. Huang RY, Neagu MR, Reardon DA, Wen PY. Pitfalls in the neuroimaging of glioblastoma in the era of antiangiogenic and immuno/targeted therapy – detecting illusive disease, defining response. Front Neurol. 2015;6:33. doi:10.3389/fneur.2015.00033
2. Inda M-M, Bonavia R, Seoane J. Glioblastoma Multiforme: A Look Inside Its Heterogeneous Nature. *Cancers*. 2014;6(1):226–239. doi:10.3390/cancers6010226.

3. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN, Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17(1):98–110.
4. Mang A, Schnabel JA, Crum WR, Modat M, Camara-Rey O, Palm C, Brasil Caseiras G, Rolf Jäger H, Ourselin S, Buzug TM, Hawkes DJ. Consistency of parametric registration in serial MRI studies of brain tumor progression. Int J Comput Assist Radiol Surg. 2008;3(3 – 4):201–11.
5. Brem SS, Bierman PJ, Brem H, Butowski N, Chamberlain MC, Chiocca EA, DeAngelis LM, Fenstermaker RA, Friedman A, Gilbert MR, et al. Central nervous system cancers. J Natl Compr Canc Netw. 2011;9(4):352–400.
6. Kesari S. Understanding glioblastoma tumor biology: the potential to improve current diagnosis and treatments. Semin Oncol. 2011;38 Suppl 4:S2–10.
7. Purdy JA. Current ICRU definitions of volumes: limitations and future directions. Semin Radiat Oncol. 2004;14(1):27–40.
8. Du J, Teng RJ, Guan T, Eis A, Kaul S, Konduri GG, Shi Y. Role of autophagy in angiogenesis in aortic endothelial cells. Am J Physiol Cell Physiol. 2012;302(2):C383–91.
9. Petit I, Jin D, Rafii S. The SDF-1-CXCR4 signaling pathway: a molecular hub modulating neoangiogenesis. Trends Immunol. 2007;28(7):299–307.
10. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436–44.
11. Revert Ventura AJ, Sanz-Requena R, Martí-Bonmatí L, Jornet J, Piquer J, Cremades A, Carot JM. [Nosological analysis of MRI tissue perfusion parameters obtained using the unicompartmental and pharmacokinetic models in cerebral glioblastomas]. Radiologia. 2010;52(5):432–41.
12. Bulakbasi N, Kocaoglu M, Farzaliyev A, Tayfun C, Ucoz T, Somuncu I. Assessment of diagnostic accuracy of perfusion MR imaging in primary and metastatic solitary malignant brain tumors. AJNR Am J Neuroradiol. 2005;26(9):2187–99.
13. Awasthi R, Rathore RK, Soni P, Sahoo P, Awasthi A, Husain N, Behari S, Singh RK, Pandey CM, Gupta RK. Discriminant analysis to classify glioma grading using dynamic contrast-enhanced MRI and immunohistochemical markers. Neuroradiology. 2012;54(3):205–13.
14. Sanz-Requena R, Revert-Ventura A, Martí-Bonmatí L, Alberich-Bayarri A, García-Martí G. Quantitative MR perfusion parameters related to survival time in high-grade gliomas. Eur Radiol. 2013;23(12):3456–65.
15. Provenzale JM, Mukundan S, Barboriak DP. Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response. Radiology. 2006;239(3):632–49.
16. Price SJ, Jena R, Burnet NG, Hutchinson PJ, Dean AF, Peña A, Pickard JD, Carpenter TA, Gillard JH. Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: an image-guided biopsy study. AJNR Am J Neuroradiol. 2006;27(9):1969–74.
17. Bulik M, Jancalek R, Vanicek J, Skoch A, Mechl M. Potential of MR spectroscopy for assessment of glioma grading. Clin Neurol Neurosurg. 2013;115(2):146–53.
18. Law M, Yang S, Wang H, Babb JS, Johnson G, Cha S, Knopp EA, Zagzag D. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. AJNR Am J Neuroradiol. 2003;24(10):1989–98.
19. Weybright P, Sundgren PC, Maly P, Gomez Hassan D, Nan B, Rohrer S, Junck L. Differentiation between brain tumor recurrence and radiation injury using MR spectroscopy. AJR Am J Roentgenol. 2005;185(6):1471–6.
20. Padma MV, Said S, Jacobs M, Hwang DR, Dunigan K, Satter M, Christian B, Ruppert J, Bernstein T, Kraus G, Mantil JC. Pre-diction of pathology and survival by FDG PET in gliomas. J Neurooncol. 2003;64(3):227–37.
21. Singhal T, Narayanan TK, Jacobs MP, Bal C, Mantil JC. 11c-methionine PET for grading and prognostication in gliomas: a comparison study with 18f-FDG PET and contrast enhancement on MRI. J Nucl Med Off Publ Soc Nucl Med. 2012;53(11):1709–15.
22. Ullrich RT, Kracht LW, Jacobs AH. Neuroimaging in patients with gliomas. Semin Neurol. 2008;28(4):484–94.
23. Spence AM, Muzi M, Swanson KR, O'Sullivan F, Rockhill JK, Rajendran JG, Adamsen TC, Link JM, Swanson PE, Yagle KJ, Rostomily RC, Silbergeld DL, Krohn KA. Regional hypoxia in glioblastoma multiforme quantified with [18f]fluoromisonidazole positron emission tomography before radio-therapy: correlation with time to progression and survival. Clin Cancer Res Off J Am Assoc Cancer Res. 2008;14(9):2623–30.
24. Mendichovszky I, A Jackson. Imaging Hypoxia in Gliomas. The British Journal of Radiology. 2011;84 (special_issue_2):145–58
25. Bauer S, Wiest R, Nolte LP, Reyes M. A survey of MRI-based medical image analysis for brain tumor studies. Phys Med Biol. 2013;58(13):R97–129.
26. Gordillo N, Montseny E, Sobrevilla P. State of the art survey on MRI brain tumor segmentation. Magn Reson Imaging. 2013;31(8):1426–38.
27. Menze B, Reyes M, van Leemput K. The multi-modal brain tumorImage segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2014; 34(10):1993–2024.
28. Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley; 2001.

29. Juan-Albarracín J, Fuster-Garcia E, Manjón JV, Robles M, Aparici F, Martí-Bonmatí L, García-Gómez JM. Auto-mated glioblastoma segmentation based on a multiparametric structured unsupervised classification. PLoS One. 2015;10(5):e0125143.
30. Huse JT, Phillips HS, Brennan CW. Molecular subclassification of diffuse gliomas: seeing order in the chaos. Glia. 2011;59(8):1190–9.
31. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz Jr LA, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. Science. 2008;321(5897):1807–12.
32. H Phillips, T Sandmann, C Li, T Francis Cloughesy, OL. Chinot, W Wick, R Nishikawa, WP. Mason, R Henriksson, F Saran, A Lai, N Moore, PS. Hegde, LE. Abrey, R Bourgon, J Gar-cia, C Bais. Correlation of molecular subtypes with survival in AVAglio (bevacizumab [Bv] and radiotherapy [RT] and temozolomide [T] for newly diagnosed glioblastoma [GB]). J Clin Oncol. 2014;32:5s(suppl; abstr 2001^).
33. Le Mercier M, Hastir D, Moles Lopez X, De Nève N, Maris C, Trepant AL, Rorive S, Decaestecker C, Salmon I. A simplified approach for the molecular classification of glioblastomas. PLoS One. 2012;7(9):e45475.
34. Bellouquid A, De Angelis E, Knopoff D. From the modeling of the immune hallmarks of cancer to a black swan in biology. Math Models Methods Appl Sci. 2012;23(05):949–78.
35. De Angelis E. On the mathematical theory of post-Darwinian mutations, selection, and evolution. Math Models Methods Appl Sci. 2014;24(13):2723–42.
36. Bellouquid A. Mathematical tools towards the modelling of biological systems. Nuclei Online. 2013;1(1):1–3
37. Wolkenhauer O, Auffray C, Brass O, Clairambault J, Deutsch A, Drasdo D, Gervasio F, Preziosi L, Maini P, Marciniak-Czochra A, Kossow C, Kuepfer L, Rateitschak K, Ramis-Conde I, Ribba B, Schuppert A, Smallwood R, Stamatakos G, Winter F, Byrne H. Enabling multiscale modeling in systems medicine. Genome Med. 2014;6(3):21.
38. Scianna M, Preziosi L, Wolf K. A Cellular Potts Model simulating cell migration on and in matrix environments. Math Biosci Eng MBE. 2013;10(1):235–61.
39. Stamatakos GS, Georgiadi EC, Graf N, Kolokotroni EA, Dionysiou DD. Exploiting clinical trial data drastically narrows the window of possible solutions to the problem of clinical adaptation of a multiscale cancer model. PLoS One. 2011;6(3):e17594.
40. Wright A, Sittig DF. A Four-Phase Model of the Evolution of Clinical Decision Support Architectures. International journal of medical informatics. 2008;77(10):641-649. doi:10.1016/j.ijmedinf.2008.01.004.
41. Kane B, Luz S. Achieving diagnosis by consensus. Comput Supported Coop Work. 2009;18(4):357–92.
42. E McLoughlin, DO'Sullivan, M Bertolotto, DC. Wilson. MEDIC: MobilE Diagnosis for Improved care. In: Proceedings of the 2006 ACM Symposium on Applied Computing, SAC '06, New York: ACM; 2006. p. 204–8..
43. T David Wang, K Wongsuphasawat, C Plaisant, and B Shneiderman. Visual information seeking in multiple electronic health records: design recommendations and a process model. In: Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10. New York: ACM; 2010. p. 46–55.
44. Waterson P, Glenn Y, Eason K. Preparing the ground for the 'paperless hospital': a case study of medical records management in a UK outpatient services department. Int J Med Inform. 2012;81(2): 114–29.

Use Case III: Imaging Biomarkers in Breast Tumours. Development and Clinical Integration

17

M.A. Marino, K. Pinker, P. Baltzer, and T.H. Helbich

17.1 Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related death among women worldwide [1]. It is a heterogeneous disease with distinct molecular and genetic subtypes, each with characteristic clinical-biological behavior and imaging patterns. A substantial proportion of tumor markers or biomarkers, both fluid and tissue based, are currently used in the management of patients with breast cancer. Serum biomarkers, such as CA 15–3 and carcinoembryonic antigen (CEA), are not recommended in any senological guidelines due to a lack of sensitivity for early disease and a lack of specificity [2]. Rather, genetic tests are regularly performed in populations at high risk of developing breast cancer. By means of a minimally invasive blood test, women are told whether a germ line mutation in cancer susceptibility genes is present (BRCA). Being diagnosed with a BRCA 1 or 2 gene mutation has a dramatic impact on the life course of a woman. About 5–10 % of all breast cancers are caused by germ line mutations in the two breast cancer susceptibility genes, BRCA-1 and BRCA-2, and women carrying BRCA1 or BRCA2 mutations have an increased risk of developing breast cancer of approximately 50–80 % at 70 years of age [3–5]. Various histopathological and immunohistochemical staining-derived features of breast cancer are used to clinically establish the prognostically relevant subtype, including hormonal receptor expression, architectural growth patterns, and nuclear grades (low, intermediate, or high). Subsequently, the individual management of breast cancer patients is adapted according to these subtypes. Despite the enormous advances in breast imaging, the aforementioned clinically relevant subtyping of breast lesions is still based on invasive procedures, such as core needle biopsy or surgery. However, imaging is increasingly used to assess not only the morphologic features of the pathological process but also to assess the function of tumor tissues or to characterize individual phenotypes for targeted drug therapies, building on

M.A. Marino
Department of Biomedical Imaging and Image-guided Therapy, Division of Molecular and Gender Imaging, Medical University of Vienna, Vienna, Austria

Department of Biomedical Sciences and Morphologic and Functional Imaging, Policlinico Universitario G.Martino, University of Messina, Messina, Italy

K. Pinker • P. Baltzer • T.H. Helbich (✉)
Department of Biomedical Imaging and Image-guided Therapy, Division of Molecular and Gender Imaging, Medical University of Vienna, Vienna, Austria
e-mail: thomas.helbich@akhwien.at

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_17

developments in genomics and molecular biology features [6–9]. Imaging biomarkers can be defined as any anatomic, physiologic, biochemical, or molecular parameter that is detectable with one or more imaging methods used to help establish the presence and/or severity of disease. The specific term "quantitative imaging biomarkers" corresponds to parameters that are objectively and quantitatively measured noninvasively, are less susceptible to subjective judgment, and are resolved spatially and temporally [10]. Moreover, prerequisites for the effective use of imaging biomarkers are standardization and validation [10–12]. The aim of this chapter is to describe breast imaging biomarkers and their objective, quantifiable features. We consider here only imaging biomarkers that require an element of quantification and standardization, to demonstrate their contribution in the management of breast cancer patients. First, we briefly summarize the heterogeneous group of subtypes of breast cancer, both invasive and noninvasive, to give the reader a synopsis of the complexity of this disease and an insight into the molecular biomarkers of breast cancer. However, a complete overview of the different histologic breast cancer subtypes is beyond the aim of this chapter. Second, we discuss the role of breast imaging biomarkers in terms of risk prediction for the development of breast cancer. These include quantitative imaging techniques for the assessment of breast density based on mammography and the measure of fibroglandular tissue (FGT) and background parenchymal enhancement (BPE) based on magnetic resonance imaging (MRI). Third, we will explain the central role of MRI in noninvasively providing quantitative information about tissue characteristics, such as cell density, tumor angiogenesis, and metabolism, leading to an improved differentiation of benign and malignant breast lesions and to a better management of breast cancer patients in monitoring and predicting response to treatment. Finally, we discuss the potential of other imaging techniques, and emerging techniques, which could improve diagnostic accuracy and have the potential for the development of new imaging biomarkers.

17.2 Molecular Biomarkers of Breast Cancer

Molecular subtyping of breast cancer is increasingly used in clinical practice for the management of breast cancer patients. The assessment of the molecular subtype influences the prediction of prognosis and guides oncologists in choosing the most effective treatment [13]. The classification of ductal carcinoma in situ (DCIS) is still based on the traditional histopathological features, such as the growth pattern (solid, cribriform, papillary), the nuclear grade, and the presence of necrosis, and new molecular classifications still struggle to be implemented in clinical use [6]. In contrast, molecular subtyping of invasive cancer has been integrated into breast cancer treatment. Invasive breast cancer can be classified into the *luminal* category, if neoplastic cells present histologic features related to the epithelial cells lining the milk ducts, or the *basal* category, if tumor cells have characteristics similar to the myoepithelial cells of the basement membranes. Moreover, *luminal* and *basal* cancers have been further classified on the basis of the expression of pathological markers (estrogen [ER] and progesterone receptors [PR] and human epidermal growth factor 2 [HER2]) [14].

The major molecular subtypes of invasive cancer are clinically differentiated as follows:

- *Luminal A* breast cancer is characterized by the expression of both ER and PR, without amplification of the HER2neu antigen. This subtype is the most common and has the best prognosis, in part because these are more frequently low-grade tumors and, in part, because of their excellent response to hormonal therapy [15, 16].
- *Luminal B* breast cancer expresses hormone receptors similar to *Luminal A*, but differs because of the higher proliferative activity, assessed through Ki-67 levels, and for being often HER2neu enriched. These cancers are usually mid- to high-grade tumors, relatively insensitive to hormone therapy, and associated with an increased risk of relapses. In case of metastatic disease, bone is the target

organ. Consequently, this subgroup presents a poorer prognosis compared to *Luminal A* neoplasms [17].

– *Her2neu* breast cancer is characterized by overexpression of HER2, without the expression of hormone receptors. The amplification of HER2neu leads to increased cellular aggressiveness and rapid growth, resulting in generally intermediate- to high-grade tumors, a high recurrence rate, and metastases involving the brain [18]. Despite these unfavorable factors, this subgroup of breast cancer presents treatment options using antibodies against HER2neu [19, 20].
– *Basal-like* breast cancer is characterized by the lack of expression of ER, PR, and HER2neu markers and also by the overexpression of oncogenes, such as c-kit and the epidermal growth factor receptor gene, EGFR. The most common type of basal-like breast cancer is triple-negative cancer, occurring more commonly in young black and Hispanic women and in carriers of the BRCA-1 germ line mutation. These tumors are more frequently high-grade tumors with early nodal involvement, with a higher rate and earlier onset of recurrences compared to the other subgroups. Preferred localization of metastases is to the lungs and brain, and usually the prognosis is very poor [21, 22].

This clinical classification scheme does not take into account many other molecular biomarkers, such as the tumor suppressor gene *p53* and *claudin*, a protein involved in the harmonious formation of epithelium, which are not routinely investigated, although their role in carcinogenesis is well recognized. Thus, translating the clinical subtype of breast cancer into specific imaging phenotypes should only be performed considering that the clinical subgroups of breast cancers are still molecularly heterogeneous. In addition to this classic scheme, over the last decade, progress has been achieved in the field of biomarkers and genomics with both prognostic and predictive capabilities, to identify patients who will potentially benefit from additional therapy. Genotype testing is now available and can, therefore, relevantly change the management of breast cancer patients, providing additional information beyond the established clinicopathological features [23]. Some of these assays, such as Oncotype Dx and MammaPrint, have been approved by advisory and regulatory committees, while others are still in the approval process. In clinical practice, these gene-based predictors can be used in neoadjuvant and adjuvant therapies, to predict late recurrences and to tailor the duration of endocrine therapy for breast cancer without compromising its efficacy [23, 24]. Table 17.1 summarizes all the available multigene assays.

17.3 Quantitative Breast Imaging Biomarkers

17.3.1 Risk Prediction for the Development of Breast Cancer

17.3.1.1 Breast Density on Mammography

By expert consensus, mammographic breast density is one of the strongest known risk factors for the disease and is comparable to or exceeds values identified through many other known factors, such as alcohol or obesity [25–33] (Table 17.2).

Mammographic breast density is defined as the proportion of the breast area in the mammogram that is occupied by radiologic dense breast tissue. It refers to the variations in the radiological appearance of the breast among women, which reflect variations in breast tissue composition [25, 27, 28]. Mammographic density mainly consists of two components of tissues: fibroglandular and fatty tissue. Fibroglandular tissue is a mixture of fibrous connective tissue, specifically the stroma and the functional or glandular epithelial cells that line the ducts of the breast (the parenchyma). Fat has a lower X-ray attenuation coefficient than fibroglandular tissue and, thus, is more transparent on X-rays and appears darker on a mammogram. Regions of brightness that are associated with fibroglandular tissue are referred to as "mammographic densities."

Table 17.1 Comparison of major current genomic tests in breast cancer

Genomic test	Target	Specimen requirements	Technique	Characteristics
OncotypeDx	21 genes	FFPE	RT-PCR	Recurrence score (0–100) predicts likelihood of recurrence at 10 years. Role in predicting response to adjuvant and neoadjuvant chemotherapy validated
MammaPrint	70 genes	Unfixed tissue	Microarray	10 years survival probability divided into low and high risk
PAM50 ROR	50 genes	FFPE	RT-PCR, nCounter technology	Links intrinsic breast cancer subtype and 5 years risk of recurrence in ERpositive disease
76-Gene signature	76 genes	Unfixed tissue	Microarray	Predicts outcomes for LN-negative patients independent of hormone receptor status
H:I ratio/breast cancer index	HOXB13 to IL17BR ratio and molecular gene index	FFPE	RT-PCR	Particularly useful in predicting late recurrence (5–10 years from diagnosis)
IHC4	Levels (%) of ER, PR HER2, and Ki67	FFPE	IHC	Modeling factors in ER, PR, HER2, and Ki67 ICH. Predictor for distant recurrence
Genomic grade index	97 genes	Unfixed tissue	Microarray	Classifies histologic grades into low and high recurrence risk. High GGI is associated with increased risk of recurrence compared to low GGI

FFPE formalin fixed, paraffin embedded, *HOX* homeobox, *IL* interleukin, *ER* estrogen receptor, *PR* progesterone receptor, *HER* human epidermal growth factor receptor, *LN* lymph node, *RT-PCR* reverse transcription polymerase chain reaction, *IHC* immunohistochemistry, *GGI* genomic grade index

The relationship between the different mammographic density phenotypes (absolute dense area, absolute nondense area, and percentage dense area) and breast cancer risk seems to indicate that percentage dense area has been estimated to be a stronger risk factor than absolute dense area for breast cancer [28]. Women with a high percentage of dense area of >75 % had a 4.6-fold increased breast cancer risk (95 % confidence interval, 3.6–5.9) compared with women with little (<5 %) dense tissue [27–30, 34]. The level of risk is largely unknown because tissue density is influenced by multiple factors, both genetic and environmental. For example, breast density decreases with an increasing body mass index and decreases with aging and multiparity [26, 32, 33]. Other factors also affect a woman's tissue density, such as tamoxifen therapy, hormone replacement therapy, weight, and diet intake changes [35]. Nevertheless, regardless of the influence of other risk factors, breast density is now an established independent risk factor for the development of breast cancer [25, 26, 36] (Table 17.3).

Breast dense tissue is common in women, with 31–43 % of the general screening population having heterogeneously dense or extremely dense

Table 17.2 List of risk factors for breast cancer and relative risk

Factors that increase the relative risk for breast cancer		
Relative risk		Factors
	Gender	Female
	Age	>65
	Race	Caucasian
>4.0	Certain genome changes	BRCA1/BRCA 2
	Family history	Two or more fist-degree relatives with breast cancer diagnosed at an early age
	Personal health history	Prior diagnosis of breast cancer or high risk lesions
	Breast density	High percentage of dense tissue
	Family history	One first-degree relative with breast cancer
2.1–4.0	Personal health history	High-dose bone radiation to chest
		High bone density (postmenopausal)
		Early menarche (<12)
		Late age at first full-term pregnancy (>30)
		Late menopause (>55)
		No full-term pregnancy
	Factors that affect circulating hormones	Never breastfed a child
		Recent oral contraceptive use
1.1–2.0		Recent and long-term use of hormone replacement therapy
		Personal history of endometrium, ovary, or colon cancer
		Obesity (postmenopausal)
		Lack of physical activity
		Poor diet
	Other factors	Being overweighted or obese
		Alcohol consumption

Table 17.3 For percentage density measured using prediagnostic mammograms, there are combined relative risks of incident breast cancer in the general population

Risk of developing breast cancer associated to breast density				
	BI-RADS ACR a	BI-RADS ACR b	Bl-RADS ACR c	BI-RADS ACR d
% Breast density measured on mammography	5–24 %	25–49 %	50–74 %	>75 %
Combined relative risks	179(148–2.16)	2.11 (1.70–2.63)	2.92 (2.49–3 42)	4.64 (3.64–5.91)

breasts at mammography [27]. Therefore, breast density assessment has been an important component of screening mammography reports and has been propelled to the forefront by recent legislation in several states that requires patients to be informed of breast density and the potential for the increased cancer risk [31]. Breast density affects mammographic screening also because of the masking effect on underlying cancers. It is well known that mammographic sensitivity decreases with increasing density, from a level of 85.7–88.8 % in patients with almost entirely fatty tissue to 62.2–68.1 % in patients with extremely dense breast tissue [37, 38]. The masking effect of breast density leads to an increased percentage of interval cancers (cancers that manifest within 1 year of a normal mammogram) in women with dense breasts. It has been suggested that, because dense breasts may make a woman more likely to be diagnosed with an interval cancer, women with

dense breasts might benefit from intensified screening [27, 39]. Therefore, the ability to identify women at greater risk from breast cancer could be used as a prognostic indicator and as a selection criterion for high-risk groups, who would benefit from special surveillance measures [29, 30, 40–42]. Hence, breast density might have great potential as a *biomarker*, to be used as an intermediate end point in studies of breast cancer etiology and as a component of overall risk-assessment models. To date, breast density assessment, both qualitative and quantitative, is based on mammography. Mammography is a 2D technique, which examines the breast, a 3D structure, employing both ionizing radiation and compression [40, 41, 43]. Calculation of breast density based on mammography can be done either qualitatively, by subjective visual radiologist review, or quantitatively, by either semiautomated 2D or fully automated 3D computer analysis.

Qualitative assessment of mammographic density is the estimation of the percentage of mammographic dense tissue of the breasts by subjective visual assessment done by the reporting radiologist. The density scheme used is the one proposed by the Breast Imaging Reporting and Data System (BI-RADS) and American College of Radiology (ACR), which classifies density into four categories of percentage mammographic dense area [44]:

- BI-RADS ACR *a*: the breasts are almost entirely fatty, with dense tissue less than 25 % (Fig. 17.1a).
- BI-RADS ACR *b*: there are scattered areas of fibroglandular densities with dense tissue less than 50 % (Fig. 17.1b).
- BI-RADS ACR *c*: the breasts are heterogeneously dense with dense tissue less than 75 % that could obscure small masses (Fig. 17.1c).
- BI-RADS ACR *d*: the breasts are extremely dense breast with more than 75 % dense tissue, which lowers the sensitivity of mammography (Fig. 17.1d).

Subjective visual assessment is used to assign an overall breast composition rating to convey the likelihood of lesion obscuration, but this is not a standardized method, with several studies reporting only moderate interobserver agreement [45–47]. Therefore, computer-based quantitative and objective methods of density measurement have been developed. *Quantitative* breast density measurement methods provide a reproducible, calculated density percentage (either the area or the volume, depending on the method used) assessed either semiautomatically or fully automatically by a computer program using either a 2D or 3D approach. The density percentage is determined by dividing the calculated area (or volume) of the fibroglandular tissue by the calculated total breast area (or volume). The most commonly used 2D approach to calculate the area of density percentage is the semiautomated, interactive, thresholding technique (the Cumulus™). Through the software application of interactive gray-level thresholding values to digitized mammograms, this method allows the separation of mammographically dense tissue from fatty tissue and, thus, calculates the area of density percentage [26, 43]. The interactive thresholding method is not completely objective because it requires user input (Fig. 17.2).

Mammographic breast density measurement can be performed in 3D as well by directly assessing the "volume" or fractional volume of dense tissue. These methods create a physics model of the complete image-forming system, including the X-ray source, X-ray scattering and scatter removal, and the image receptor, and calculate a quantity which is referred to as the thickness of "interesting" fibroglandular tissue.

Currently, software programs (Table 17.4), approved by the US Food and Drug Administration (FDA), are available and provide fully automated volume density percentages that can be eventually converted to the appropriate BI-RADS density category [49–53].

The main drawback of all the existing breast density assessment methods is that they are based on mammography, a technique that requires the use of ionizing radiation, as well as strong compression of the breast, which is often uncomfortable or even painful for the patient. Mammography is a 2D method by which to assess the breast, which is a 3D structure, and therefore, mammography will

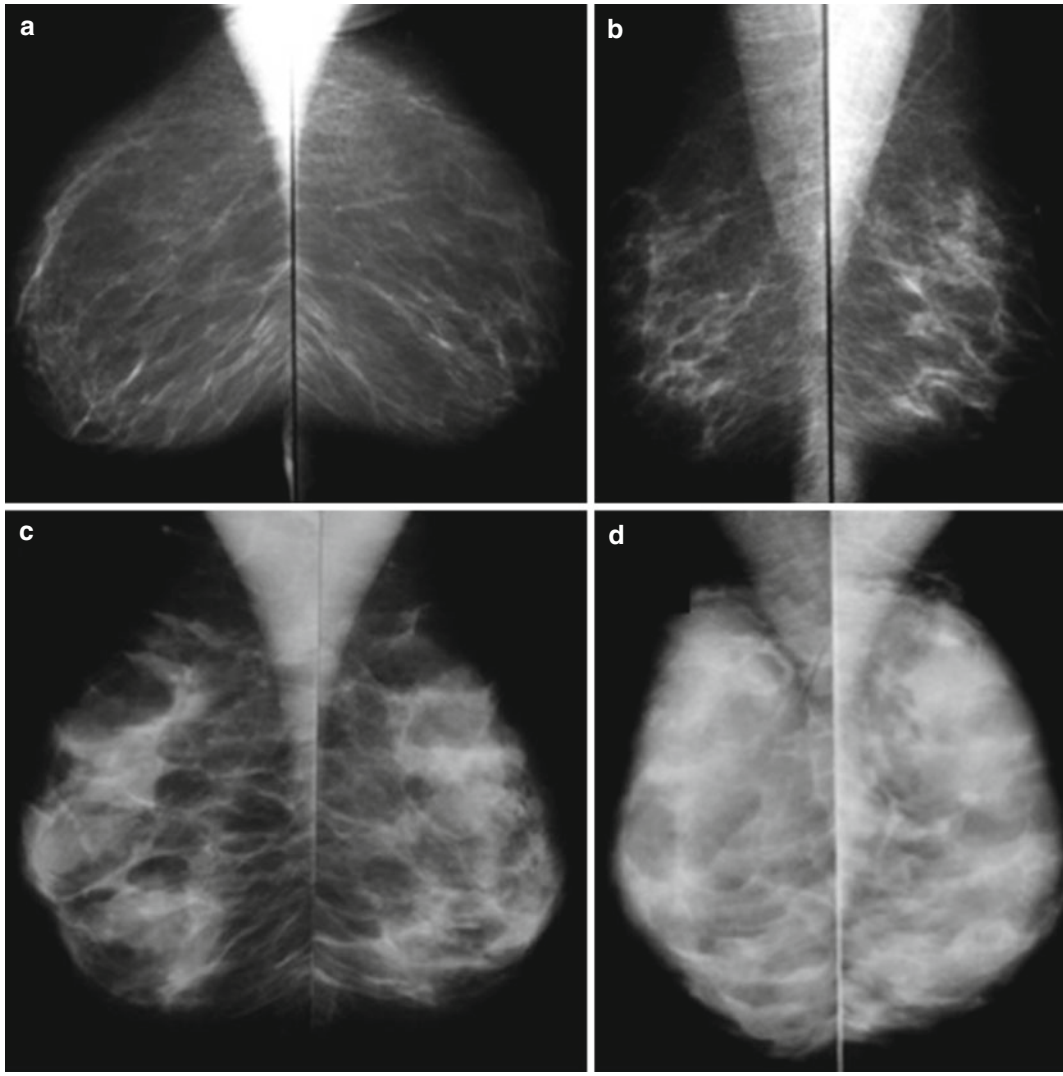


Fig. 17.1 The four-category ACR BI-RADS system for classifying mammographic density. The categories describe the fraction of fibroglandular tissue in the breast as judged by an observer and are (**a**) <25 %, (**b**) <50 %, (**c**) <75 %, and (**d**) >75 %

never be able to give an accurate absolute account of breast density, but rather a rough estimate.

17.3.1.2 Amount of Fibroglandular Tissue and Background Parenchymal Enhancement on Magnetic Resonance Imaging

The amount of fibroglandular tissue (FGT) refers to the composition of the breast, specifically the volume of non-fatty normal parenchyma assessed on MRI [54]. Thus, FGT greatly correlates to mammographic density and may provide superior accuracy in the determination of breast cancer risk [55–57].

MR imaging, with the administration of contrast media, reveals the enhancement kinetic features of normal parenchymal tissue, which is considered to reflect hormonal stimulation and proliferative activity. Background parenchymal enhancement (BPE) is defined as the intensity of FGT enhancement in the early phases of the

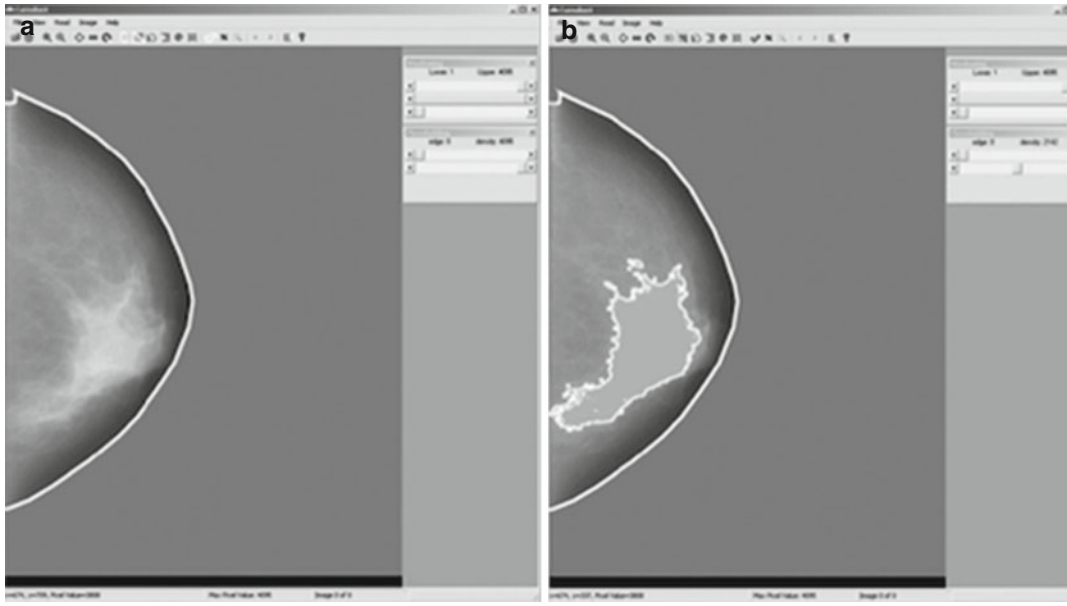


Fig. 17.2 The user interface for the interactive thresholding method for the determination of mammographic density (Cumulus ™). (**a**) The digitized or digital mammogram is displayed on the computer screen, and a threshold is selected by the operator to segment the breast from the surrounding background. (**b**) A second threshold is set to identify the regions of density. The algorithm indicates these pixels by a white overlay (Reproduced from Johns et al. [48])

Table 17.4 Quantitative assessment methods for breast density on mammography

Assessment type	Method	Device name	Specifics
Calculation of area density percentage	Semi-automated	CUMULUS™ University of Toronto, Canada	Interactive thresholding
		QUANTRA™ Hologic, Bedford, MA, USA	Fibroglandular density per pixel is estimated by using known image-acquisition parameters, including breast thickness; the pixel values are then added to determine the volume of the fibroglandular tissue
Calculation of volume density percentage (breast segmentation)	Completely automated	VOLPARA™ Matakina Technology Limited, Wellington, New Zealand	Pixel value representing fat is identified automatically by the software to provide a reference value; individual pixels in the breast are compared with the reference value to determine X-ray attenuation and generate a density map; the volume of the fibroglandular tissue and the total breast volume are calculated by adding the corresponding pixel value

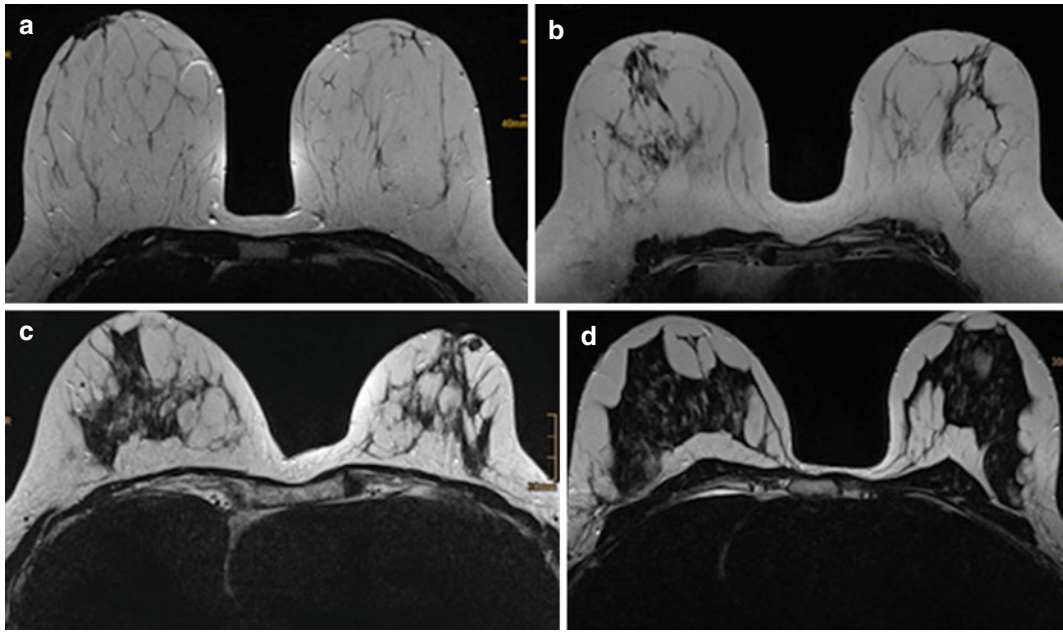


Fig. 17.3 The four-category BI-RADS system for classifying fibroglandular tissue. The categories describe the fraction of fibroglandular tissue in the breast and are (**a**) a, (**b**) b, (**c**) c, and (**d**) d

dynamic contrast-enhanced examination [58, 59]. It has been shown that BPE and, to a lesser extent, also FGT are hormonally responsive features that decrease over time with the effects of menopause, after oophorectomy, and in women who underwent treatment with tamoxifen or aromatase inhibitors [60, 61]. However, BPE is not necessarily positively associated with FGT nor has a strong correlation been shown with mammographic breast density [62, 63]. Initial research by two case-control studies in high-risk screening patients have demonstrated that BPE can be a predictive biomarker of breast cancer risk in these patients. In women who have already been determined to be at high risk, in particular, marked BPE increases the individual breast cancer risk up to tenfold [58, 64]. In addition, BPE is considered to possibly decrease the sensitivity of breast MRI by obscuring enhancing malignancies or may diminish the specificity by causing enhancement patterns that overlap with the appearance of cancers [62, 65]. However, cross-sectional studies demonstrated no major increase in either false-positive or false-negative MR findings due to BPE [59, 66].

Recently, there have been attempts to measure breast density with MRI. A qualitative assessment of both FGT and BPE is possible by following the BI-RADS classification descriptors [44]. Analogous to breast density, FGT can be classified as follows:

- BI-RADS *a*: the breasts are almost entirely fatty (Fig. 17.3a).
- BI-RADS ACR *b*: there are scattered areas of fibroglandular tissue (Fig. 17.3b).
- BI-RADS ACR *c*: heterogeneous fibroglandular tissue (Fig. 17.3c).
- BI-RADS ACR *d*: extreme fibroglandular tissue (Fig. 17.3d).

BPE can be classified in congruence with the BI-RADS lexicon according to the different level of appearance (from minimal to marked) and to the distribution within the breasts (symmetric or asymmetric) (Fig. 17.4). Because it is a cross-sectional imaging technique that provides 3D images of the breast, MR imaging provides images that are related to the fat and water composition of the breast. Since the water composition is highly

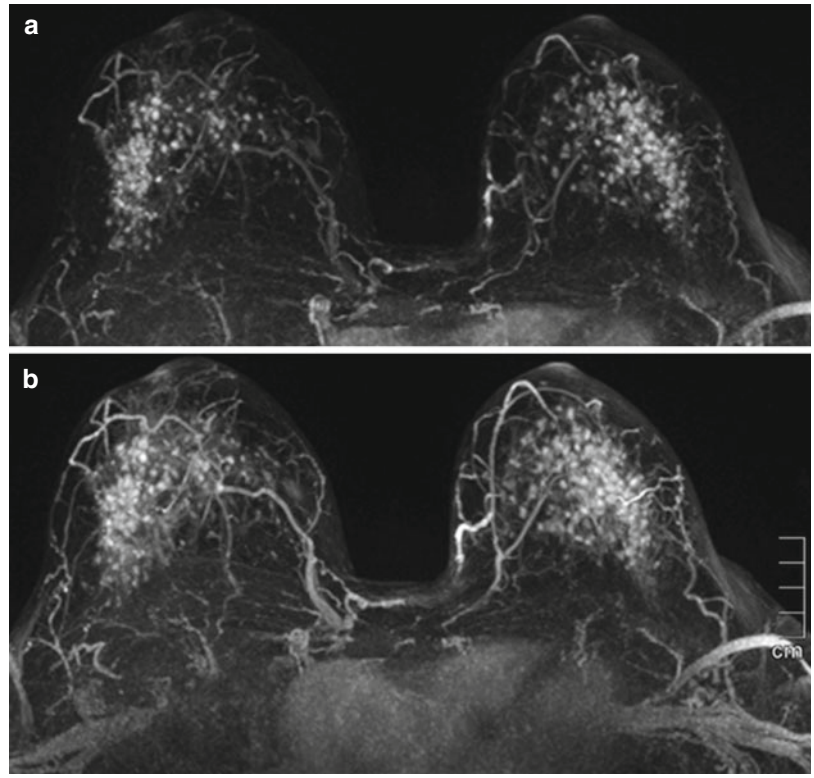


Fig. 17.4 A 44-year-old woman showing marked and symmetric nodular background parenchymal enhancement (BPE) described using the BI-RADS lexicon. (**a**) Maximum intensity projection (MIP) image after 2 min from injection of contrast agent, (**b**) MIP image after 8 min from injection of contrast agent

correlated with the prevalence of fibroglandular tissue, these images can be used for the slice-by-slice segmentation of the fibroglandular and fatty tissues for quantitative breast density assessment [57, 67, 68]. Multiple studies employing semiautomated algorithms have focused on the evaluation of the amount of FGT, demonstrating that breast MRI assessment of percent density correlates well with mammographic measures of percent density [33, 55, 56]. The current research on breast density measurement with MRI has been based on conventional T1-weighted sequences [33, 55, 56], which does not allow a fully automated separation of fatty and fibroglandular tissue due to the relatively small contrast between those two tissues. T1-weighted sequences provide gray-scale images, and therefore, the definition of a threshold by a radiologist is required to enable calculation of breast density. This human-defined threshold is a source of variability and poor sensitivity. A different approach, with the use of different sequences, has been recently proposed [69]. The Dixon technique represents an alternative to T1-weighted sequences for a fully automated, user-independent, MRI-based, quantitative volumetric breast density assessment. Dixon sequences are based on chemical shift imaging and, thus, are best suited for fat suppression in tissues containing similar amounts of lipid and water (e.g., fibroglandular tissue) [70–72] (Fig. 17.5). These sequences are fast and allow reliable and reproducible fat and water separation, and thus, a fully automated segmentation of fibroglandular and fatty breast tissue for breast density calculation is possible (Table 17.5).

17.3.2 Tumor Characterization and Prognosis for the Hallmarks of Breast Cancer with Magnetic Resonance Imaging (MRI)

The hallmarks of cancer are all the biological capabilities acquired during the multistep development of tumors which reflect genetic alterations that drive the progressive transformation of

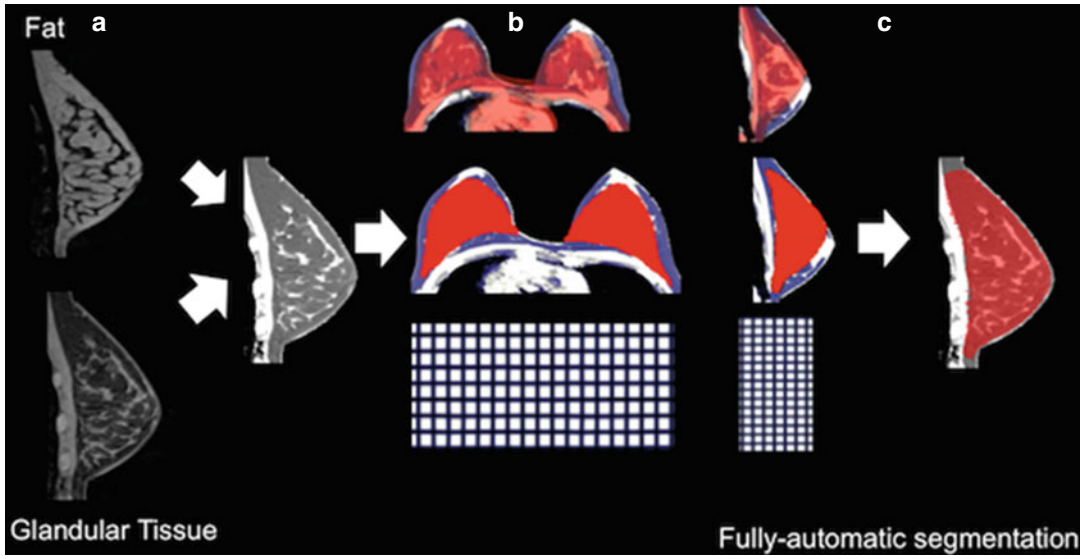


Fig. 17.5 Illustration of the automatic segmentation process of the breast. (**a**) The fat- and water-weighted Dixon images of the target subject are combined into a single image.(**b**) The template breast image (*red overlay*) with similar size and shape is registered to the subjects breast until both matched (*bottom*). (**c**) The deformation obtained from the registration process is applied on the associated manual template segmentation (*red*), resulting in an automatic segmentation of the target breast (*bottom*), excluding skin and the pectoralis muscle

Table 17.5 Interquartile ranges (25th to 75th percentile) of quantitative MG % BD estimation and the AQV MR % BD measurement system for each qualitative ACR BI-RADS BD category and the corresponding number of women

ACR	MG (IQR)	MRI (IQR)
ACR 1 (0–24 %)	8.42–33.84 % (mean 22.48 %)	3.47–7.84 % (mean 7.84 %)
ACR 2 (25–49 %)	26.71–41.38 % (mean 35.26 %)	7.84–25.88 % (mean 15.62 %)
ACR 3 (50–74 %)	41.99–63.86 % (mean 50.39 %)	26.25–44.15 % (mean 34.42 %)
ACR 4 (75–100 %)	61.75–79.91 % (mean 70.14 %)	39.86–71.20 % (mean 49.74 %)

Reproduced from Wengert et al. [69]
% indicates percentage; *ACR BI-RADS BD* American College of Radiology Breast Imaging Reporting and Data System breast density, *ACR* American College of Radiology, *AUQV MR % BD* automated quantitative volumetric magnetic resonance imaging percentage breast density, *IQR* interquartile range, *MG % BD* mammographic percentage breast density, *MG* mammography, *MRI* magnetic resonance imaging

normal human cells into highly malignant derivatives [73, 74]. They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis [73–77]. Recently, other hallmarks have also been reported, such as deregulating cellular energetics and avoiding immune destruction, and these processes play a central role in the pathogenesis of cancers [73] (Fig. 17.6). Tumors are, therefore, complex tissues consisting of multiple different cells that contribute to the development and expression of distinctive and complementary capabilities and which enable tumor growth and metastatic dissemination [73–77].

The understanding of these dynamic variations is crucial to improve tumor characterization and to develop new target therapies. However, a reasonably complete depiction of the network of microenvironmental signaling interactions is still far from complete, as the great majority of signaling molecules and pathways remain to be identified [73]. To elucidate these complexities of neoplastic disease, imaging techniques have to be equally sophisticated and multilayered. MRI may provide a wide range of information for the detection and

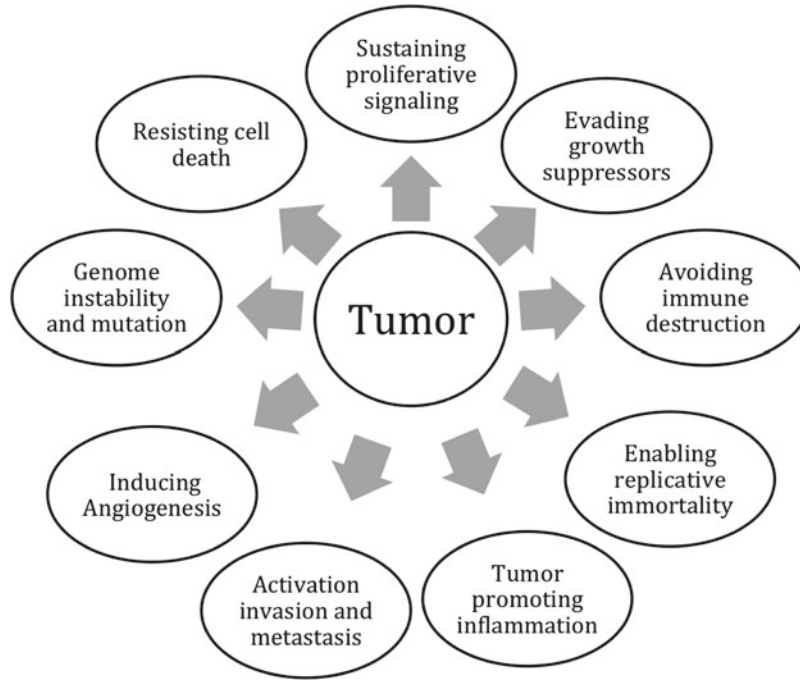


Fig. 17.6 The hallmarks of cancers. The capabilities of cancer involved in tumor growth and progression of some and perhaps all cancers (Modified from Hanahan et al. [73])

characterization of tumors as well as insight into monitoring the response to treatment [78–85]. MRI could provide excellent contrast because of variations in water proton density and relaxation times (which reflect macromolecular content) and provide the ability to probe abnormal vessel permeability, tissue microstructure (diffusion imaging), and chemical composition (e.g., concentration of amides or lipids, or metabolites) [86–88]. Furthermore, the use of quantitative measurements from MR images contributes to reporting and monitoring specific biological changes: for example, a decrease in vascularity as a result of antiangiogenic treatement or an increase/decrease in cell density as biomarker of changes in proliferations and an increased lactate production as an indicator of anaerobic metabolism indicator [88]. Thus, MRI assays tumor characteristics and provides a means to noninvasively characterize tumors for prognostication, for individualization and optimization of treatment, and for monitoring therapeutic response. In the next subchapters, we will focus on the currently available MR techniques that allow the assessment of some of the aforementioned hallmarks of breast cancer.

17.3.2.1 Imaging of Tumor Neoangiogenesis with Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI)

The hallmark of tumor growth and metastatic potential is neoangiogenesis, expression of dedicated vasculature that supports the high metabolic demand for oxygen especially of aggressive tumors [79, 89, 90]. Specific peptide hormones released by cancer cells promote the increased neoangiogenesis as soon as they exceed 2 mm [90, 91]. Being a limiting factor for both tumor growth and metastases, it has been assumed that angiogenesis correlates with tumor aggressiveness [92]. In particular, it has been widely demonstrated that microvascular density (MVD) (i.e., the number of endothelial clusters in a high-power microscopic field) correlates with the presence of metastases at time of diagnosis and with decreased patient survival times. Compared to conventional imaging, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has been so far the most sensitive method for detection of breast cancer with a negative

predictive value ranging between 89 and 99 % in general [5, 78, 93–95]. Breast DCE-MRI can be used in preoperative and postoperative staging, in screening settings of women at high-risk, in patients with cancers of unknown primary (CUP) syndrome, and as a reliable problem-solving tool to accurately excludes malignancy [82, 83, 96]. This high sensitivity technique allows the detection of lesions visible only on breast MRI, while several authors advocate its false-positive rate that varies but is consistently higher than that of conventional imaging [78]. As the causes for false-positive findings are known – small and non-mass lesions [97] – protocols that focus on improving temporal and spatial resolution at field strengths of 1.5 and 3 Tesla showed their feasibility but did not focus on the problem makers in breast MRI yet [72, 98, 99]. High-field 7-T MR imagers also were introduced to offer even higher temporal and spatial resolution imaging because of the increased intrinsic signal-to-noise ratio [100–102]. However, with appropriate techniques and experienced readers, specificities not lower than 80 % can be achieved using breast DCE-MRI at any fields (1.5, 3, 7 T) [79, 83, 99]. DCE-MRI of the breast highlights lesions after intravenous injection of a contrast agent based on morphologic assessment and the enhancement characteristics of lesions [100–102]. The enhancement of a lesion is closely related to the microscopic vascular architecture as the agent is injected intravenously and washes in into the lesion via sprouted vessels [79, 89]. DCE-MRI is able to visualize tumor vasculature as a tumor-specific feature derived from angiogenesis [96]. Neoangiogenetically introduced vessels typically presents abnormal vessel permeability that is mediated by cytokines. Microstructural tissue properties can be characterized through assessing kinetic enhancement curves obtained from DCE-MR imaging [103]. To assess these, low molecular contrast agents on gadolinium basis that are not actively metabolized in the human body and that do not enter the intracellular space are intravenously injected as a rapid bolus. The distribution in the examined tissue is measured by repetitive T1-weighted sequences with preferably high temporal resolution. Signal intensity time curves can be measured either visually, per regions of interest or pixelwise parametric mapping [104, 105]. Biological factors influencing these curves in breast lesions are microvessel density, permeability, and extracellular leakage space [106, 107]. The percentage of maximal signal increase correlates well with the microvessel count [108–110].

In addition to semiquantitative assessment approaches, quantitative evaluation of contrast kinetics can be performed through pharmacokinetic modeling techniques. Contrast media exchange between the intravascular and the interstitial space, providing a measure of tumor blood flow, microvasculature, and capillary permeability. The most commonly used pharmakokinetical model introduced by Tofts et al. can be used to interpret signal intensity time curves. The initial phase describes tissue perfusion (e.g., the exchange between plasma and interstitium (K^{trans}) that is influenced by microvessel density and permeability); the delayed phase is associated with the leakage space that is the extracellular extravascular space, described by the interstitial space volume fraction (v_e). The higher tissue perfusion, the higher the initial enhancement; the lower the extracellular distribution space, the faster the maximum possible concentration of contrast agent is reached, leading to a transport from the extravascular extracellular space back into the vasculature, causing a washout of the contrast medium. The preferable quantification of DCE-MRI data is limited by a number of technical and biological influence factors on signal intensity time curves, e.g., scanner type and field strength and homogeneity, acquistion sequence parameters and timing, contrast injection, contrast material, and intra-individual circulation differences [104, 111]. The optimal pharmacokinetic approach deriving model parameters from nonlinear regression analysis of signal intensity time curves has not been determined yet, a fact underlined by a wide variety of models and model variations [112]. To perform DCE-MR imaging pharmacokinetic modeling and calculate quantitative parameters, two parameters should preferably be evaluated: precontrast T1 relaxation times of the tumor or tissue and the arterial

input function (AIF). Precontrast T1 mapping requires acquisition of additional series of images with varying flip angle or inversion recovery sequences leading to inaccuracies due to B1 inhomogeneity especially at high field strengths [113]. This approach implicates additional minutes to the overall examination; thus, it is time-consuming [79, 103]. Most models require that the AIF be measured directly for each subject [106, 114], which is often challenging to perform and sensitive to patient motion between dynamic acquisitions. The most common approach was proposed by Huang and colleagues [114]. The authors proved the accuracy of an average AIF, i.e., shutter-speed approach calculated from a larger population for whom the injection site, dose, and rate were kept constant, with standard approach DCE-MR imaging pharmacokinetic analysis for breast cancer diagnosis. Yankeelov and colleagues [115] have proposed another method, estimating the AIF using a reference region model and found that such an approach correlated well with direct AIF measurement. Although model-based pharmacokinetics are challenging and time-consuming, they have the potential advantage over semiquantitative assessments of providing standardized measurements [116]. As discussed elsewhere in detail, pharmacokinetic modeling of DCE-MRI data is not as robust as often suggested [112]. In clinical practice, semiquantitative analysis of kinetic features in a breast lesion is usually measured using modest temporal resolution with at least 2 to 3 post-contrast T1-weighted acquisitions, with k-space centered at approximately 90–120 s after contrast injection for the first post-contrast images [103].

Tumor Characterization By the pharmacokinetic information on tissue vasculature available through DCE-MRI, breast lesions can be characterized noninvasively: a highly vascularized and highly proliferative cancer with a relatively low extracellular, extravascular space will show a fast signal increase (reflected by high k^{trans} or plasma flow values), followed by an early signal decrease, the typical washout curve (quantified by a low v_e or distribution volume). On the other hand, poorly vascularized benign lesions with a high extravascular extracellular space will show a slow (relfected by low k^{trans} or plasma flow values) and persistent (reflected by high v_e and distribution volume values) signal increase. Depending on the underlying pathology, variations of these characteristics occur (Fig. 17.7). These basic phenomena were first described by Kaiser [117, 118]. Empirical evidence for quantitative DCE-MRI in breast lesion characterization is difficult to assess in general due to the wide variety of approaches. Huang and colleagues [114] found that low K^{trans} values could avoid biopsy and proposed a potential K^{trans} cutoff to decrease false-positive MR examinations. Li and colleagues [119] evaluated the value of morphologic examination in conjunction with dynamic contrast-enhanced MRI (DCE-MRI) for more precise diagnosis of breast cancer, as well as their correlation with angiogenesis and proliferative activity. They found that K^{trans} and k_{ep} values were significantly higher in invasive ductal carcinoma and DCIS than ductal dysplasias.

The official BI-RADS approach is a simplification of the pharmacokinetic approach, accounting for inconsistent techniques used worldwide: the initial enhancement is classified as either slow, intermediate, or fast and the delayed enhancement curve type as either persistent, plateau, or washout. A highly aggressive cancer will thus present with a fast washout curve, while a slow-growing invasive lobular cancer can present a slow and persistent enhancement [44, 104]. However, considerable overlap of semiquantitative kinetic curve types among benign and malignant lesions exists [120]. Plateau curves might be seen in both malignant and benign lesions and are thus not decisive for differential diagnosis [80, 117, 121–124]. Rarely, washout can be associated to benign conditions, such as papilloma [125]. Furthermore, less common subtypes of invasive breast cancers, such as inflammatory carcinomas or lobular carcinomas, can show an enhancement pattern, which differs from the classical malignant type. Lobular invasive cancer demonstrates a washout enhancement pattern less frequently compared to invasive ductal cancer [79, 126]. Ductal carcinoma in situ (DCIS) is a biologically very heterogeneous entity

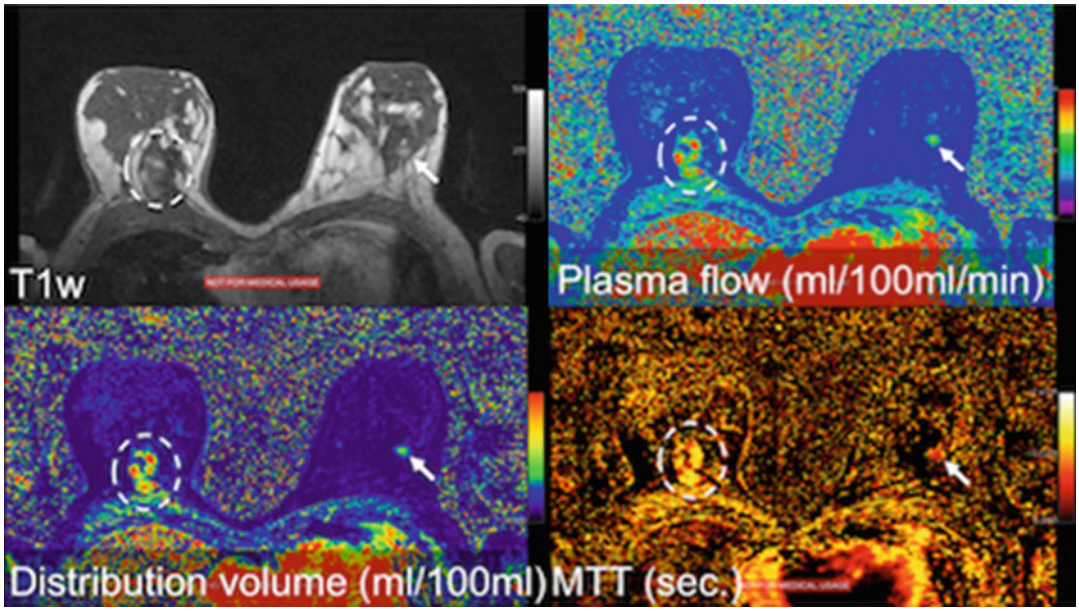


Fig. 17.7 DCE-MRI scan of a 48-year-old woman presenting with two lesions. DCE-MRI data was acquired using a 1.5 T device with a 16-channel dedicated bilateral breast coil at a temporal resolution of 6 s. at full breast coverage with a spatial resolution of 1.5 by 1.5 by 1.5 mm. Two nonpalpable lesions are present. The bigger lesion in the right breast (*dashed circle*) shows a high perfusion (plasma flow map, *upper right*), while the distribution volume is also high (*upper left*) as reflected by a long mean transit time (MTT, *lower right*). This combination is rather atypical for a malignant lesion; here, one would expect a high plasma flow, a low distribution volume, and a low mean transit time in such lesions. Biopsy revealed a highly vascularized fibroadenoma, consistent with the DCE-MRI findings. The smaller lesion on the left side (*white arrow*) shows an intermediate plasma flow, combined with an intermediate distribution volume and a low mean transit time. Findings are in line with a malignant lesion showing intermediate proliferation rate. In case of a fibrous fibroadenoma, one would expect a high distribution volume and longer MTT. Histology revealed a hormonal receptor-positive G2 invasive ductal carcinoma

exhibiting highly heterogeneous enhancement patterns with varying kinetic characteristics: in only 61–69 % of the DCIS lesions, a washout pattern has been found [127–129]. Evaluation of enhancement patterns in lesions suspicious for DCIS is, therefore, not regarded as a reliable basis for their differential diagnosis [130].

Tumor Prognosis Despite improving delineation between benign and malignant lesions, enhancement analysis correlates also to prognostic factors [131]. Specifically, a number of studies have demonstrated that high-grade breast cancers or triple-negative cancers show different perfusion parameters, namely, stronger and faster initial enhancement and higher rate of washout compared with low-grade breast cancers or luminal-type cancers [92, 108, 114, 132]. These surrogate markers hint at a poorer prognosis, and their connection to DCE-MRI parameters let assume that also DCE-MRI is a direct biomarker for breast cancer patient's outcome. Baltzer et al. could demonstrate a direct connection between computer-assisted quantification of a breast cancer lesion's worst washout and the individual risk of developing metachronous metastases [133] (Fig. 17.8). They could differentiate a group from patients at very low risk from patients at risk of developing metastases. It does then not astonish that other authors showed that disease-specific survival in breast cancer is associated to more malignant enhancement kinetics [131, 132, 134–137].

Therapy Monitoring Another application of DCE-MRI is in women receiving neoadjuvant chemotherapy. It has been demonstrated that MRI is superior to clinical assessment and

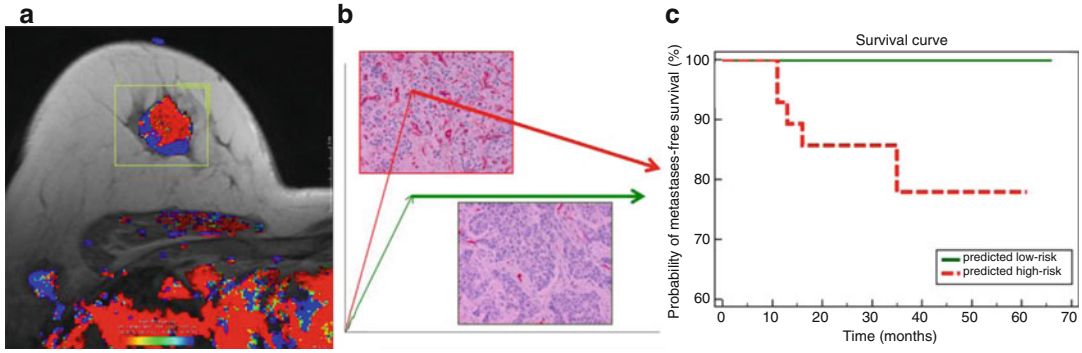


Fig. 17.8 Direct association between DCE-MRI and patient outcomes as demonstrated by Baltzer et al. [133]: (**a**) computer-assisted semiquantitative lesion analysis. Every voxel passing an initial threshold for early enhancement is color coded according to its initial enhancement and delayed enhancement curve type following BI-RADS criteria. The computer then automatically detects the most suspicious washout curve in the *yellow rectangular* volume of interest indicated by the user. Curve types are extracted, reflecting tumor biology (**b**): a breast cancer with higher proliferative activity and microvessel density (confirmed by immunohistochemical staining in this example) shows a more aggressive curve type (*red*) as compared to a less vascularized cancer (*green*). Baltzer et al. demonstrated that by the semiquantitative amount of washout, patients at high-risk for developing metachronous metastases (**c**, *red*) could be differentiated from low-risk patients (**c**, *green*)

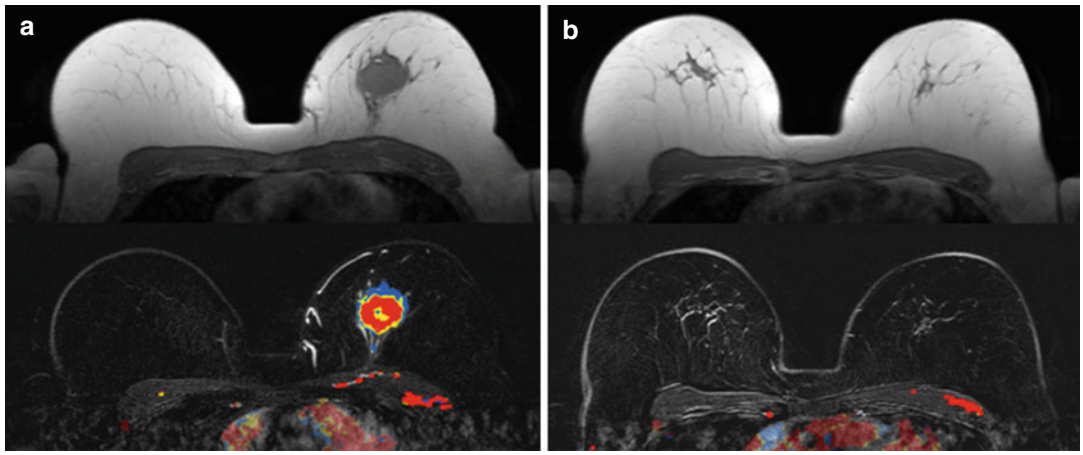


Fig. 17.9 MRI before (**a**) and after (**b**) neoadjuvant chemotherapy in a 51-year-old patient with her-2-enriched breast cancer. The parametric map of semiquantitative enhancement curve types defined by relative initial enhancement (signal intensity) and relative washout (*color*) shows a highly aggressive breast cancer lesion before treatment (**a**), demonstrated by a high percentage of bright red voxels, corresponding to fast washout enhancement curves. After chemotherapy, the lesion shows a substantial reduction in size, combined with the absence of any threshold-passing enhancement indicated by the absence of color-coded voxels on the parametric map

conventional imaging in measuring response to treatment and in predicting pathologic complete response (pCR) [138]. MRI has the unique ability of a volumetric evaluation of tumor size as well as a quantitative assessment of enhancement, which is reflective of the intra-tumoral microvascularization. As this vascularization is responsive to neoadjuvant therapy, a number of reports have demonstrated that DCE-MRI provides diagnostic information to assess breast cancer response to treatment (Fig. 17.9). Regarding quantitative pharmacokinetic parameters, decreases in k^{trans} [139, 140], k^{ep} [140, 141], and the relative initial enhancement [139, 141] are associated with treatment response. Recently, it has been demonstrated that simple analysis of enhancement curve

types contains similar accurate diagnostic information for prediction of treatment response as compared to more advanced pharmacokinetic parameters. Although not fully quantitative by nature, simple time curve-type analysis is faster and does not require complex post-processing [142]. It should be mentioned that besides technical factors, several additional criteria could influence the accuracy of DCE-MRI in the neoadjuvant setting. The intrinsic molecular subtypes show distinct different responses [143–145], and also the prediction of pCR is related to the molecular subtypes. HER2-positive and triple-negative cancers more often demonstrate a complete response than hormone receptor-positive breast cancers [146–148]. Additionally, the accuracy of the assessment of response to neoadjuvant chemotherapy with MRI is dependent on the type of neoadjuvant chemotherapy regimen used. For instance, taxane-containing chemotherapy has an antiproliferative and antiangiogenesis effects, thus reducing contrast enhancement of both benign and malignant breast lesions and even the BPE [149].

In conclusion, there is broad consensus in the literature concerning the diagnostic value of DCE-MRI measurements for noninvasive characterization and prognostication of breast lesions as well as for therapy monitoring in breast cancer patients undergoing neoadjuvant therapies. However, the broad variety of methodologies to analyze dynamic enhancement data does yet preclude cutoff recommendations that could be used in prospective studies or in everyday clinical practice.

17.3.2.2 Imaging of Tumor Microstructure with Diffusion-Weighted Magnetic Resonance Techniques

Diffusion-Weighted Imaging (DWI)

Diffusion-weighted imaging (DWI) quantifies the random, microscopic motion of protons in tissues (i.e., Brownian movement) and noninvasively describes tissue microstructure in vivo [150]. Tumor tissue is characterized by a high cellular density, due to the increased cellular turnover, and by a high proteolytic activity which enables cancers to degrade the tissue structure and spread unhindered through adjacent tissues. DWI is able to show the degree of water diffusion in tissue that is inversely correlated with tissue cellularity and the integrity of cell membranes [151, 152]. Tumor cells, and the chronic inflammatory reaction to proteolysis through desmoplastic tissue changes, lead to a relative or absolute reduction in extracellular water content, thus limiting extracellular water diffusion [152, 153]. Molecular diffusion can be calculated by assessing the signal attenuation that occurs at least at two different b values, depending on the timing and amplitude of diffusion gradients, to measure what is known as the apparent diffusion coefficient (ADC) value in square millimeters per second. ADC is defined as the average area covered by a molecule per unit time [84, 150] and closely reflects biological tumor characteristics, such as cellularity and water content. Typically, low ADC values in breast tissue are associated with malignancy as a result of the increased restrictions in water diffusion in highly cellular regions, in contrast to normal glandular tissue or nonrestricted diffusion in cysts [154–156]. In biological tissue, microscopic motion from diffusion and perfusion both contribute to ADC. However, microstructural changes that influence water diffusion in neoplastic breast tissue are still poorly understood [151]. ADC values derived from DWI are also strongly influenced by the choice of b value. Bogner et al. [157] compared the diagnostic quality of different DWI protocols with regard to ADC accuracy, ADC precision, and DW contrast-to-noise ratio (CNR) for different types of breast tumors. Those authors found that the optimum ADC determination and DWI quality could be achieved with a combination of b value protocol of 50 and 850 s/mm^2 (Fig. 17.10). This provided diagnostic accuracy of 96 % for differentiation of benign and malignant breast tumors. A systematic meta-analysis from Dorrius et al. [158], evaluating 26 published articles on DWI studies, confirmed that the wide variety of b value combinations applied in different studies significantly affects the ADC of breast lesions. These investigators suggested the

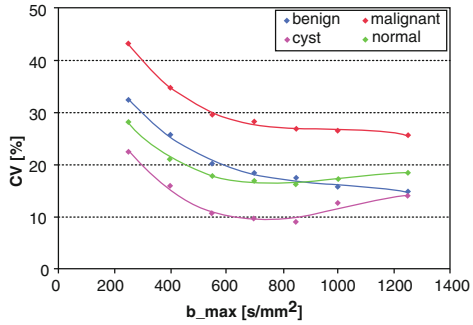


Fig. 17.10 Proper choice of bmax (polynomial trendlines) improves fitting precision, as illustrated by decreasing mean coefficient variation (CV). ADC calculation precision improves only slowly for benign or malignant lesions with b values higher than 850 mm²/s, while both normal tissue and cysts showed maximum precision at this point (Reproduced from Bogner et al. [157])

combinations of $b=0$ and 1,000 s/mm² to achieve the best improvement in differentiation between benign and malignant lesions. In general, the specificity of DWI is higher (75–84 %) than that achieved in DCE-MRI (67–72 %), especially when b values >600 s/mm² are used. Prior studies have shown promising results from DWI images and ADC values over conventional DCE-MRI in the characterization of breast lesions and further proved that the additional use of DWI significantly improved the diagnostic accuracy of DCE-MR examinations [85, 159–161]. Various investigations [160, 162–164] have already pursued the idea of incorporating the ADC thresholds into the standardized system of reporting the breast imaging examinations (BI-RADS Breast Imaging Reporting And Data System) [44]. Partridge and colleagues demonstrated an improvement in diagnostic accuracy over DCE-MRI alone by using an ADC threshold combined to conventional MRI assessment. Namely, they demonstrated an improvement in PPV by using DWI combined with DCE-MRI, with a 33 % reduction of false positives. Pinker et al. [165] developed a BI-RADS-adapted system for breast MR imaging, useful in clinical practice, by improving significantly the specificity of the readers (89.4 %). Baltzer et al. [166] investigated the improvement in specificity of breast MRI by integrating ADC values with DCE-MRI using a simple sum score. The additional integration of ADC scores achieved an improved specificity (92.4 %) compared to DCE-MR only reading (specificity of 81.8 %), without causing false-negative results.

Tumor Characterization and Prognosis DWI and ADC values have demonstrated an excellent ability to differentiate benign and malignant breast lesions [151, 152]. A meta-analysis of 13 studies that evaluated the diagnostic performance of DWI in 615 malignant and 349 benign lesions reported a pooled sensitivity of 84 % (95 % CI, 82–87 %) and a specificity of 79 % (95 % CI, 75–82 %) [167]. Several authors have recommended cutoff thresholds to discriminate benign and malignant lesions [156, 168–171]. A very low ADC value ($<1.2\times10^{-3}$ mm²/s) is specific for malignancy (Fig. 17.11), and high ADC values ($>1.4\times10^{-3}$ mm²/s) are likely expected for benignity (Fig. 17.12). Intermediate ADC values are unspecific and, thus, less helpful in clinical practice. Rigid interpretation criteria, such as single cutoff values, are not feasible due to an overlap of ADC values in benign and malignant lesions. Analogous to serum blood markers, such as serum C-reactive protein concentrations, different levels of suspicion can be raised based on ADC values: low (likely benign), medium (indeterminate), and high (likely malignant) [166]. Thus, high ADC thresholds ($>1.4\times10^{3}$ mm²/s) can reliably exclude malignancy and help to obviate unnecessary false-positive MRI-guided biopsies [172]. ADC values can also be used as a noninvasive imaging biomarker to predict the probability of cancer invasiveness [173]. Bickel et al. found that DCIS showed significantly higher ADC values ($<1.24\times10^{-3}$ mm²/s) compared to invasive cancers ($<1.14\times10^{-3}$ mm²/s). These results have great clinical implications, since they could result in a change in the management of noninvasive cancers, often overdiagnosed and therefore overtreated [174]. ADC values have also shown differences according to tumor subgroups. Martinicich et al. [175] found that breast cancers characterized by high cellularity (or a higher number of mitoses) showed a lower ADC value at DWI. It is also plausible that the median ADC values correlate to HER2/neu status, with ADC values significantly higher in the subgroup of HER2-/neu-enriched

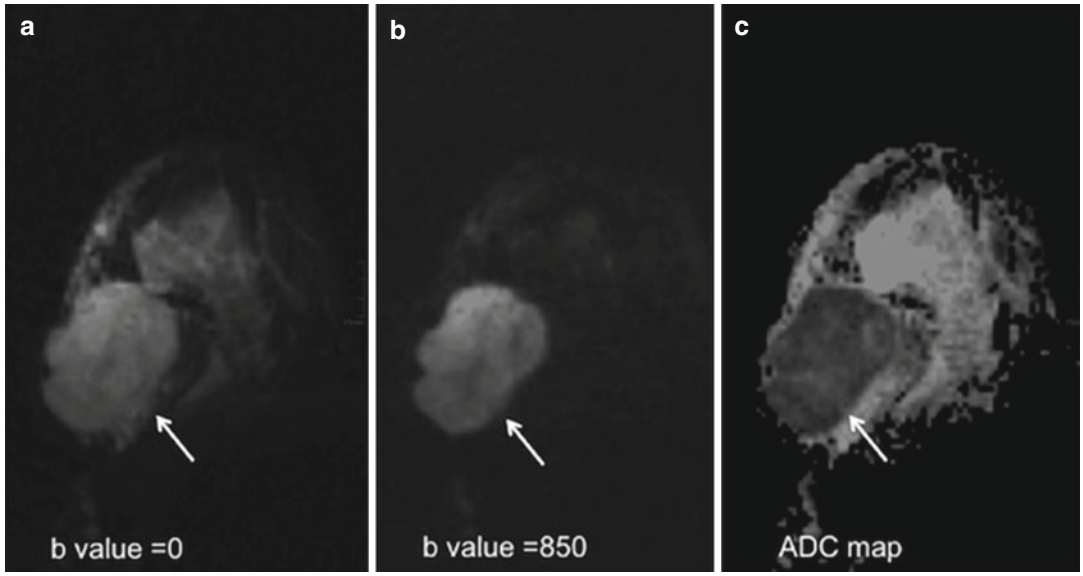


Fig. 17.11 Invasive ductal cancer G3 (triple negative) in (0.998×10^{-3} mm²/s) highly suggestive of malignancy a 29-year-old woman at 2 o'clock of the right breast. (**a**, **b**) (BI-RADS 5) The lesion is marked with an *arrow*. Diffusion weighing at different *b* values and the respective (**c**) ADC values

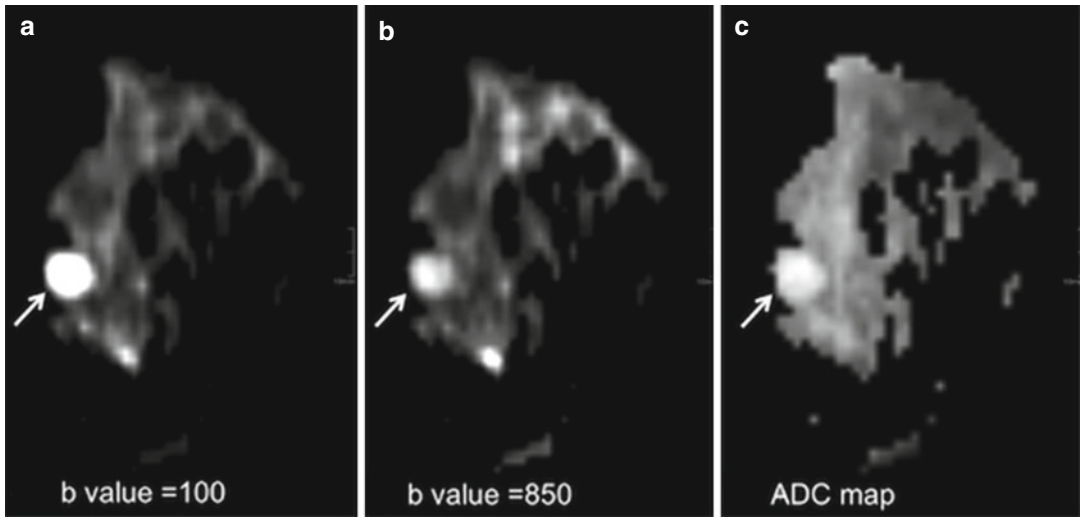


Fig. 17.12 Cystic lesion in a 45-year-old woman at 2 o'clock in the right breast. (**a**, **b**) Diffusion weighing at different *b* value. (**c**) The lesion (*arrow*) demonstrated ADC values (2.675×10^{-3} mm²/s) above the threshold for malignancy. The lesion was classified as BI-RADS 2 (benign)

compared to HER2-/neu-negative tumors [168, 175]. Similar observations have been published in triple-negative breast cancers, associated with higher ADC values on DWI compared to ER+ and HER2-/neu-enriched tumors [21, 176–178]. In contrast, mucinous carcinoma usually presents ADC values similar to those of benign lesions ($>1.4 \times 10^{-3}$ mm²/s), most likely due to the presence of both low cellularity and mucin-rich compartments [179]. No significant correlation between ADC values and tumor grading has been found as yet [173, 180, 181].

Therapy Monitoring DWI has been proposed as a "biomarker for treatment response in oncology" [12]. In patients undergoing anticancer therapy, ADC values are very sensitive to changes in tumor cellularity and necrosis, and early ADC changes in responding patients may predict tumor regression even months ahead of the currently employed imaging techniques [12, 182]. The cytotoxic effect associated with the administration of these agents increases the Brownian movements in the damaged tissues, reflecting an increase in ADC value that may appear earlier than lesion size changes [182, 183]. Park et al. [184] studied the potential of DWI in the prediction of response to neoadjuvant chemotherapy in patients with breast cancer and found that patients with breast cancer and a low pretreatment ADC tended to respond better to chemotherapy, yielding a sensitivity of 94 % and a specificity of 71 % when the ADC cutoff pretreatment was very low ($<10^{-3}$ mm^2/s). Richards et al. [185] evaluated the accuracy of ADC in predicting the response to neoadjuvant chemotherapy according to breast tumor phenotypes. In 118 women undergoing neoadjuvant chemotherapy for locally advanced breast cancer, the authors found that pretreatment tumor ADC values differed between intrinsic subtypes and were predictive of pathologic response in triple-negative tumors. However, the assessment of prechemotherapy ADC has yielded divergent results in the literature, and its role as potential biomarker for predicting the response to treatment in locally advanced breast cancer needs further validation [153, 182, 183, 186].

Advanced Methods of Breast DWI

Diffusion Tensor Imaging (DTI) Considered as an extension of DWI, diffusion tensor imaging (DTI) provides information about water motion in six or more directions, characterizing the motion of water in more detail [86, 187]. DTI provides measurements of two parameters: mean diffusivity (MD) and fractional anisotropy (FA). MD reflects the average anisotropy, while FA describes the degree of anisotropy [87, 187]. The diffusion of water molecules in the mammary glandular/ductal system presents a particular example of restricted movement: free diffusion parallel to the walls of the ducts and lobules and restricted diffusion in the perpendicular directions, leading to an anisotropic diffusion [188]. Therefore, any changes of this tissue structure, through either benign or malignant tumor growth, should be reflected by changes in diffusion anisotropy [86, 187, 188]. Accordingly, based on histopathological knowledge, most breast diseases result in decreased structuring compared to healthy tissue. Partridge et al. [189] investigated whether DTI measures of anisotropy in breast tumors are different from those of normal breast tissue and whether these measures could improve the discrimination between benign and malignant lesions. They found that diffusion anisotropy is significantly lower in breast cancers than in normal tissue, which may reflect alterations in tissue organization, but cannot reliably differentiate between benign and malignant. Baltzer et al. [86] proved that DTI can visualize microanatomical differences between malignant and benign breast lesions as well as breast parenchyma, but FA did not show an incremental diagnostic accuracy compared to ADC (Fig. 17.13). Although the results concerning the diagnostic accuracy of FA are divergent [86, 159, 187], DTI has proven to be diagnostically relevant in cancer detection [86, 188, 189], having the potential to serve not only as an adjunct method to DCE examination but also as an alternative method when DCE imaging is contraindicated.

Diffusion-Weighted Kurtosis Imaging (DKI) Diffusion-weighted kurtosis imaging (DKI) is able to quantify the Brownian incoherent motion and microperfusion or blood flow that shows non-Gaussian phenomena [190], which are more characteristics of the diffusion-weighted MR signal in living tissues. Compared to DWI, DKI has demonstrated substantially higher sensitivity in breast cancer detection and characterization [191, 192]. By enabling differentiation and characterization of breast lesions and providing insights into microstructural complexity, DKI may be used as a biomarker for the prediction of histologic malignancy grade and proliferative activity of breast cancer. Sun and colleagues [192] investigated the diagnostic accuracy of

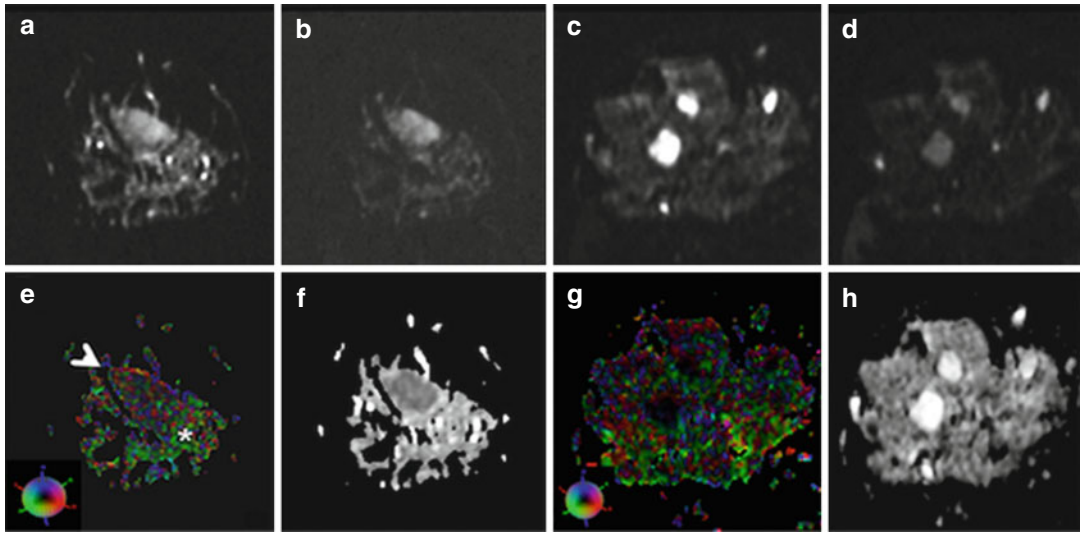


Fig. 17.13 (**a**–**d**) Comparing *b* 0 s/mm^2 with the *b* 1,000 s/mm^2 image in two different lesions, one malignant (**a**, **b**) and one benign (**c**, **d**). In both cases diffusion-dependent signal loss of fibroglandular tissue is stronger than that of the lesions. For the malignant lesion, the corresponding ADC map (**f**) shows low signal intensity. Color-coded FA map (**e**) shows the lesion (*arrowhead*) a little more hypointense compared to breast tissue (*asterisk*), the lesion does not have a predominant diffusion direction. These findings are typical for invasive breast cancer, which was verified by histopathology (*invasive ductal cancer G2*). For the benign lesion, the corresponding ADC map (**h**) shows high-signal intensity. Color-coded FA map (**g**) demonstrates very low anisotropy (hypo-intensity) inside the lesion without predominant diffusion direction. These findings are typical for *fibroadenoma*, which was verified by histopathology (Reproduced from Baltzer et al. [86])

DKI in assessing breast cancer clinicopathological factors. They found that the quantitative analysis of DKI-derived parameters significantly improved the specificity of the characterization of benign versus malignant lesions compared to ADC values and DWI. In addition, DKI correlated with tumor histologic grades and Ki-67 tumor expression. The authors also proved an increase in diagnostic accuracy when diffusivity and kurtosis were added to DCE-MR, enhancing the potential role of DKI in the management of breast cancer patient, especially in monitoring the response to treatment.

17.3.2.3 Imaging of Tumor Metabolism with Proton Magnetic Resonance Spectroscopic Imaging (^{1}H MRS)

Proton magnetic resonance spectroscopic imaging (^{1}H-MRSI) is a noninvasive diagnostic tool that enables the assessment of biochemical tissue properties and provides metabolic information about tumors [193, 194]. ^{1}H-MRSI detects the choline peak at 3.2 ppm, which is not regularly identified in healthy breast tissue at field strengths of 1.5 and 3 T [195–198]. Choline metabolites, such as choline, phosphocholine, and glycerol-phosphocholine, simply referred to as total choline (tCho), are components of cell membranes [193, 199–201]. The increased levels of tCho detected in cancers can be explained by the increase in tumor cell turnover [154, 202–205]. It has been demonstrated that the additional application of ^{1}H-MRSI helps in the characterization of breast tumors [206, 207] (Fig. 17.14). A variety of spectroscopic techniques at different field strengths can be used. Higher field strengths provide higher signal-to-noise ratios, and spectroscopic imaging allows spatially resolved examinations with small voxels [208]. Although a diagnostic benefit may thus be expected, comparison with studies at 1.5 T have not revealed such an advantage as yet [206].

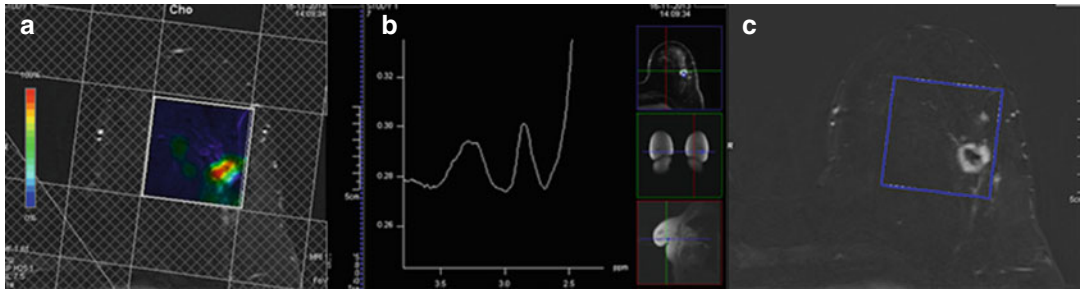


Fig. 17.14 Invasive ductal carcinoma G3 in the 3 o'clock position of the left breast in a 40-year-old woman: (**a**) On proton magnetic resonance spectroscopic imaging, there was a choline peak at 3.2 ppm (signal-to-noise ratio 10.5). (**b**) Metabolic map (**c**) DCE-MRI demonstrated an irregularly shaped mass lesion with spiculated margins. The lesion was classified as malignant, BI-RADS_ 5 (highly suggestive of malignancy)

Baltzer et al. [209] estimated the diagnostic performance of ^{1}H-MRSI in differentiating benign from malignant lesions in a recent meta-analysis of 19 studies, performed at 1.5 T and at 3 T. They found a pooled sensitivity and specificity of 73 % and 88 %, respectively, with a homogenous high specificity and a variable sensitivity in the studies (42–100 %). ^{1}H-MRSI showed limitations in the diagnosis of early breast cancer and small breast tumors because of low sensitivity and heterogeneous performance in non-mass-like enhancing lesions. Gruber et al. [195] evaluated the diagnostic accuracy of a high-spatial-resolution 3D ^{1}H-MRSI protocol at 3 T, designed to cover a large fraction of the breast in a clinically acceptable measurement time of 12–15 min, for the differentiation of benign and malignant breast lesions, on the basis of choline (Cho) signal-to-noise ratio (SNR) threshold levels. They concluded that 3D-MRS at 3 T is possible in patients with breast lesions, with excellent data quality, and has the potential to become a valuable adjunct to DCE-MRI for the differentiation of benign and malignant breast lesions. Furtherore, Pinker et al. [85] found that the additional value of ^{1}HMRS to DW and DCE-MR yielded a diagnostic accuracy significantly higher than DCE-MRI and also resulted in the elimination of false-negative lesions and significantly reduced the false positives.

Qualitative and quantitative in vivo measurements of choline metabolites have been used as a diagnostic test in the work-up of neoplastic breast lesions [206]. Quantitative assessment of the recorded tCho, indicator of breast malignancy, can be performed either intra-voxel, with water used as an internal reference, or using an external standard. For accurate measurement of the relative amount of a metabolite, the resonance area (i.e., the calculated integral of the number of molecules measurable by ^{1}H-MRSI) must be corrected for the relaxation properties of the signal and not overlap with any other resonances in the spectrum [207]. Qualitative assessment of the recorded tCho involves the subjective determination by an observer as to whether a distinct resonance at approximately 3.2 ppm is present [194]. The sensitivity of qualitative tCho detection, and hence, cancer diagnosis, is significantly decreased for smaller cancers, which might be diagnosed as benign due to insufficient tCho signal for detection [194].

Tumor Characterization and Prognosis The human breast demonstrates a number of distinct resonances attributable to choline, glycerides (esters of fatty acids and glycerols), saturated and unsaturated fatty acids, and water. The variations in cellular metabolites that occur during tumor development can be monitored using ^{1}H-MRSI [194, 196]. Malignant disease has a typical peak at 3.23 ppm, whereas benign lesions, such as fibroadenomas, have a resonance at the frequency of 3.28. The same spectrum can be seen in healthy volunteers and lactating women [203, 207]. Therefore, based

on the different choline peaks, it is plausible to differentiate benign and malignant lesions and also use ^{1}H-MRSI as a biomarker of tumor aggressiveness. Shin et al. [210] showed that tumor tCho measures were significantly higher for invasive ductal carcinoma versus in situ cancers and correlated with several prognostic factors, including nuclear grade, histologic grade, and estrogen receptor status. To date, tCho-containing compounds are not only considered for the expression of increased proliferation rather as general hallmarks of malignant transformation [200, 211]. Therefore, the role of ^{1}H-MRSI for breast cancer risk stratification has been investigated [81, 212, 213]. Recently, Ramadan and colleagues [214] have proved that healthy breast tissue in patients who are carriers of a mutation in one of the two genes related to breast cancer (BRCA-1 and BRCA-2) is likely to differ from each other and from non-mutation carriers regarding levels of triglycerides, unsaturated fatty acids, and cholesterol, even in the absence of any other imaging findings. Further studies are needed to prove these hypotheses, which, if supported by comprehensive results, could open unique frontiers for the screening programs for high-risk women.

Therapy Monitoring ^{1}H-MRSI of the breast is a promising tool for the assessment of the direct effect of anticancer agents. Jagannathan et al. demonstrated that, after completion of neoadjuvant chemotherapy, a change in the tCho concentration was observed and confirmed by histopathology [215–217]. Meisamy et al. [196] demonstrated that ^{1}H-MRSI of the breast was able to detect a change in Cho concentration from baseline (before receiving therapy) within 24 h of administration of the first dose of the regimen. This change had a statistically significant positive correlation with a change in the final size of the lesion ($p = 0.001$). Therefore, it can be expected that the addition of ^{1}H-MRSI of the breast will offer a substantial advantage over contrast-enhanced MRI of the breast alone in the prediction of response to neoadjuvant chemotherapy.

17.3.2.4 Multiparametric Magnetic Resonce Imaging

MRI of the breast is an established noninvasive imaging modality for the detection, characterization, and staging of breast tumors. DCE-MRI is the backbone of any given MRI protocol. DCE-MRI provides mainly morphologic and, to some extent, functional information, i.e., perfusion and vascular integrity, which results in an excellent sensitivity and good specificity [54, 99, 213, 218]. Recently, several functional MRI parameters have been assessed for breast imaging. Available data suggest that functional MRI parameters can provide detailed information about the cellular, microcellular, and subcellular biological systems and, thus, may provide additional specificity [54, 99, 103, 165, 208, 213, 218]. The combination of DCE-MRI and other functional MRI parameters is defined as multiparametric (MP) MRI. MP MRI of the breast simultaneously and noninvasively acquires multiple imaging biomarkers at different levels and, thus, has the potential to significantly improve breast cancer diagnosis, staging, and assessment of treatment response. In several recent studies, the diagnostic value of MP MRI using two parameters, i.e., DCE-MRI and DWI, was assessed. It was demonstrated that MP MRI with two parameters achieves an increased diagnostic accuracy in breast cancer diagnosis [103, 159, 162, 165, 166]. Recently Pinker et al. compared the diagnostic accuracy of DCE-MRI as a single parameter to MP MRI with two (DCE-MRI and DWI) and three (DCE-MRI, DWI, and ^{1}H-MRSI) parameters in breast cancer diagnosis (Figs. 17.14 and 17.15). MP MR with three MRI parameters yielded significantly higher AUCs (0.936) compared to DCE-MRI alone (0.814) ($p < 0.001$). MP MRI with just two parameters at 3 T did not yield higher AUCs (0.808) than DCE-MRI alone (0.814). MP MRI with three parameters resulted in elimination of false-negative lesions and significantly reduced the false positives ($p = 0.002$). The authors concluded that MP MRI with three parameters increases the diagnostic accuracy of breast cancer compared to DCE-MRI alone and MP MRI with two parameters.

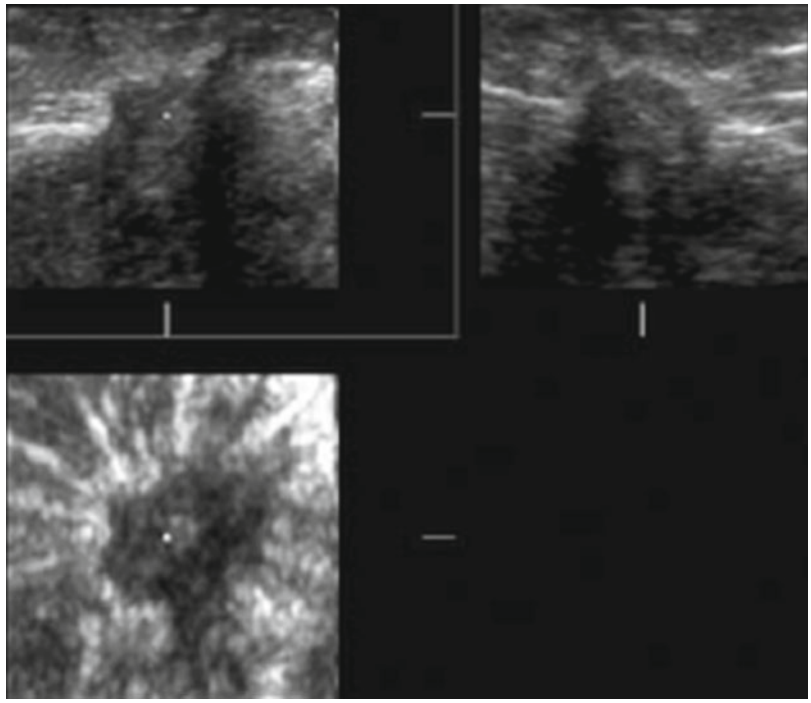


Fig. 17.15 Three-dimensional (3D) ultrasound scan of a malignant breast lesion. The plane parallel to the array of the probe, equivalent to a conventional two-dimensional picture, is displayed in the left upper quadrant. The reconstructed plane perpendicular to this plane is shown in the right upper quadrant and can be achieved in conventional ultrasound by a 90° turn of the probe. The "coronal" plane, parallel to the skin and depicted in the left lower quadrant, is unique to 3D ultrasound and cannot be achieved with a conventional probe. A strong retraction phenomenon is visible in this plane (Reproduced from Watermann et al. [219])

17.3.3 New Frontiers in Breast Imaging

17.3.3.1 Quantitative Imaging Biomarkers with Ultrasound

Ultrasound (US) is an established adjunct in breast imaging for further characterization of mammographically suspicious breast lesions, for lesion detection in women with dense breasts, and for the assessment of tumor response and is the method of choice in younger women, due to the lack of ionizing radiation [220–222]. B-mode US demonstrated a high diagnostic performance with high accuracy and sensitivity, but suffers from varying sensitivity, with ranges between 20.7 and 98.5 % [223–226]. In addition, interobserver variability in breast US examinations poses a further limitation [227, 228]. High-resolution 2D and 3D B-mode imaging allows an accurate depiction of anatomy and the pathomorphology of the tumor and its interactions with the surrounding tissue [219, 229] (Fig. 17.15). Automated 3D breast ultrasound (ABUS) has gained a lot of interest recently especially in screening of dense breasts, where the sensitivity of mammography is poor [230]. Although a promising tool for its high reproducibility [231], ABUS is time-consuming and subtle abnormalities may be missed [232, 233]. Apart from B-mode US, several US-based techniques have been implemented to collect functional and kinetic information about breast tumors [234, 235]. In the following paragraphs, we focus on these techniques, such as elastography, Color and Power Doppler (CD and PD), and contrast-enhanced US (CEUS). Furthermore, we briefly describe a new promising technology, i.e., molecular US imaging, as a method of noninvasive imaging of tumor angiogenesis using microbubbles targeted to molecular markers as contrast agents [236, 237].

Elastography

The use of elastography in breast ultrasound is based on the fact that different tissues are expected to respond differently based on the amount of compression [225, 238, 239]. Generally, breast cancer tissue is harder than the adjacent normal breast tissue [238]. Given that compression of the breast using a probe results in displacement (strain) of the underlying tissues, stiffer tissues are expected to present a smaller strain than softer ones, and thus, calculation of the strain induced by probe compression can help differentiate stiffer from softer components, namely, breast cancer from surrounding healthy breast parenchyma [223, 224, 238]. Elastography allows the assessment of qualitative and quantitative information about solid breast lesions [83]. To characterize breast lesions and to allow differentiation between malignant and benign lesions, Itoh et al. [238] introduced a five-point elasticity score. The authors found that cutoff points between 3 and 4, to rule out malignancy, yielded a sensitivity of 86.5 % and a specificity of 89.8 %. Further efforts have been made in order to quantify the elastographic results, using strain ratio (i.e., the fat-to-lesion ratio, indicating the stiffness of a lesion) [240] and shear wave (SW) elastography. Strain elastography calculates the degree of deformation (strain) in a direction perpendicular to the tissue surface in response to an externally applied force and, thus, is limited by significant interobserver variability [239, 241]. SW elastography is less user dependent, with better reproducibility, because the operator does not apply any kind of pressure on the breast, and a short duration, high-intensity acoustic "pushing pulse" is used for tissue compression instead [242–244]. Then, either a series of diagnostic intensity pulses, used to track the subsequent displacement of the tissue (acoustic radiation force impulse (ARFI) technology), or a very fast US acquisition sequence that captures the propagation of shear waves (supersonics imaging) follows the "pushing pulse" [245]. Both technologies provide quantitative measurements of tissue stiffness, either in m/s of SW velocity (ARFI) or in kPa (supersonics).

Tumor Characterization and Prognosis

Several authors have proved that elastography can help characterize breast lesions [229, 240, 242]. Ophir et al. [246] first described the potential use of elastography to differentiate breast cancer from benign lesions. Garra et al. [247] presented the first clinical study about the value of elastography in distinguishing benign and malignant diseases of the breast and reported a sensitivity of 100 % and a specificity of 56 % for solid breast lesions. The two most important strain elastographic characteristics in evaluating breast lesions are size and stiffness criteria. Invasive breast cancers have been shown to be stiffer than benign or normal tissues and, thus, measure larger and darker than they do on corresponding B-mode images [248]. Fibroadenomas, softer than malignant lesions, appear smaller in size and brighter than on the B-mode image [241]. However, fibroadenomas sometimes have elastographic size or stiffness features that are more typical of malignancy. Such false-positive elastographic findings tend to occur in fibroadenomas that are larger than 2 cm and contain calcification [249]. Areas of ductal carcinoma in situ (DCIS) have static elastography values that fall between those seen in invasive cancer and in fibroadenomas. SW provides a color map of tissue elasticity that is superimposed on the real-time gray-scale ultrasound images, to provide a general overview of the examined area. SWs travel faster in stiffer tissue, and thus, the velocity measurements in ARFI imaging (Fig. 17.16) can help differentiate between benign (softer, lower SW velocity) and malignant lesions (stiffer, higher SW velocity). The quantitative data are also presented in the form of a color-coded map, which is superimposed on the B-mode images to provide a general overview of the examined area. Several studies have demonstrated positive results for the differentiation of benign and malignant lesions [225, 250], although the proposed cutoff values are quite divergent. Recently, ARFI technology has also demonstrated, in a preliminary study [251], a cutoff value of 3.23 m/s to show a sensitivity of 82.4 % and a specificity of 80.4 % in the

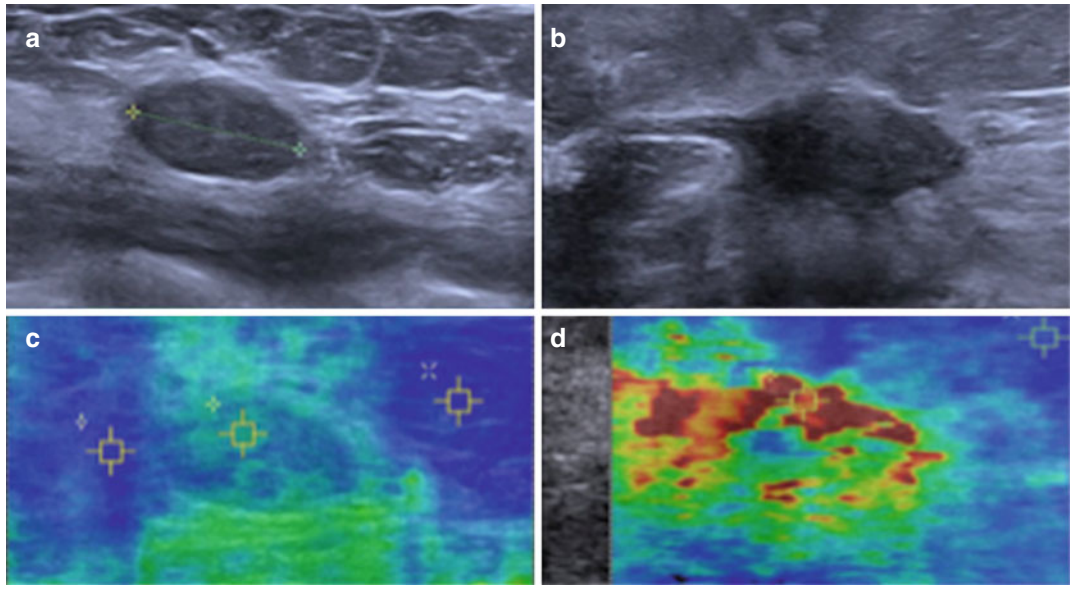


Fig. 17.16 ARFI: (**a**–**c**) Fibroadenoma in a 35-year-old woman on the right breast. (**a**) On conventional B-mode image, lesion was classified as BI-RADS category 3. (**c**) On elasticity image, the hypoechoic lesion was *blue* and *green*, indicating a softness. (**b**–**d**) Invasive ductal carcinoma G2 in a 50-year-old woman on the right breast. (**b**) On conventional B-mode image, lesion was classified as BI-RADS category 5. (**d**) The diagnostic hypothesis was confirmed due to the hardness of the tissue hypoechoic lesion, *red* and *yellow* areas

differentiation of benign and malignant breast lesions. Further studies are needed to validate these promising results.

Therapy Monitoring Both strain and SW elastography might add information to help predict tumor response to neoadjuvant chemotherapy [242, 252] and for monitoring the response to therapy, by providing insight about the altered mechanical properties of tumors and the surrounding tissue under neoadjuvant chemotherapy [253, 254].

Color Doppler (CD) and Power Doppler (PD)

One of the hallmarks of malignant tumors is neoangiogenesis, so an early hypothesis was that breast cancer could be differentiated from benign lesions by studying the vascularization features of each lesion. CD/PD [255–257] of the breast enables the depiction of increased microvascular density in breast tumors as a marker of tumor neoangiogenesis. In particular, CD/PD can allow the evaluation of potential differentiation parameters, such as the vascular density of the tumor, the pattern of the vessels inside and around the tumor, flow velocities, as well as pulsatility index (PI) and resistance index (RI) [226, 255, 256, 258, 259]. In a recent meta-analysis, Bruening and colleagues [260] evaluated 13 articles about CD and PD ultrasound, six for CD and seven for PD, in the work-up of women with a suspicious finding in the breast. For color Doppler, pooled sensitivity and specificity were, respectively, 88.5 % and 76.4 %. In the power Doppler studies, sensitivity and specificity were lower (70.8 % and 72.6 %, respectively).

Tumor Characterization and Prognosis Doppler examinations offer the possibility of quantification by acquiring a spectral flow analysis of the examined vessel and measuring systolic and diastolic peak velocities, as well as calculating indices such as the PI and RI. Although there is no consesus about threshold values, it seems that higher PI and RI are mostly found in malignant lesions [255, 261]. A further exploitation of Doppler ultrasonography has been made possible by the introduction of sonographic contrast agents, which have demonstrated promising results [262–264].

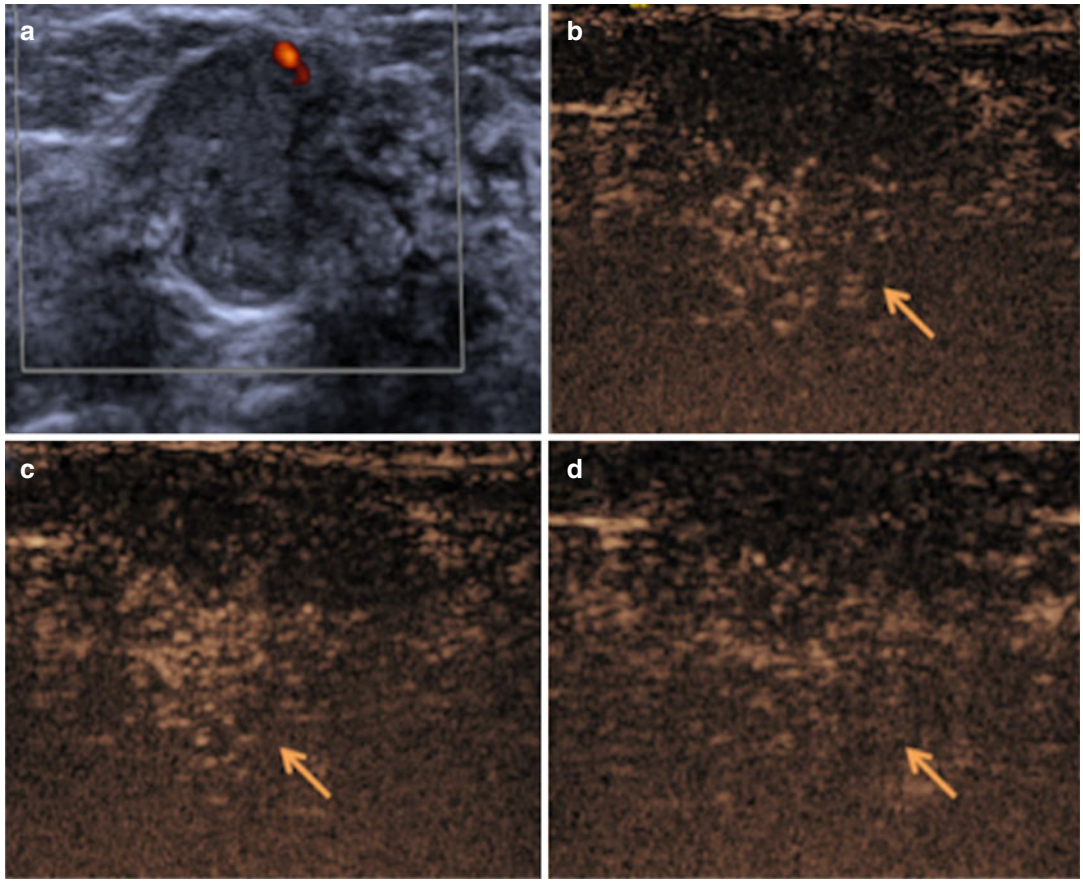


Fig. 17.17 Invasive ductal carcinoma G3 in a 46-year-old women at 6/7 o'clock of the right breast. (**a**) Color Doppler US of a 3 cm round-shaped lesion with indistinct margins demonstrates minimal internal vascularity. (**b–d**, lesion marked with *arrow*) Contrast enhancement ultrasound (CEUS) at 20 (**b**), 40 (**c**), and 65 (**d**) seconds demonstrates rapid wash-in and wash-out, kinetic characteristics likely malignant. The lesion was correctly classified as BI-RADS 5 (highly suggestive of malignancy)

Therapy Monitoring CD and PD have also found interesting applications in the prediction and monitoring of response to neoadjuvant chemotherapy. CD and PD might serve as predictive factors for tumor response to neoadjuvant chemotherapy [265], as well as in the monitoring of response, with an observed reduction in tumor vessels, PI, and RI [266].

Contrast-Enhanced Ultrasound (CEUS) and Molecular Ultrasound

The introduction of sonographic contrast agents has increased the applications for Doppler ultrasonography. The potential use of such agents in breast diagnostics has been studied, with promising results [256, 262, 263]. Huber et al. [263] have shown that benign and malignant breast lesions behave differently regarding the degree and kinetics of Doppler US contrast enhancement. The implementation of harmonic imaging has made the introduction of second-generation contrast agents possible. These require a low mechanical index, remain intact for a longer period of time, and facilitate real-time imaging through both early (arterial) and late vascular phases [267]. Thus, a more sophisticated observation of tumor kinetics has been made possible, allowing for better characterization of malignant lesions [268, 269] (Figs. 17.17 and 17.18). More recently, targeted microbubble-enhanced

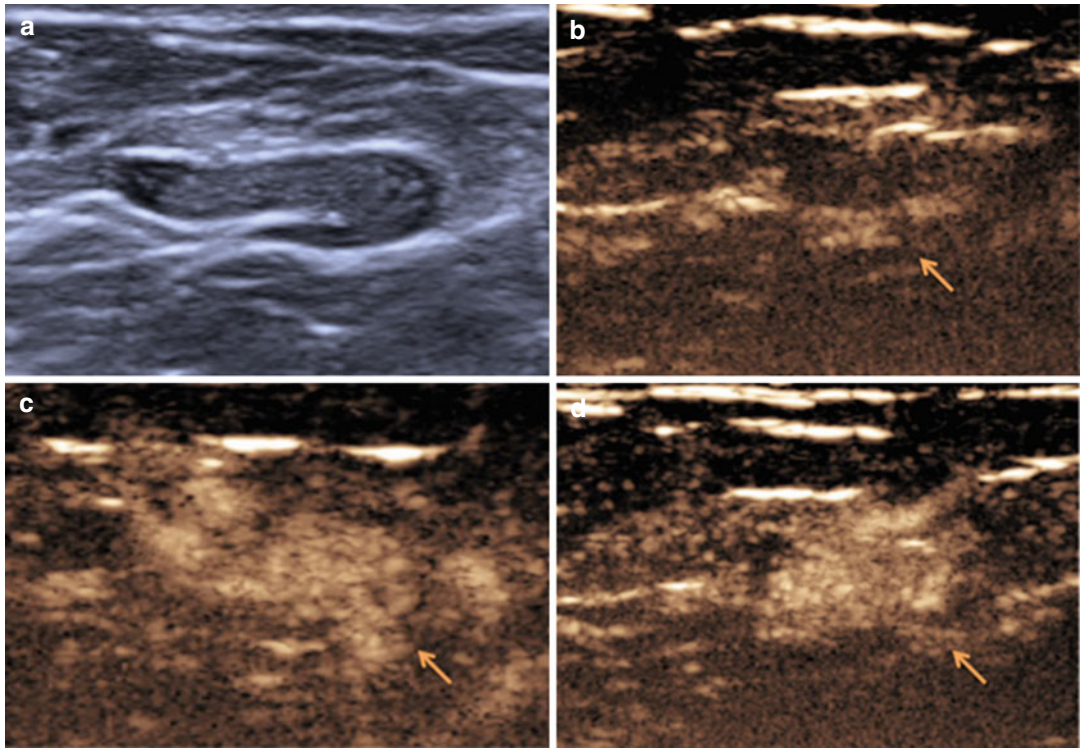


Fig. 17.18 Benign breast tumor (papilloma) in a 36-year-old woman at 6 O'clock of the left breast. (**a**) B-mode US shows a 1 cm intraductal lesion. (**b–d**, *arrow* in **b–d**) Contrast enhancement ultrasound (CEUS) at 20 (**b**), 40 (**c**), and 65 (**d**) seconds demonstrates rapid wash-in and an homogenous plateau enhancement (indeterminate). The lesion was correctly classified as BI-RADS 4a (low suspicious of malignancy)

US, i.e., molecular US, has proved to be a powerful modality with which to distinguish between normal and pathologic tissues [270, 271]. Molecular US uses contrast agents that bear adhesion ligands designed to bind tissue markers specific for a disease process. Such agents can be detected by ultrasound with a great degree of sensitivity, providing both anatomical reference information, as well as additional data, such as molecular characteristics of the interrogated region. Different ligands have been tested in various preclinical studies [271–273], and the purely vascular confinement of the microbubbles makes them ideal tools for the development of contrast agents targeted to receptors expressed on the endothelial lining of vessels [272]. Targeting microbubbles to P-selectin, vascular endothelial growth factor receptor 2 (VEGFR2), and $\alpha_v\beta_3$-integrin angiogenic molecular markers has been shown to effectively increase visualization of the tumor vasculature by 60 % over single-targeted strategies and 40 % over dual-targeted strategies in preclinical breast cancer models [274]. BR55 (Bracco, Milan, Italy) is one of the most promising VEGFR2-targeted US contrast agents that does not require, in contrast to other VEGFR2-targeted microbubbles, binding proteins like streptavidin or antibodies that are potentially immunogenic. BR55 is the first molecular contrast agent used in clinical trials [275].

Tumor Characterization and Prognosis
Benign and malignant breast lesions behave differently with regard to the degree and kinetics of Doppler US contrast enhancement [263]. A high maximum intensity of ultrasound signal during bolus transit [276] and a high regional blood flow have been described as being characteristic of carcinomas, with a PPV of 91 % [277]. Molecular US imaging has the potential to be

used as a tool to better characterized malignant lesions. Current investigations have also shown the advantages of using targeted microbubble strategies through enhanced visualization and assessment of tumor vascularity compared to traditional contrast-enhanced US [236, 237, 278]. Bachawal et al. [279] tested the sensitivity and specificity of BR55 to differentiate benign from malignant entities in the breast in a transgenic mouse model. The authors found a sensitivity of 84 % and a specificity of 89 %. This new approach seems to be a promising tool for early breast cancer detection. Thus, molecular US can serve as additional screening strategy, especially in women with dense breasts [280].

Therapy Monitoring Differences in the morphology and location of blood vessels inside and in the periphery of benign and malignant tumors have also been observed, but with some variation of results partially due to differences in the contrast agents and dosages used [268, 269, 276, 281]. CEUS has shown promising results in several xenograft model studies [282, 283]; however, to date, there are no published data about CEUS for the evaluation of tumor response to neoadjuvant chemotherapy. Recently, multitargeted microbubbles have also shown the potential to evaluate the early tumor response to antiangiogenic therapy [271], enabling a noninvasive approach to determine the early tumor response to antiangiogenic therapy through molecular US imaging.

17.3.3.2 Ultrahigh-Field MRI of Breast Tumors

Recently, ultrahigh-field MR scanners, operating at a field strength of 7 T, have become clinically available. Compared to MRI at 3 T, ultrahigh-field MRI at 7 T provides a significantly increased intrinsic SNR [101] that can be translated either into even higher temporal and spatial resolution [100, 284] or into functional and metabolic imaging [100, 285]. Initial studies investigating unilateral DCE-MRI of the breast at 7 T have demonstrated the feasibility and encouraged the implementation of further advanced bilateral coil concepts to fully explore the diagnostic potential of DCE-MRI at 7 T [286–288]. In the first clinical study, Pinker et al. investigated the application of bilateral high-resolution DCE-MRI at 7 T in patients with breast lesions [95, 289, 290]. In that study, high-resolution DCE-MRI of the breast at 7 T had a sensitivity of 100 % and a specificity of 90 %, resulting in a diagnostic accuracy of 96.6 % with an AUC of 0.95 (Figs. 17.19 and 17.20). Overall image quality was excellent in the majority of cases and examinations were not hampered by artifacts. The authors concluded that bilateral high-resolution DCE-MRI of the breast at 7 T is clinically applicable and enables a breast cancer diagnosis with a high diagnostic accuracy and excellent inter-rater agreement and image quality. As the increase in intrinsic SNR can also be exploited for functional imaging, several studies have investigated the application of DWI in breast cancer diagnosis at 7 T [101, 291]. DWI at 7 T yields high-quality ADC maps and high-spatial-resolution T2-weighted MR images that can be used to assess tumor and breast morphology, and ADC quantification enables an excellent differentiation of benign and malignant breast lesions. Recently, the concept of multiparametric MRI has been extended to 7 T. Pinker et al. evaluated whether multiparametric MRI of the breast, combining high-resolution DCE-MRI and DWI at 7 T, is feasible and could improve diagnostic accuracy [102]. Multiparametric MRI, combining high-resolution DCE-MRI and DWI at 7 T, yielded a sensitivity of 100 %, a specificity of 88.2 %, and an AUC of 0.941, which was significantly greater than DCE-MRI ($p=0.003$) with a sensitivity of 100 %, a specificity of 53.2 %, and an AUC of 0.765 and DWI, with a sensitivity of 93.1 %, a specificity of 88.2 %, and an AUC 0.907 (Fig. 17.21). In that study, multiparametric MRI of the breast at 7 T accurately detected all cancers, reduced false positives from eight with DCE-MRI to two (Fig. 17.22), and, thus, could have obviated unnecessary breast biopsies ($p=0.031$). Therefore, multiparametric MRI of the breast with DCE-MRI and DWI at 7 T is feasible in clinical practice and seems to enable breast cancer diagnosis with a high diagnostic accuracy and excellent inter-rater agreement. In a quite recent study, Schmitz et al. investigated multiparametric MRI with three parameters, i.e., DCE-MRI, DWI, and 31-phosphorus spectroscopy (^{31}P MRSI) at 7 T for the characterization of breast

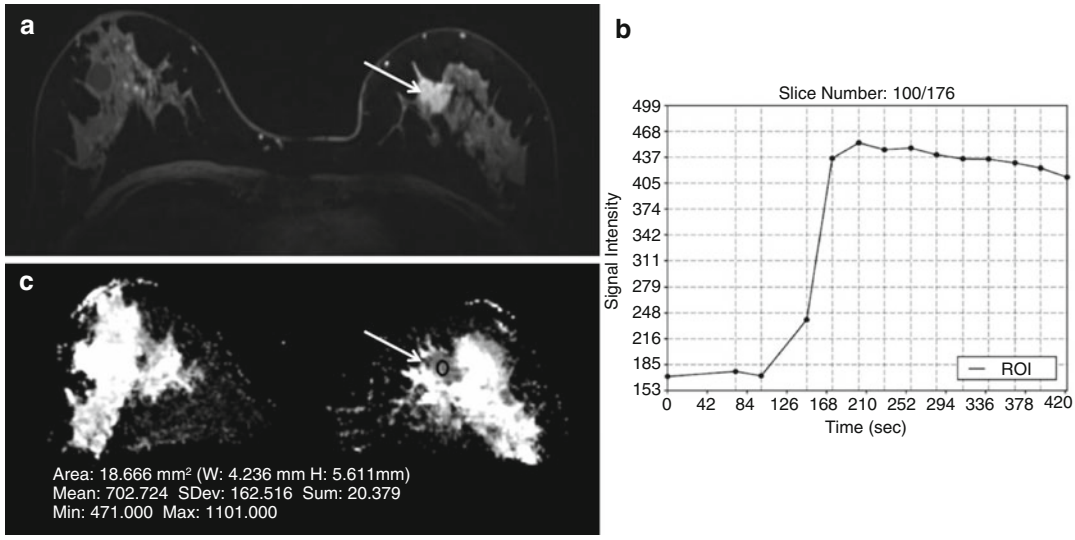


Fig. 17.19 Multiparametric MRI at 7 T with DCE- MRI and DWI. Invasive ductal carcinoma G2 in a 68-year-old woman at 11 o'clock in the left breast. (**a**) The irregular-shaped mass with spiculated margins (*arrow*) demonstrates (**b**) heterogeneous initial strong enhancement followed by a washout. (**c**) On diffusion-weighted imaging (DWI), the ADC values (0.702×10^{-3} mm²/s) were below threshold for malignancy (*arrow*) and were accurately classified as BI-RADS 5 (highly suggestive of malignancy)

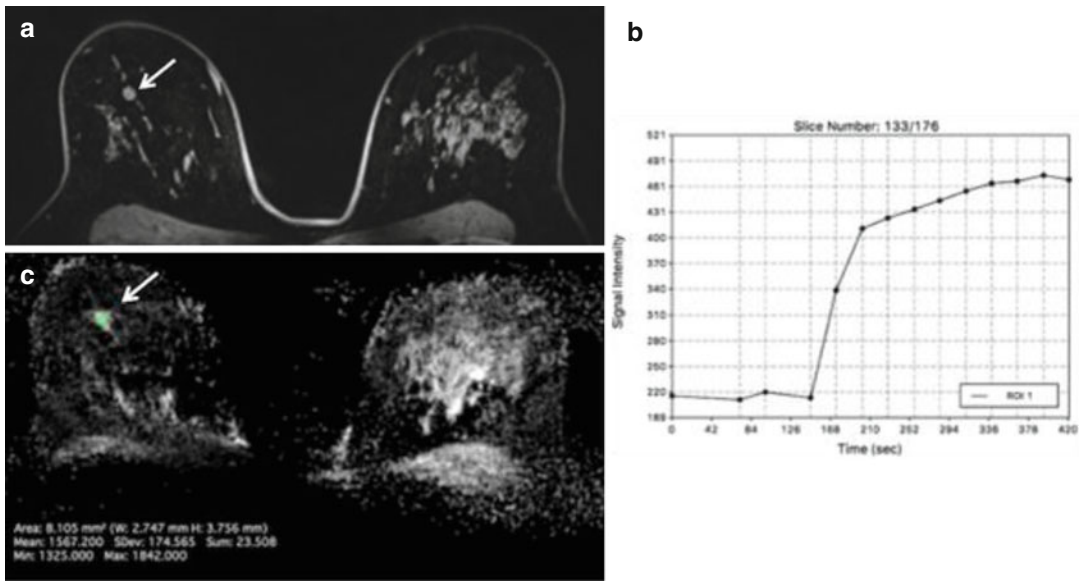


Fig. 17.20 Multiparametric MRI at 7 T with DCE- MRI and DWI. Fibroadenoma in a 50-year-old woman at 2 o'clock in the right breast. (**a**) The round and circumscribed mass (*arrow*) demonstrated (**b**) an initial moderate/persis- tent (II) homogenous contrast enhancement. (**c**) On diffusion-weighted imaging (DWI), the ADC values (1.567×10^{-3} mm²/s) were well above threshold for malignancy (*arrow*) allowing an accurate classification as a benign finding with the BI-RADS-adapted reading

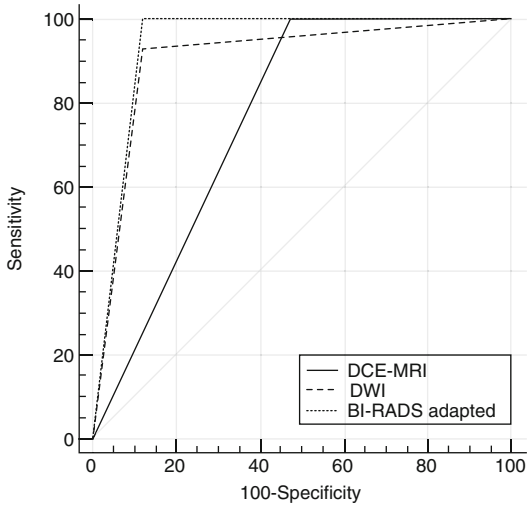


Fig. 17.21 Receiver operator characteristic curves illustrate the higher diagnostic value of multiparametric MR imaging compared with DCE imaging and DWI of the breast (Reproduced from Pinker et al. [102])

cancer [161]. The authors concluded that multiparametric 7 T breast MRI with three parameters is feasible in the clinical setting and showed an association between ADC and tumor grade and between ^{31}P MRSI and mitotic count.

Emerging Parameters for Ultrahigh-Field MRI

Sodium MR Imaging (^{23}NA MRI)

Sodium (^{23}Na) MR imaging has emerged as novel functional MRI parameter for the tumor characterization and therapy monitoring of breast cancer. ^{23}Na MRI provides information about the physiological and biochemical state of tissue, and the sodium concentration is a sensitive indicator of cellular and metabolic integrity, as well as ion homeostasis [292, 293]. In healthy cells, a low intracellular sodium concentration

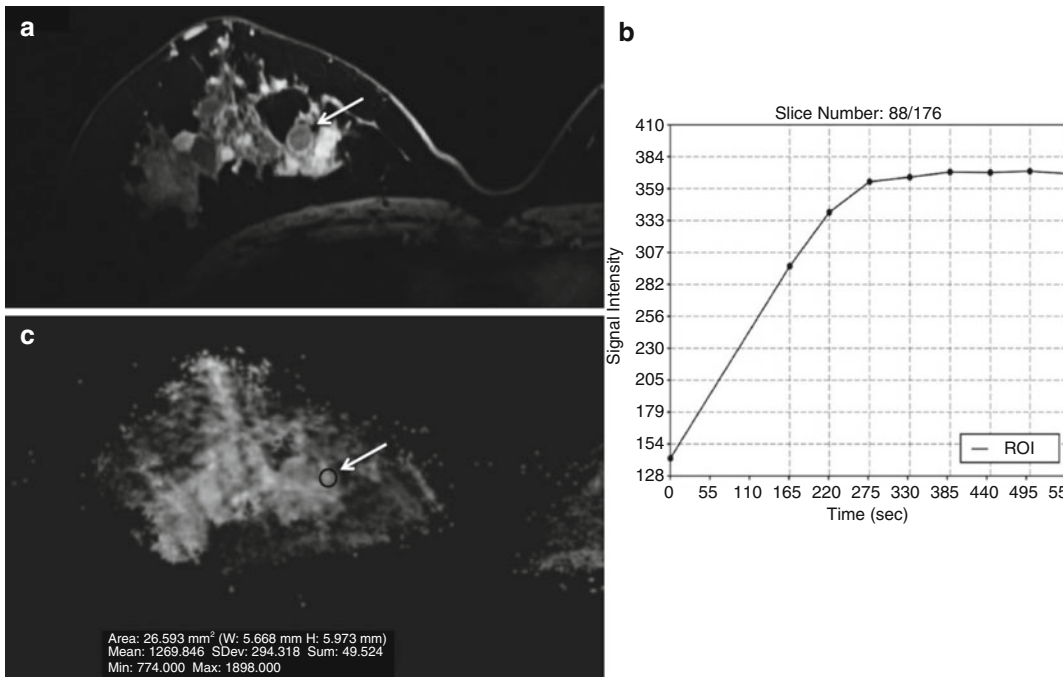


Fig. 17.22 Multiparametric MRI at 7 T with DCE-MRI and DWI. Fibroadenomatous hyperplasia in a 39-year-old woman at 3 o'clock in the right breast adjacent to a cyst (*arrow*). (**a**) The irregularly shaped and partly irregularly marginated mass lesion demonstrates (**b**) homogenous initial strong enhancement followed by a plateau is classified by DCE-MRI of the breast as BI-RADS 4 (probably malignant). (**c**) However, the ADC values are not below the cutoff for malignancy (1.269×10^{-3} mm^2/s) allowing an accurate classification as a benign finding with the BI-RADS-adapted reading

is maintained by the Na^+/K^+ ATPase pump actively pumping sodium out of the cell against a concentration gradient formed by the much higher extracellular sodium concentration. ^{23}Na MRI detects increased sodium levels that occur when there is a failure of the Na^+/K^+-ATPase pump secondary to the breakdown of cell membranes, as seen in malignancy. Ouwerkerk et al. investigated the potential of ^{23}Na MRI for the differentiation of benign and malignant breast lesions at 1.5 Tesla [294]. Those authors demonstrated that an increased total sodium concentration in breast tumors appears to be a sensitive cellular-level indicator of malignancy and has the potential to increase the specificity of breast MRI. At field strengths of 1.5 and 3 T, ^{23}Na MRI is limited. Recently, ^{23}Na MRI has been established at 7 T with promising results. Zaric et al. investigated quantitative ^{23}Na MRI at 7 T and compared it to DWI at 7 T. That study demontrated that quantitative ^{23}Na MRI at 7 T could be accomplished with good resolution and image quality, within clinically acceptable measurement times in patients with breast tumors. ^{23}Na MRI allowed good discrimination of benign and malignant breast lesions ($p = 0.002$) (Figs. 17.23 and 17.24) that was similar to that based on DW-MRI ($p = 0.002$). ^{23}Na MRI can reliably provide additional information about pathophysiologic changes in tumor tissue, with the potential to improve the detection, characterization, and treatment monitoring of breast lesions [295].

Phosphorus Spectroscopy (^{31}P MRSI)

Phosphorus spectroscopy (^{31}P MRSI) measures tissue bioenergetics and membrane phospholipid metabolism. The signals of precursors and catabolites of phospholipids can serve as an imaging biomarker for tumor characterization, prognosis, and treatment moitoring [199, 296]. Several in vitro and in vivo ^{31}P MRSI studies have proven that elevated levels of phosphocholine (PC)/phosphoethanolamine (PE) can be detected in several cancers inclduing breast cancers. Barzilai et al. demonstrated a significant decrease in the PE/PC ratio in malignant, breast tumors compared to benign lesions [297], and changes in PE/PC ratios are observed in response to therapy. However, at field strengths of 1.5 and 3 T, the application of ^{31}P MRSI is limited to relatively large and superficial tumors [296]. Recently, the feasibility of ^{31}P MRSI at 7 T has been investigated, with promising results. Quantitative ^{31}P MRSI of breast tumors at 7 T is feasible, with excellent quality of ^{31}P-MR spectra, and can be used for clinical research. Quantitative ^{31}P MRSI enables a specific analysis of the biochemical status of breast tissue and, thus, can be expected to serve as a valuable imaging biomarker for breast cancer diagnosis and treatment monitoring [285, 298].

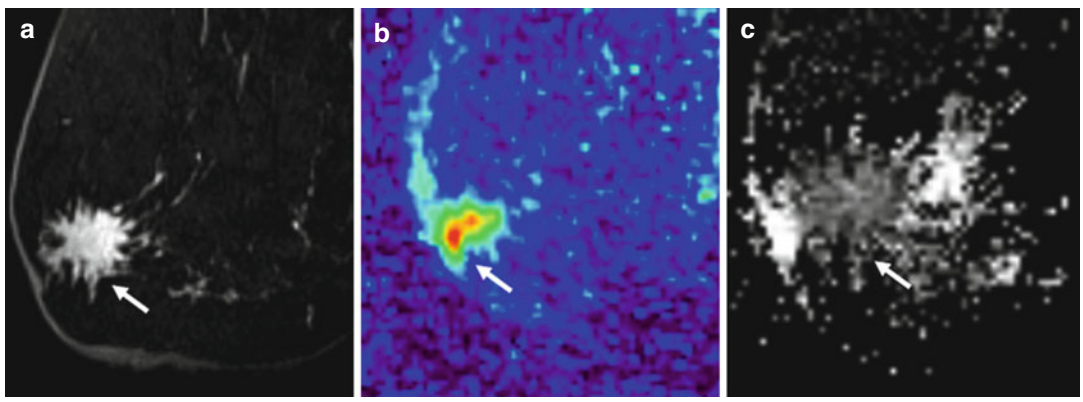


Fig. 17.23 Invasive ductal carcinoma G3 in the left breast of a 49-year-old female patient. (**a**) On DCE MRI there is a strongly enhancing irregular tumor (*arrow*) with spiculated margins. (**b**) Sodium (^{23}Na) MRI reveals an increase in sodium content as result of altered metabolism in the tumor tissue. (**c**) Diffusion-weighted image (DWI) shows low apparent diffusion coefficient (ADC) values indicating high cellularity and tumor malignancy.

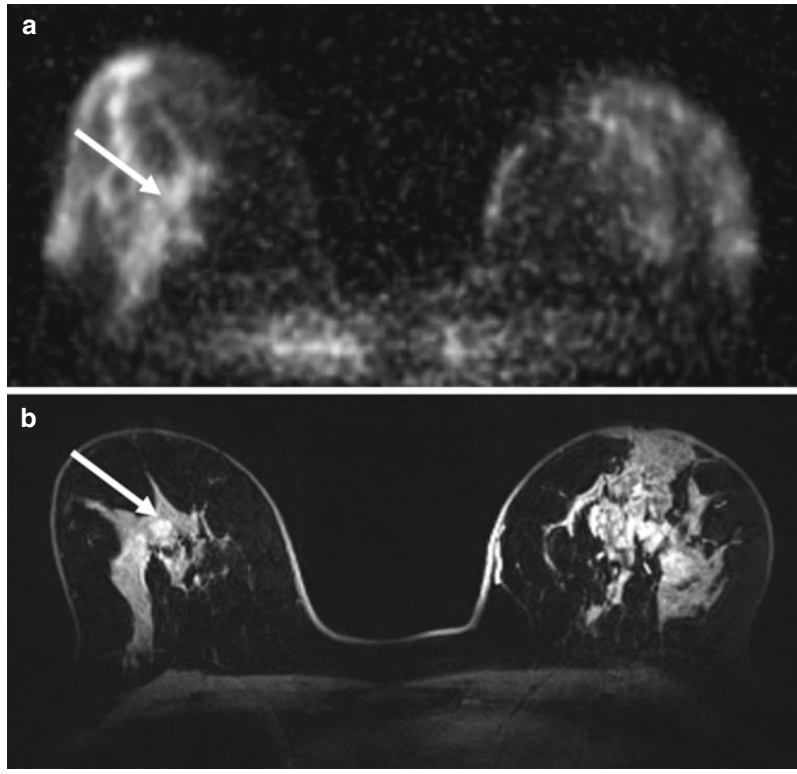


Fig. 17.24 Fibroadenoma in the right breast of a 57-year-old patient. (**a**) In Sodium (^{23}Na) MR the tumor (*arrow*) shows slightly increase in signal compared to glandular tissue; 1.5 in- plane resolution image enables good delineation between morphologic structures (glandular and adipose tissue). (**b**) On DCE the tumor (*arrow*) shows a strong enhancement but regular margins.

Chemical Exchange Saturation Transfer (CEST) Imaging

Chemical exchange saturation transfer (CEST) is an MRI parameter that enables visualization of chemical exchange processes between protons bound to solutes and surrounding bulk water molecules [299, 300]. It has been demonstrated that the endogenous CEST can discriminate tumor from healthy breast tissue based on the information about protons associated with mobile proteins through the amide proton transfer (APT) effect and also has been implicated as a prognosticator of response to therapy [299]. Recently, Schmitt et al. investigated ATP CEST-MRI at 3 T for the detection of breast cancer, with promising results.

In that study, the detection of lesions was equally possible with DCE-MRI and ATP CEST-MRI. The results of this initial feasibility study indicate a significant potential for ATP CEST-MRI to discriminate cancer from fibroglandular tissue in the human breast through the CEST contrast generated by endogenous solute molecules.

Recently, animal studies have investigated CEST contrasts other than ATP, exploiting the entire CEST spectrum. Desmond et al. found that imaging of the amide, amine, and aliphatic signal (aaaCEST) allows a noninvasive differentiation of areas of apoptosis and/or necrosis from actively progressing tumor [301]. In addition, similar to [^{18}F]fluorodeoxyglucose positron emission tomography ([^{18}F]FDG), dynamic CEST after the administration of glucose (glucoCEST) has been shown to enable the noninvasive evaluation of the kinetics of glycolysis, which are typically increased in malignant lesions [19]. Initial results indicate that glucoCEST potentially might serve as a substitute for PET/CT or PET/MRI in the clinic for the detection of tumors and metastases, distinguishing between

malignant and benign tumors and monitoring tumor response to therapy, without the need for radiolabeled isotopes [302–305].

Nevertheless, further studies will be necessary to explore the true potential of CEST imaging in breast cancer.

Hyperpolarized Magnetic Resonance Spectroscopic Imaging (HP MRSI)

Hyperpolarized magnetic resonance spectroscopic imaging (HP MRSI) is one of the most recent advances in molecular imaging. HP MRSI allows a rapid, radiation-free, noninvasive investigation of tumor metabolism by exploiting exogenous contrast agents that have been "hyperpolarized." Conventional MR imaging depends on nuclear spins that have been polarized on the order of a few parts per million, whereas, in HP MRSI, nuclear spins reach near-unity polarization, resulting in an extensive increase in signal intensity [306, 307]. With HP MRSI, nuclear spins are polarized in an amorphous solid state at ~1.2 K through coupling of the nuclear spins with unpaired electrons that are added to the sample via an organic free radical. In the solid state, the high electron spin polarization is, in part, transferred to the nuclear spins by microwave irradiation, and then the sample is rapidly dissolved for injection into the system of interest. Recently, ^{13}C-labeled substrates have been polarized to obtain dramatic enhancements of the ^{13}C nuclear MR signals (>50,000 fold at 3 T) of the substrate, as well as subsequent metabolic products [308]. The enhancement that is achieved is lost in time as a function of the spin–lattice relaxation time of the nucleus (T1). The HP ^{13}C probes can be injected into living systems and their metabolism observed in real time by chemical shift. (^{13}C) pyruvate has been the most widely used probe for HP MR studies since it polarizes well, has a long T_1, and is rapidly taken up by the cell and metabolized at the juncture of glycolysis, TCA, amino acid biosynthesis, and other critical pathways. Several animal studies have confirmed real-time measurement of the relative transformation of pyruvate into lactate and alanine using HP MRSI, which allows differentiation of tumor from normal tissue [309–311]. In addition to the differentiation between malignant and nonmalignant cells [310, 312, 313], HP MRSI using ^{13}C pyruvate has been used for the assessment of cancer progression [314, 315]. Recently, other novel probes for redox (^{13}C dehydroascorbate), necrosis (^{13}C fumarate), and glutamine metabolism (^{13}C glutamine) have been developed to interrogate other metabolic pathways and have shown promising results [316]. To our knowledge, to date, there has been no specific clinical application of HP MRSI to breast cancer; however, several preclinical and initial studies in cancer, including breast cancer [317], suggest that this technique may be applicable for the detection of breast cancer and for the assessment of treatment response in the future.

17.3.3.3 Nuclear Imaging of the Breast

Gamma Camera Imaging: ^{99}mTc-Sestamibi Scinti-mammography (SM) and Breast-Specific Gamma Imaging (BSGI)

^{99}mTc-sestamibi scintigraphic mammography (^{99}mTc-MIBI-SM) was first introduced in the 1990s and can be used as an alternative diagnostic method in patients at high risk for breast cancer. In ^{99}mTc-MIBI-SM, the radiotracer ^{99}mTc-MIBI is injected intravenously and accumulates in tissues with increased perfusion, permeability, and metabolic activity, such as breast cancer tissues. Prior studies have investigated the application of ^{99}mTc-MIBI for the assessment of breast tumors using both planar and single-photon emission computed tomographic (SPECT) radionuclide imaging, with a general-purpose gamma camera, and reported sensitivities ranging from 84 to 93% [318, 319]. However, ^{99}mTc-MIBI-SM has a relatively low spatial resolution, which impedes the depiction of small cancers and is limited in the detection of low-grade breast tumors [320–322]. To overcome the limitations of ^{99}mTc-MIBI-SM, breast-specific gamma imaging (BSGI) has been developed [323–328] (Fig. 17.25). In a recent meta-analysis, Sun et al. evaluated the role of BSGI as an adjunct modality to mammography [329]. The

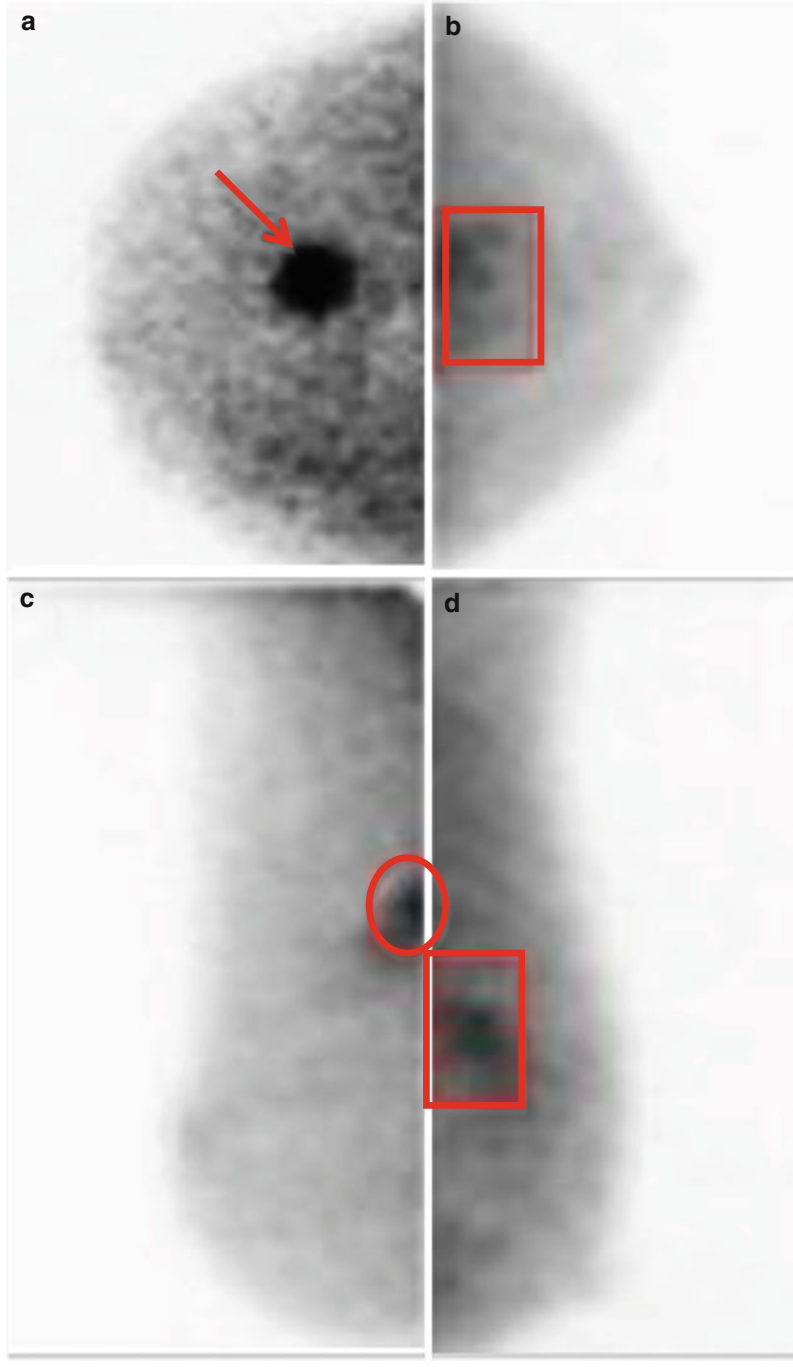


Fig. 17.25 Mammography and breast-specific gamma imaging (BSGI) in a 39-year-old woman who presented with a palpable abnormality of the left breast. BSGI (**a**) right craniocaudal and (**c**) left mediolateral oblique views demonstrated one focus of abnormal uptake in the left breast corresponding to the spiculated mass (*red arrow*). BSGI (**b**) right craniocaudal and (**d**) right mediolateral views revealed an occult focus of focal increased radiotracer uptake (*red rectangle*). Ultrasound-guided core biopsy yielded infiltrating carcinoma in the left breast at the 11:30 axis and in the right breast at the 1:00 axis (Reproduced from Brem et al. [324])

pooled sensitivity and specificity of BGSI was 95 % and 80 %, respectively. In patients with normal mammography findings, 4 % were diagnosed with breast cancer by BSGI, and in those with a suspicious imaging finding on mammography or biopsy-proven breast cancer, 6 % were diagnosed with multifocal disease by BSGI. The authors concluded that BSGI is an excellent adjunct modality to mammography for detecting breast cancer.

Positron Emission Tomography (PET)/ Computed Tomography (CT)

Positron emission tomography (PET) is a noninvasive diagnostic, nuclear medicine imaging method that enables the assessment of physiological processes using radiotracers. Among the multiple radiotracers investigated, [^{18}F]fluorodeoxyglucose ([^{18}F]FDG) is the most commonly used [330]. [^{18}F]FDG PET allows an assessment of tissue glucose consumption, which is typically increased in neoplastic processes such as breast cancer [331–333]. Radiotracer uptake can be quantified using a so-called standardized uptake value (SUV), which is roughly proportional to the rate of phosphorylated FDG. However, [^{18}F]FDG uptake is variable based on the organ of origin, tumor type, and grade, and a critical mass of tumor cells is necessary for visualization [332, 334, 335]. In addition, [^{18}F]FDG is not tumor specific, and some benign conditions, such as inflammatory processes, can also be [^{18}F]FDG avid. As [^{18}F]FDG PET alone provides limited anatomical information and low spatial resolution that frequently results in difficulties in lesion localization and in assessment of potential tumor infiltration into adjacent organs, hybrid imaging systems, such as PET/CT, were developed and established for the clinical routine (Fig. 17.26).

Although imaging studies performed with whole-body PET imaging scanners have established the feasibility of using [^{18}F]FDG PET/CT to identify and characterize breast cancer, with sensitivities and specificities of 88 % and 80 %, respectively, these studies also highlight the limitations that are inherent in currently available whole-body [^{18}F]FDG PET/CT imaging techniques, such as the limited ability to depict small lesions and also particular subtypes of breast cancer [335, 336]. Magometschnigg et al. [336] compared the diagnostic accuracy of [^{18}F]FDG PET/CT with DCE-MRI at 3 T in suspicious breast lesions and evaluated the influence of tumor size on diagnostic accuracy, as well as the use of SUV_{MAX} thresholds to differentiate malignant and benign breast lesions. In that study, the patients were scanned in the prone position with a state-of-the-art combined PET/CT, which likely explains the higher sensitivity and diagnostic accuracy that was achieved. Both [^{18}F]FDG PET/CT and DCE-MRI demonstrate an equal diagnostic accuracy for breast cancer diagnosis of 93 %. Neither sensitivity ($p = 0.125$), specificity ($p = 0.344$), nor diagnostic accuracy ($p = 1$) were significantly different. In lesions <10 mm, diagnostic accuracy deteriorated to 91 % for both [^{18}F]FDG PET/CT and DCE-MRI. In lesions <10 mm, CE-MRI at 3 T is more sensitive, but less specific, than [^{18}F]FDG PET/CT. Quantitative assessment using an SUV_{MAX} threshold for the differentiation of benign and malignant lesions is not helpful in breast cancer diagnosis (Figs. 17.27 and 17.28). The authors concluded that [^{18}F]FDG PET/CT can be considered an alternative imaging modality in patients who are not candidates for DCE-MRI.

Tumor Characterization and Prognosis

Several studies have investigated [^{18}F]FDG PET and [^{18}F]FDG PET/CT in breast imaging [331–333, 337–339]. FDG uptake can manifest a wide variation based on the different histologic subtypes and size of breast cancers [205, 332, 340]. In general, a weak FDG uptake has been reported in ductal carcinoma in situ and in invasive lobular carcinoma, while infiltrating ductal carcinoma has the highest FDG uptake among breast neoplasms [331, 333]. Other authors studied the correlation between grades and molecular characteristics of breast cancer and SUV uptake, demonstrating a significantly higher FDG uptake in tumors with unfavorable prognostic characteristics [9, 330]. Groheaux et al. [340] prospectively studied the relationship between tumor characteristics and the SUV in 132 women with

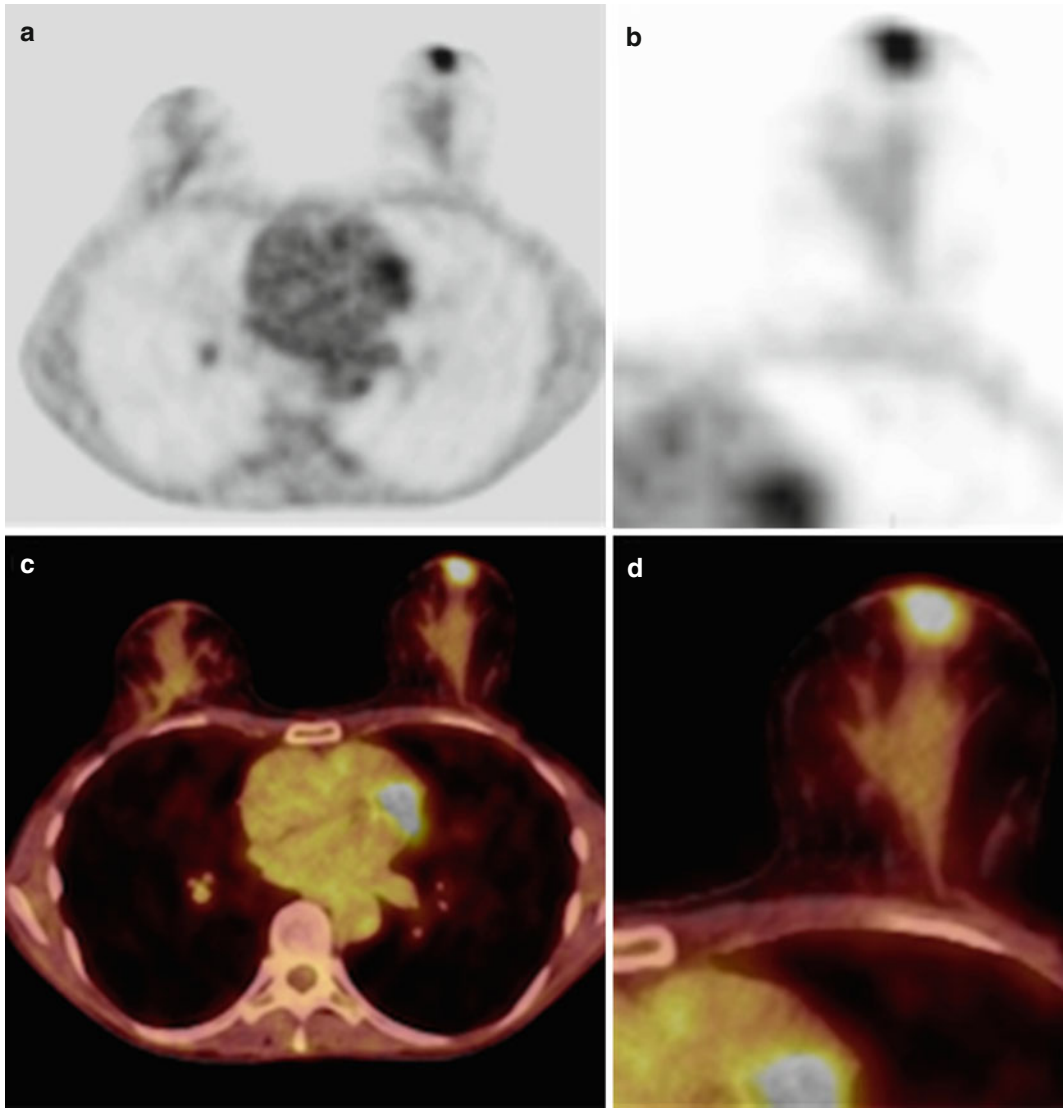


Fig. 17.26 ^{18}FDG PET/CT of the breast. Invasive ductal carcinoma G2 of the left breast in a 46-year-old woman: (**a**, **b**) ^{18}FDG PET images, (**c**, **d**) ^{18}FDG PET/CT fused images. The lesion is highly ^{18}FDG avid, with a SUV_{MAX} of 4.4. *^{18}FDG* [F-18]-fluorodeoxyglucose, *SUV* standardized uptake value

breast tumors larger than 20 mm. They found that FDG uptake in primary tumors correlated with several factors known to confer a poorer prognosis, such as high tumor grade, estrogen receptor-negative tumors, and triple-negative cancers. Considering histologic tumor type, lower tumoral uptake was found in invasive lobular carcinoma (median SUV_{max}, 3.4) compared with invasive ductal carcinoma (median SUV_{max}, 6.6). According to the literature, however, the diagnostic yield for [^{18}F]FDG PET/CT in patients with early-stage breast cancer is low, and there is currently no clinical role for [^{18}F]FDG PET/CT in the detection of primary breast cancer or in the evaluation of axillary lymph nodes in stage I and early-stage II disease, but these are areas of active research [205, 208].

Therapy Monitoring [^{18}F]FDG PET/CT is currently considered a valuable adjunct to other

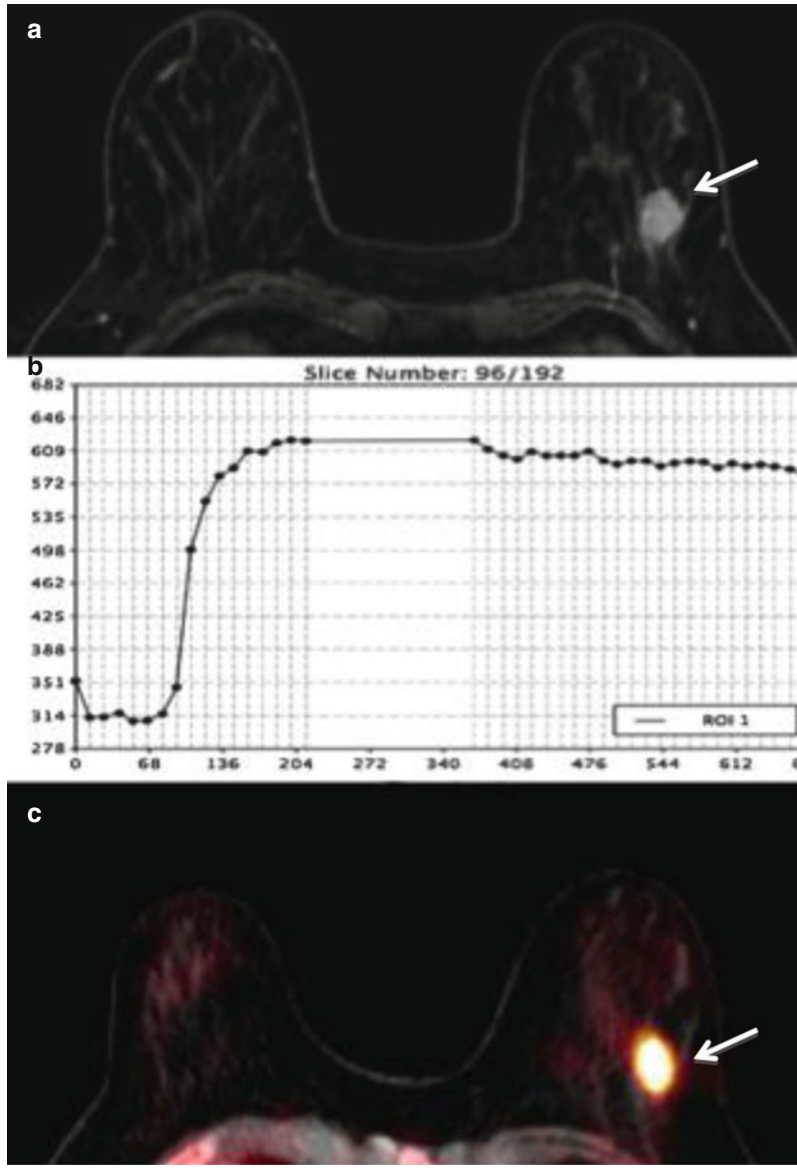


Fig. 17.27 Invasive ductal carcinoma G3 laterally in the left breast in a 40-year-old woman. (**a**) The round, irregularly marginated mass lesion (*arrow*) shows a heterogeneous (**b**) initial strong enhancement followed by a washout phase and was classified by CE-MRI as BI-RADS 5 (suspicious). (**c**) On the ^{18}F-FDG PET/CT image, the lesion is strongly ^{18}F-FDG avid (*arrow*), with an SUV_{MAX} of 16.83. The lesion was true positive by both modalities (Reproduced from Magometschnigg et al. [336])

imaging techniques for the staging of locally advanced metastatic or recurrent breast cancer and in evaluating the response of locally advanced and metastatic breast cancer to treatment. In several studies, the role of metabolic evaluation with [^{18}F]FDG PET and [^{18}F] FDG PET/CT was examined, demonstrating a correlation between early changes in the SUV_{max} value (after one or two courses of chemotherapy) and the final pathologic response at completion of chemotherapy [338, 341, 342]. An optimal threshold value for a decrease in SUV (ΔSUV) has been proposed for discriminating metabolic responders (diminution of SUV superior to the

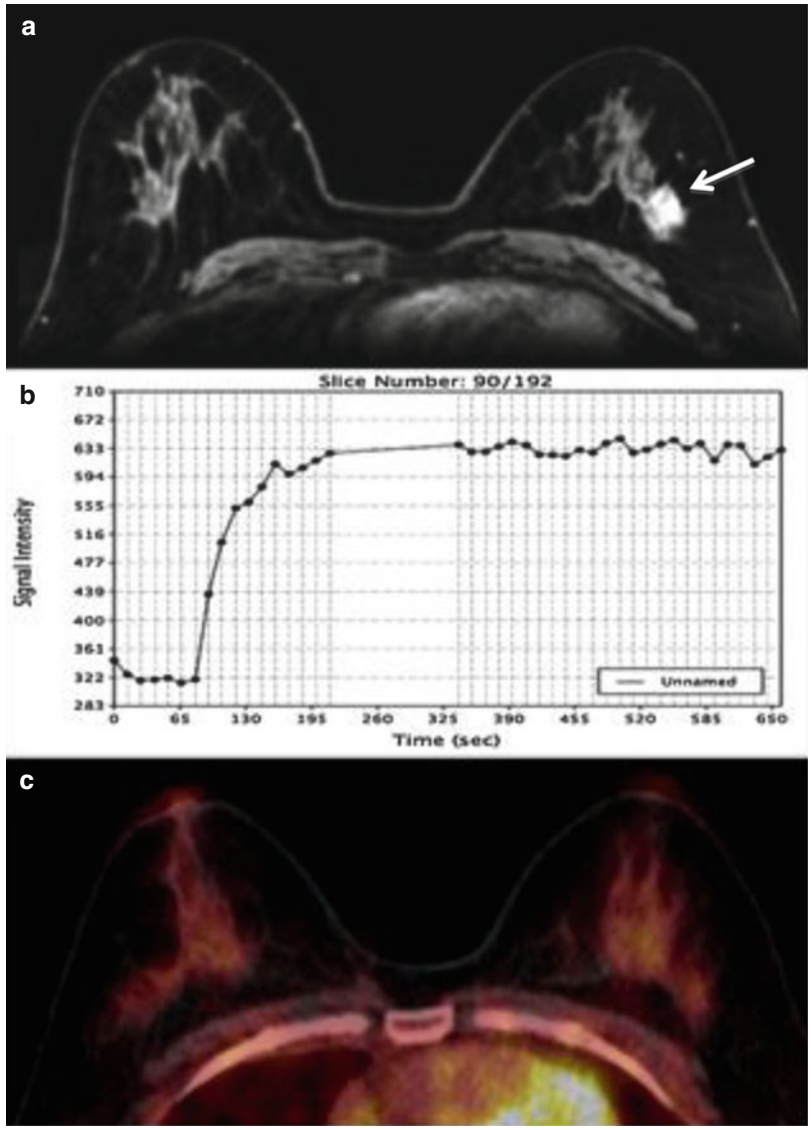


Fig. 17.28 Invasive lobular carcinoma G2 laterally in the left breast in a 53-year-old woman. (**a**) The irregularly shaped and not circumscribe mass lesion (*arrow*) shows a heterogeneous (**b**) medium initial enhancement followed by a plateau and was classified by CE-MRI as BI-RADS 4 (suspicious). (**c**) On the 18F-FDG PET/CT image, the mass is not significantly ^{18}F-FDG avid, with an SUVMAX of 1.9. The lesion was a true-positive finding by DCE-MRI at 3 T and false negative by ^{18}F-FDG PET/CT (Reproduced from Magometschnigg et al. [336])

threshold value) from nonresponders. However, the cutoff varies across studies and more studies are necessary to better define the criteria for evaluation. Available data also suggest that [^{18}F] FDG PET/CT is valuable in the evaluation of advanced axillary disease and nodal spread in locally advanced breast cancer [331–334, 337–339]. In a patient suspected of having a breast cancer recurrence, [^{18}F]FDG PET/CT seems to perform better than conventional imaging (whole-body CT and bone scanning for distant recurrences, US and mammography for local recurrences) and better than PET alone, with a sensitivity ranging between 85 and 97 %; the specificity, between 52 and 100 %; and the accuracy, between 60 and 98 % [340, 343, 344].

PET Mammography (PEM)

To overcome the limitations of whole-body [^{18}F]FDG PET/CT in breast imaging, dedicated breast PET systems have been developed. Positron emission mammography (PEM) is a high-resolution, breast-specific device enabling co-registration of mammographic and emission [^{18}F]FDG images of the breast using two flat detectors integrated in the system on either side of the breast. The more recently developed breast-specific MAMography with MolecularImaging (MAMMI)-PET utilizes a small ring of detectors, which yields an improved contrast and signal-to-noise ratio (SNR). Dedicated breast PET systems have been shown to have both a high sensitivity and specificity in detecting breast malignancy (<1 cm). Kalles et al. reviewed the role of [^{18}F]FDG PEM in breast cancer imaging and concluded that [^{18}F]FDG PEM can successfully complement conventional imaging in breast cancer detection by providing information about tumor biology. In a recent study, Kalinyak et al. [345] compared [^{18}F]FDG PEM with whole-body [^{18}F]FDG PEM or [^{18}F]FDG PET/CT in 178 breast cancer patients and demonstrated that ^{18}FDG PEM was more sensitive than whole-body ^{18}FDG PET or PET/CT by showing index and additional ipsilateral breast tumors smaller than 10 mm.

The current data suggest that PEM might not be far from being included in the first-line modalities for breast cancer screening [330].

Molecular Imaging with Multiparametric PET/MRI

Multiparametric MRI and PET of the breast can visualize different processes involved in cancer development and progression and have been proven to be valuable tools in breast imaging by gathering and combining morphologic, functional, metabolic, and molecular information. To overcome the limitations of morphologic and functional imaging techniques, hybrid imaging systems have been developed and introduced into clinical routine. Initial studies investigating fused [^{18}F]FDG PET and DCE-MRI for breast cancer diagnosis demonstrated that fused [^{18}F]FDG PET/MRI provides accurate morphologic and functional data and has the potential to emerge as an all-encompassing alternative to conventional multitechnique tumor staging [346, 347]. However, in these studies, the functional information that [^{18}F]FDG PET provided was only combined with the morphologic and partly functional information of DCE-MRI and the true potential of multiparametric [^{18}F]FDG PET/MRI was not fully explored. In a recent feasibility study, Pinker et al. investigated multiparametric PET-MRI using DCE-MRI, DWI, ^{1}H-MRSI, and [^{18}F]FDG for the assessment of breast tumors at 3 T [342]. The authors demonstrated that MP [^{18}F]FDG PET/MRI enables an improved differentiation of benign and malignant breast tumors when several MRI and PET parameters are combined (Figs. 17.29 and 17.30). In addition, they concluded that MP [^{18}F]FDG PET/MRI may lead to a reduction in unnecessary breast biopsies.

Specific Radiotracers

To date, nuclear imaging and combined nuclear and radiological imaging in breast cancer is mainly performed using the radiotracer [^{18}F]FDG and 99mTc-MIBI for PET and GAMMA imaging. The advances in molecular cancer biology have led to an increased understanding of the cancer biomarkers that contribute to cancer progression and, thus, led to the rapid development of more personalized and specifically tumor-targeted treatments. Recently, more specific radiotracers to target processes involved in cancer evolution and progression have been developed and introduced into breast imaging. [^{18}F]Fluoromisonidazole ([^{18}F]FMISO) for the assessment of tumor hypoxia, radioactive-labeled annexin V for the assessment of tumor neoangiogenesis, [^{18}F]fluoro-L-thymidine ([^{18}F]FLT) for the assessment of nucleic acid metabolism, and [^{18}F]fluoroestradiol ([^{18}F]FES) and radiolabeled trastuzumab for the assessment of tumor receptor status are some of those radiotracers.

- Radiotracters for imaging tumor hypoxia

[^{18}F]FMISO has a high affinity for hypoxic cells with active nitroreductase enzymes and accumulates in activated tumor cells compared to

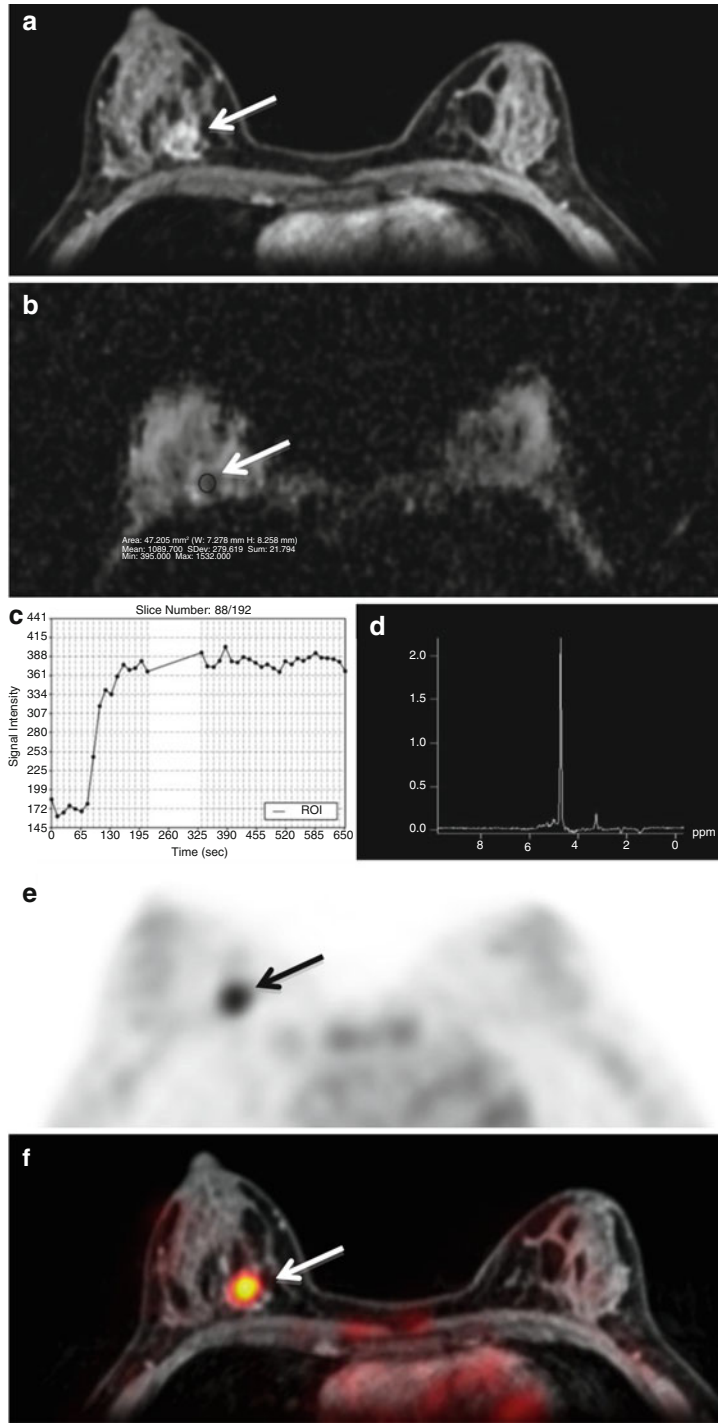


Fig. 17.29 Multiparametric MRI at 3 T DCE-MRI, DWI, 3D ^{1}H-MRSI, and [^{18}F]FDG. Invasive ductal carcinoma G3 in a 46-year-old woman at 2 o'clock in the right breast. (**a**) The irregular shape and irregularly marginated mass (*arrow*) showed an (**c**) initial fast enhancement followed by a plateau and was classified by DCE-MRI of the breast as BI-RADS 5 (suspicious finding). (**b**) The lesion demonstrates decreased ADC values (*arrow*, 1.089×10^{-3} mm^2/s) , and (**d**) there is a distinct choline peak at 3.2 ppm and, therefore, considered malignant with DWI and 3D ^{1}H-MRSI. (**e**) On [^{18}F]FDG PET, the lesion is highly [^{18}F] FDG avid (*arrow*). (**f**) Multiparametric PET/MRI accurately classified the tumor as malignant.

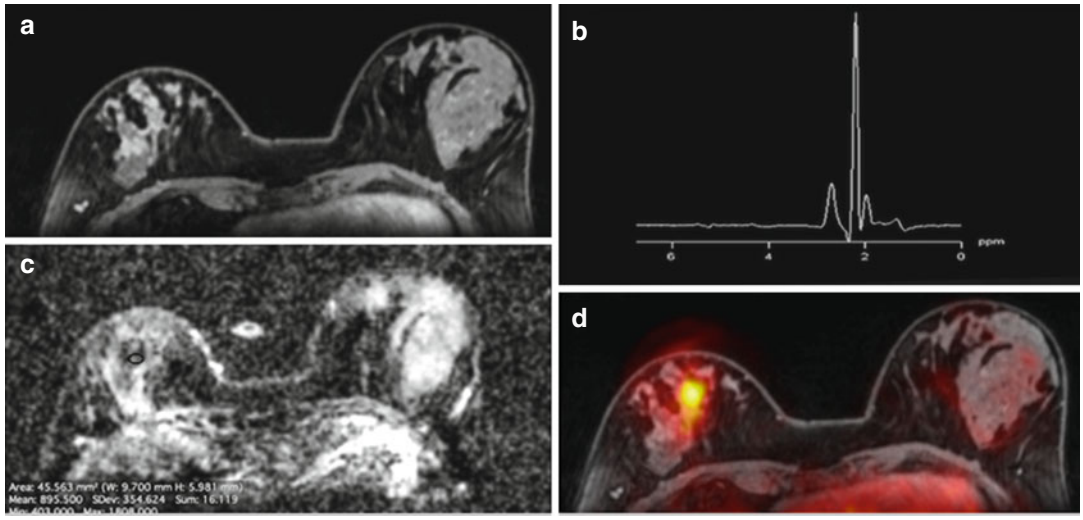


Fig. 17.30 Multiparametric MRI at 3 T DCE-MRI, DWI, 3D ^{1}H-MRSI, and [^{18}F]FDG. Invasive ductal carcinoma G3 in a 44-year-old woman in the right breast retroareolar. (**a**, *arrow*) The irregular heterogeneous non-mass enhancement was classified by DCE-MRI of the breast as BI-RADS 4 (suspicious finding). (**b**) On 3D ^{1}H-MRSI, the lesion is false negative, as there were no elevated Cho levels. (**c**) The lesion demonstrates decreased ADC values (0.895×10^{-3} mm^2/s) and is, therefore, considered malignant. (**d**) On [^{18}F]FDG PET, the lesion is highly [^{18}F]FDG avid

necrotic cells. [^{18}F]FMISO therefore may serve as a potential imaging biomarker for tumor grading and assessment of treatment response. Cheng et al. [348] investigated whether [^{18}F]-MISO PET/CT could predict primary resistance to endocrine therapy in estrogen receptor-positive breast cancer and found a significantly positive correlation between baseline [^{18}F]-MISO uptake and clinical outcomes after ≥3 months of primary endocrine therapy with letrozole. The data suggests that [^{18}F]-MISO PET/CT may be used as an effective method for monitoring the response to endocrine therapy and has the potential for early identification of nonresponders.

- Radiotracers for imaging apoptosis

Apoptosis has a significant role in tumorgenesis, progression, and therapy. Apoptosis induces a cascade of enzymatic processes that eventually lead to cell death. The acitvation of caspases enables the externalizaion of phosphatidylserine (PS), which is usually located on the inside of the cell membrane. The protein annexin V binds to PS with a high affinity and, therefore, is a marker for apoptosis. To date, annexin V has been labeled with multiple radiotracers for PET imaging [349].

- Radiotracers for imaging cell proliferation

The radiotracer [^{18}F]FLT accumulates in proliferating cells. Its accumulation is regulated by the thymidine salvage pathway and by the activity of thymidine kinase 1 and, thus, reflects DNA synthesis. This highly specific radiotracer demonstrated promising results for the detection of treatment response in preclinical breast cancer mouse models and is now investigated in several clinical trials [350].

- Radiotracer imaging amino acid trasporters and protein synthesis

Tumor cells take up and consume more amino acids to sustain their uncontrolled growth compared with normal cells. Amino acid-based radiotracers can enter the tumor via amino acid transporters and Na+-dependent transport systems. 99mTc-labeled methionine has been successfully used to detect breast cancer in a recent clinical trial performed by Sharma et al. [351].

The radiotracer 99mTc-methionine provides a more simple and affordable way to image breast tumors using conventional scinti-mammography. The 99mTc-methionine was synthesized by conjugating methionine with diethylenetriaminepentaacetic acid and 99mTc, and the radiochemical yield was >95 %. The sensitivity, specificity, and positive predictive value of 99mTc-methionine in this clinical trial with 47 patients were 87.8 %, 92.8 %, and 96.6 %, respectively. Another radiotracer, 11C-methionine (11C-MET), has recently been evaluated by Lindholm et al. [352] for its potential to assess early response to therapy in advanced breast cancer. The SUV in all responding metastatic sites decreased by 30–54 % ($P < 0.05$), while that of nonresponding sites did not decrease significantly (11–13 %, P = NS).

- Radiotracers for imaging the receptor status

[^{18}F]FES PET imaging allows the noninvasive visualization and quantification of estrogen receptor expression of both the primary tumor and metastases [353]. In addition, [^{18}F]FES PET/CT provides valuable information about the response to endocrine therapy both in the neoadjuvant and adjuvant setting. In a recent publication by van Kruchten et al., the authors provide a comprehensive overview about the role of [^{18}F]FES PET/CT in breast cancer [354]. The authors concluded that [^{18}F]FES PET/CT has the potential to significantly influence patient management.

Radiolabeled trastuzumab allows the noninvasive visualization and quantification of human epidermal growth factor receptor 2 (HER2) status. In initial studies, Smith-Jones et al. confirmed the noninvasive measurement of HER2 expression and therapy-induced changes using a ^{68}gallium-labeled fragment of trastuzumab in an animal model [341, 355]. Recent clinical PET/CT studies with ^{64}copper (^{64}Cu)-DOTA-trastuzumab demonstrated that ^{64}Cu-DOTA-trastuzumab PET/CT enables the detection of the primary tumor as well as metastases with excellent sensitivity and, therefore, has the potential to further improve HER2-targeted therapies [356, 357].

- Radiolabeled imaging of angiogenesis

Integrins are glycoproteins located on the cell surface that are extremely important in angiogenesis to mediate cell-cell or cell-extracellular matrix interactions [358]. Among the 24 integrins reported, integrin alpha-V beta-3 is the best-studied subtype as the molecular marker for targeting the angiogenic cascade. Two major radiotracers have been recently introduced, 99mTc-NC100692 for SPECT imaging and [^{18}F]-galacto-RGD for PET imaging. NC100692 is a cyclic synthetic ligand containing an RGD binding site with a high affinity for $\alpha_v\beta_3$ and $\alpha_v\beta_5$, which are upregulated during angiogenesis. Bach-Gansmo et al. [359] found in their study that 19 of 22 malignant lesions from 20 breast cancer patients could be detected by 99mTc-NC100692 scintigraphy (86 %).

The radiotracer [^{18}F]-galacto-RGD was first developed in 2001 [360]. Beer et al. performed ^{18}F-galacto-RGD -PET in 16 patients with invasive ductal breast cancer. All the invasive carcinomas could be identified (SUV = 3.6 ± 1.8, tumor-to-muscle ratios = 6.2 ± 2.2); however, only three of eight lymph node metastases were detected. According to several preclinical and clinical studies, [^{18}F]-galacto-RGD-PET may not be suitable for the differentiation of tumor from inflammation, because [^{18}F]-galacto-RGD is also highly expressed in macrophages and other inflammatory lesions [361–364].

It can be expected that these tailored radiotracers will play a major role in the detection, characterization, staging, and therapy monitoring of breast cancer in the future.

References

1. DeSantis CE, Bray F, Ferlay J, et al. International variation in female breast cancer incidence and mortality rates. Cancer Epidemiol Biomarkers Prev. 2015;24(10):1495–506. doi:10.1158/1055-9965.EPI-15-0535.
2. Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value? Clin Chem. 2006;52:345–51. doi:10.1373/clinchem.2005.059832.
3. Phi X-A, Houssami N, Obdeijn I-M, et al. Magnetic resonance imaging improves breast screening sensitivity in

BRCA mutation carriers age ≥50 years: evidence from an individual patient data meta-analysis. J Clin Oncol. 2015;33:349–56. doi:10.1200/JCO.2014.56.6232.
4. Riedl CC, Ponhold L, Gruber R, et al. New information on high risk breast screening. Radiologe. 2010;50:955–6. doi:10.1007/s00117-010-2011-z, 958–963.
5. Riedl CC, Luft N, Bernhart C, et al. Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. J Clin Oncol. 2015;33:1128–35. doi:10.1200/JCO.2014.56.8626.
6. Clark SE, Warwick J, Carpenter R, et al. Molecular subtyping of DCIS: heterogeneity of breast cancer reflected in pre-invasive disease. Br J Cancer. 2011;104:120–7. doi:10.1038/sj.bjc.6606021.
7. Grimm LJ, Johnson KS, Marcom PK, et al. Can breast cancer molecular subtype help to select patients for preoperative MR imaging? Radiology. 2015;274:352–8. doi:10.1148/radiol.14140594.
8. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. Cancer Res. 2008;68:3108–14. doi:10.1158/0008-5472.CAN-07-5644.
9. Yoon H-J, Kang KW, Chun IK, et al. Correlation of breast cancer subtypes, based on estrogen receptor, progesterone receptor, and HER2, with functional imaging parameters from ^{68}Ga-RGD PET/CT and ^{18}F-FDG PET/CT. Eur J Nucl Med Mol Imaging. 2014;41:1534–43. doi:10.1007/s00259-014-2744-4.
10. European Society of Radiology (ESR). White paper on imaging biomarkers. Insights Imaging. 2010;1:42–5. doi:10.1007/s13244-010-0025-8.
11. Van Beers BE, Vilgrain V. Biomarkers in abdominal imaging. Abdom Imaging. 2009;34:663–7. doi:10.1007/s00261-008-9480-9.
12. Padhani AR, Liu G, Koh DM, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia. 2009;11:102–25.
13. Huber KE, Carey LA, Wazer DE. Breast cancer molecular subtypes in patients with locally advanced disease: impact on prognosis, patterns of recurrence, and response to therapy. Semin Radiat Oncol. 2009;19:204–10. doi:10.1016/j.semradonc.2009.05.004.
14. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52. doi:10.1038/35021093.
15. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100:8418–23. doi:10.1073/pnas.0932692100.
16. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98:10869–74. doi:10.1073/pnas.191367098.
17. Tran B, Bedard PL. Luminal-B breast cancer and novel therapeutic targets. Breast Cancer Res. 2011;13:221. doi:10.1186/bcr2904.
18. Bertos NR, Park M. Breast cancer – one term, many entities? J Clin Invest. 2011;121:3789–96. doi:10.1172/JCI57100.
19. Perez EA, Romond EH, Suman VJ, et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. J Clin Oncol. 2014;32:3744–52. doi:10.1200/JCO.2014.55.5730.
20. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med. 2005;353:1673–84. doi:10.1056/NEJMoa052122.
21. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363:1938–48. doi:10.1056/NEJMra1001389.
22. Schneider BP, Winer EP, Foulkes WD, et al. Triple-negative breast cancer: risk factors to potential targets. Clin Cancer Res. 2008;14:8010–8. doi:10.1158/1078-0432.CCR-08-1208.
23. Gupta A, Mutebi M, Bardia A. Gene-expression-based predictors for breast cancer. Ann Surg Oncol. 2015;22:3418–32. doi:10.1245/s10434-015-4703-0.
24. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351:2817–26. doi:10.1056/NEJMoa041588.
25. Martin LJ, Boyd NF. Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence. Breast Cancer Res. 2008;10:201. doi:10.1186/bcr1831.
26. Byng JW, Yaffe MJ, Jong RA, et al. Analysis of mammographic density and breast cancer risk from digitized mammograms. Radiographics. 1998;18:1587–98. doi:10.1148/radiographics.18.6.9821201.
27. McCormack VA, dos Santos SI. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15:1159–69. doi:10.1158/1055-9965.EPI-06-0034.
28. Pettersson A, Graff RE, Ursin G, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst. 2014; 106(5). pii: dju078. doi: 10.1093/jnci/dju078.
29. Byrne C. Mammographic density: a breast cancer risk factor or diagnostic indicator? Acad Radiol. 2002;9:253–5.
30. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol. 2005;6:798–808. doi:10.1016/S1470-2045(05)70390-9.
31. Lee CI, Bassett LW, Lehman CD. Breast density legislation and opportunities for patient-centered outcomes research. Radiology. 2012;264:632–6. doi:10.1148/radiol.12120184.
32. Gierach GL, Ichikawa L, Kerlikowske K, et al. Relationship between mammographic density and breast cancer death in the Breast Cancer Surveillance Consortium. J Natl Cancer Inst. 2012;104:1218–27. doi:10.1093/jnci/djs327.

33. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. Radiology. 2008;246:348–53. doi:10.1148/radiol.2461070309.
34. Kerlikowske K, Ichikawa L, Miglioretti DL, et al. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. J Natl Cancer Inst. 2007;99:386–95. doi:10.1093/jnci/djk066.
35. Vachon CM, Kushi LH, Cerhan JR, et al. Association of diet and mammographic breast density in the Minnesota breast cancer family cohort. Cancer Epidemiol Biomarkers Prev. 2000;9:151–60.
36. Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. Breast Cancer Res. 2007;9:217. doi:10.1186/bcr1829.
37. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003;138:168–75.
38. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med. 2011;155:493–502. doi:10.7326/0003-4819-155-8-201110180-00005.
39. Freer PE. Mammographic breast density: impact on breast cancer risk and implications for screening. Radiographics. 2015;35:302–15. doi:10.1148/rg.352140106.
40. Yaffe M, Boyd N. Mammographic breast density and cancer risk: the radiological view. Gynecol Endocrinol. 2005;21 Suppl 1:6–11. doi:10.1080/09513590400030053.
41. Boyd NF, Lockwood GA, Byng JW, et al. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 1998;7:1133–44.
42. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356:227–36. doi:10.1056/NEJMoa062790.
43. Byng JW, Boyd NF, Fishell E, et al. The quantitative analysis of mammographic densities. Phys Med Biol. 1994;39:1629–38.
44. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS® atlas, breast imaging reporting and data system. Reston: American College of Radiology; 2013.
45. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. Breast. 2005;14:269–75. doi:10.1016/j.breast.2004.12.004.
46. Redondo A, Comas M, Macià F, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol. 2012;85:1465–70. doi:10.1259/bjr/21256379.
47. Timmers JMH, van Doorne-Nagtegaal HJ, Verbeek ALM, et al. A dedicated BI-RADS training programme: effect on the inter-observer variation among screening radiologists. Eur J Radiol. 2012;81:2184–8. doi:10.1016/j.ejrad.2011.07.011.
48. Johns PC, Yaffe MJ. X-ray characterisation of normal and neoplastic breast tissues. Phys Med Biol. 1987;32:675–95.
49. Ciatto S, Bernardi D, Calabrese M, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. Breast. 2012;21:503–6. doi:10.1016/j.breast.2012.01.005.
50. Yaffe MJ. Mammographic density. Measurement of mammographic density. Breast Cancer Res. 2008;10:209. doi:10.1186/bcr2102.
51. van Engeland S, Snoeren PR, Huisman H, et al. Volumetric breast density estimation from full-field digital mammograms. IEEE Trans Med Imaging. 2006;25:273–82. doi:10.1109/TMI.2005.862741.
52. Highnam R, Brady M, English R. Detecting film-screen artifacts in mammography using a model-based approach. IEEE Trans Med Imaging. 1999;18:1016–24. doi:10.1109/42.811313.
53. Highnam R, Jeffreys M, McCormack V, et al. Comparing measurements of breast density. Phys Med Biol. 2007;52:5881–95. doi:10.1088/0031-9155/52/19/010.
54. Morris EA. Diagnostic breast MR imaging: current status and future directions. Radiol Clin North Am. 2007;45:863–80. doi:10.1016/j.rcl.2007.07.002, vii.
55. Wei J, Chan H-P, Helvie MA, et al. Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images. Med Phys. 2004;31:933–42.
56. Nie K, Chang D, Chen J-H, et al. Quantitative analysis of breast parenchymal patterns using 3D fibroglandular tissues segmented based on MRI. Med Phys. 2010;37:217–26.
57. Nie K, Chen J-H, Chan S, et al. Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI. Med Phys. 2008;35:5253–62.
58. King V, Brooks JD, Bernstein JL, et al. Background parenchymal enhancement at breast MR imaging and breast cancer risk. Radiology. 2011;260:50–60. doi:10.1148/radiol.11102156.
59. Baltzer PA, Dietzel M, Vag T, et al. Clinical MR mammography: impact of hormonal status on background enhancement and diagnostic accuracy. RoFo. 2011;183:441–7. doi:10.1055/s-0029-1246072.
60. Price ER, Brooks JD, Watson EJ, et al. The impact of bilateral salpingo-oophorectomy on breast MRI background parenchymal enhancement and fibroglandular tissue. Eur Radiol. 2014;24:162–8. doi:10.1007/s00330-013-2993-9.
61. DeLeo MJ, Domchek SM, Kontos D, et al. Breast MRI fibroglandular volume and parenchymal enhancement in BRCA1 and BRCA2 mutation carriers before and immediately after risk-reducing salpingo-oophorectomy. AJR Am J Roentgenol. 2015;204:669–73. doi:10.2214/AJR.13.12146.
62. Hambly NM, Liberman L, Dershaw DD, et al. Background parenchymal enhancement on baseline

screening breast MRI: impact on biopsy rate and short-interval follow-up. AJR Am J Roentgenol. 2011;196:218–24. doi:10.2214/AJR.10.4550.
63. Ko ES, Lee BH, Choi HY, et al. Background enhancement in breast MR: correlation with breast density in mammography and background echotexture in ultrasound. Eur J Radiol. 2011;80:719–23. doi:10.1016/j.ejrad.2010.07.019.
64. Dontchos BN, Rahbar H, Partridge SC, et al. Are qualitative assessments of background parenchymal enhancement, amount of fibroglandular tissue on MR images, and mammographic density associated with breast cancer risk? Radiology. 2015;276:371–80. doi:10.1148/radiol.2015142304.
65. Uematsu T, Kasami M, Watanabe J. Does the degree of background enhancement in breast MRI affect the detection and staging of breast cancer? Eur Radiol. 2011;21:2261–7. doi:10.1007/s00330-011-2175-6.
66. DeMartini WB, Liu F, Peacock S, et al. Background parenchymal enhancement on breast MRI: impact on diagnostic performance. AJR Am J Roentgenol. 2012;198:W373–80. doi:10.2214/AJR.10.6272.
67. Thompson DJ, Leach MO, Kwan-Lim G, et al. Assessing the usefulness of a novel MRI-based breast density estimation algorithm in a cohort of women at high genetic risk of breast cancer: the UK MARIBS study. Breast Cancer Res. 2009;11:R80. doi:10.1186/bcr2447.
68. Khazen M, Warren RML, Boggis CRM, et al. A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging. Cancer Epidemiol Biomarkers Prev. 2008;17:2268–74. doi:10.1158/1055-9965.EPI-07-2547.
69. Wengert GJ, Helbich TH, Vogl W-D, et al. Introduction of an automated user-independent quantitative volumetric magnetic resonance imaging breast density measurement system using the Dixon sequence: comparison with mammographic breast density assessment. Invest Radiol. 2015;50:73–80. doi:10.1097/RLI.0000000000000102.
70. Ma J. Breath-hold water and fat imaging using a dual-echo two-point Dixon technique with an efficient and robust phase-correction algorithm. Magn Reson Med. 2004;52:415–9. doi:10.1002/mrm.20146.
71. Ma J. Dixon techniques for water and fat imaging. J Magn Reson Imaging. 2008;28:543–58. doi:10.1002/jmri.21492.
72. Clauser P, Pinker K, Helbich TH, et al. Fat saturation in dynamic breast MRI at 3 Tesla: is the Dixon technique superior to spectral fat saturation? A visual grading characteristics study. Eur Radiol. 2014;24:2213–9. doi:10.1007/s00330-014-3189-7.
73. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74. doi:10.1016/j.cell.2011.02.013.
74. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.
75. Tennant DA, Durán RV, Gottlieb E. Targeting metabolic transformation for cancer therapy. Nat Rev Cancer. 2010;10:267–77. doi:10.1038/nrc2817.
76. Tennant DA, Durán RV, Boulahbel H, Gottlieb E. Metabolic transformation in cancer. Carcinogenesis. 2009;30:1269–80. doi:10.1093/carcin/bgp070.
77. Trosko JE, Chang C-C, Upham BL, Tai M-H. Ignored hallmarks of carcinogenesis: stem cells and cell-cell communication. Ann N Y Acad Sci. 2004;1028:192–201. doi:10.1196/annals.1322.023.
78. Houssami N, Ciatto S, Macaskill P, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. J Clin Oncol. 2008;26:3248–58. doi:10.1200/JCO.2007.15.2108.
79. Kuhl C. The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. Radiology. 2007;244:356–78. doi:10.1148/radiol.2442051620.
80. Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology. 2004;233:830–49. doi:10.1148/radiol.2333031484.
81. Mann RM, Balleyguier C, Baltzer PA, et al. Breast MRI: EUSOBI recommendations for women's information. Eur Radiol. 2015;25(12):3669–78. doi:10.1007/s00330-015-3807-z.
82. Spick C, Szolar DHM, Preidler KW, et al. Breast MRI used as a problem-solving tool reliably excludes malignancy. Eur J Radiol. 2015;84:61–4. doi:10.1016/j.ejrad.2014.10.005.
83. Benndorf M, Baltzer PAT, Vag T, et al. Breast MRI as an adjunct to mammography: does it really suffer from low specificity? A retrospective analysis stratified by mammographic BI-RADS classes. Acta Radiol. 2010;51:715–21. doi:10.3109/02841851.2010.497164.
84. Baltzer PAT, Benndorf M, Dietzel M, et al. Sensitivity and specificity of unenhanced MR mammography (DWI combined with T2-weighted TSE imaging, ueMRM) for the differentiation of mass lesions. Eur Radiol. 2010;20:1101–10. doi:10.1007/s00330-009-1654-5.
85. Pinker K, Bogner W, Baltzer P, et al. Improved diagnostic accuracy with multiparametric magnetic resonance imaging of the breast using dynamic contrast-enhanced magnetic resonance imaging, diffusion-weighted imaging, and 3-dimensional proton magnetic resonance spectroscopic imaging. Invest Radiol. 2014;49:421–30. doi:10.1097/RLI.0000000000000029.
86. Baltzer PAT, Schäfer A, Dietzel M, et al. Diffusion tensor magnetic resonance imaging of the breast: a pilot study. Eur Radiol. 2011;21:1–10. doi:10.1007/s00330-010-1901-9.
87. Le Bihan D, Mangin JF, Poupon C, et al. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging. 2001;13:534–46.
88. Gore JC, Manning HC, Quarles CC, et al. Magnetic resonance in the era of molecular imaging of cancer. Magn Reson Imaging. 2011;29:587–600. doi:10.1016/j.mri.2011.02.003.

89. Kuhl CK. Current status of breast MR imaging. Part 2. Clinical applications. Radiology. 2007;244:672–91. doi:10.1148/radiol.2443051661.
90. Folkman J. Seminars in Medicine of the Beth Israel Hospital, Boston. Clinical applications of research on angiogenesis. N Engl J Med. 1995;333:1757–63. doi:10.1056/NEJM199512283332608.
91. Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med. 1971;285:1182–6. doi:10.1056/NEJM197111182852108.
92. Brasch R, Turetschek K. MRI characterization of tumors and grading angiogenesis using macromolecular contrast media: status report. Eur J Radiol. 2000;34:148–55.
93. Warner E. The role of magnetic resonance imaging in screening women at high risk of breast cancer. Top Magn Reson Imaging. 2008;19:163–9. doi:10.1097/RMR.0b013e31818bc994.
94. Burnside ES, Sickles EA, Bassett LW, et al. The ACR BI-RADS experience: learning from history. J Am Coll Radiol. 2009;6:851–60. doi:10.1016/j.jacr.2009.07.023.
95. Kuhl CK, Jost P, Morakkabati N, et al. Contrast-enhanced MR imaging of the breast at 3.0 and 1.5 T in the same patients: initial experience. Radiology. 2006;239:666–76. doi:10.1148/radiol.2392050509.
96. Kaiser CG, Reich C, Dietzel M, et al. DCE-MRI of the breast in a stand-alone setting outside a complementary strategy – results of the TK-study. Eur Radiol. 2015;25:1793–800. doi:10.1007/s00330-014-3580-4.
97. Baltzer PAT, Benndorf M, Dietzel M, et al. False-positive findings at contrast-enhanced breast MRI: a BI-RADS descriptor study. AJR Am J Roentgenol. 2010;194:1658–63. doi:10.2214/AJR.09.3486.
98. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. J Clin Oncol. 2005;23:8469–76. doi:10.1200/JCO.2004.00.4960.
99. Kuhl CK. Breast MR imaging at 3T. Magn Reson Imaging Clin N Am. 2007;15:315–20. doi:10.1016/j.mric.2007.08.003, vi.
100. Korteweg MA, Veldhuis WB, Visser F, et al. Feasibility of 7 Tesla breast magnetic resonance imaging determination of intrinsic sensitivity and high-resolution magnetic resonance imaging, diffusion-weighted imaging, and (1)H-magnetic resonance spectroscopy of breast cancer patients receiving neoadjuvant therapy. Invest Radiol. 2011;46:370–6. doi:10.1097/RLI.0b013e31820df706.
101. Gruber S, Minarikova L, Pinker K, et al. Diffusion-weighted imaging of breast tumours at 3 Tesla and 7 Tesla: a comparison. Eur Radiol. 2015;26(5):1466–73. doi:10.1007/s00330-015-3947-1.
102. Pinker K, Baltzer P, Bogner W, et al. Multiparametric MR imaging with high-resolution dynamic contrast-enhanced and diffusion-weighted imaging at 7 T improves the assessment of breast tumors: a feasibility study. Radiology. 2015;276:360–70. doi:10.1148/radiol.15141905.
103. Rahbar H, Partridge SC. Multiparametric MR imaging of breast cancer. Magn Reson Imaging Clin N Am. 2016;24:223–38. doi:10.1016/j.mric.2015.08.012.
104. Baltzer PAT, Freiberg C, Beger S, et al. Clinical MR-mammography: are computer-assisted methods superior to visual or manual measurements for curve type analysis? A systematic approach. Acad Radiol. 2009;16:1070–6. doi:10.1016/j.acra.2009.03.017.
105. Baltzer PA, Dietzel M, Vag T, et al. Can color-coded parametric maps improve dynamic enhancement pattern analysis in MR mammography? RoFo. 2010;182:254–60. doi:10.1055/s-0028-1109843.
106. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10:223–32.
107. Tofts PS, Berkowitz B, Schnall MD. Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model. Magn Reson Med. 1995;33:564–8.
108. Buadu LD, Murakami J, Murayama S, et al. Breast lesions: correlation of contrast medium enhancement patterns on MR images with histopathologic findings and tumor angiogenesis. Radiology. 1996;200:639–49. doi:10.1148/radiology.200.3.8756909.
109. Tuncbilek N, Unlu E, Karakas HM, et al. Evaluation of tumor angiogenesis with contrast-enhanced dynamic magnetic resonance mammography. Breast J. 2003;9:403–8.
110. Agrawal G, Su M-Y, Nalcioglu O, et al. Significance of breast lesion descriptors in the ACR BI-RADS MRI lexicon. Cancer. 2009;115:1363–80. doi:10.1002/cncr.24156.
111. Pabst T, Kenn W, Kaiser WA, Hahn D. Understanding why contrast enhancement in dynamic MRI is not reproducible: illustration with a simple phantom. Breast J. 2001;7:166–70.
112. Zwick S, Brix G, Tofts PS, et al. Simulation-based comparison of two approaches frequently used for dynamic contrast-enhanced MRI. Eur Radiol. 2010;20:432–42. doi:10.1007/s00330-009-1556-6.
113. Kuhl CK, Kooijman H, Gieseke J, Schild HH. Effect of B1 inhomogeneity on breast MR imaging at 3.0 T. Radiology. 2007;244:929–30. doi:10.1148/radiol.2443070266.
114. Huang W, Tudorica LA, Li X, et al. Discrimination of benign and malignant breast lesions by using shutter-speed dynamic contrast-enhanced MR imaging. Radiology. 2011;261:394–403. doi:10.1148/radiol.11102413.
115. Yankeelov TE, Luci JJ, Lepage M, et al. Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: a reference region model. Magn Reson Imaging. 2005;23:519–29. doi:10.1016/j.mri.2005.02.013.
116. Huang W, Li X, Chen Y, et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. Transl Oncol. 2014;7:153–66.

117. Kaiser WA, Zeitler E. MR imaging of the breast: fast imaging sequences with and without Gd-DTPA. Preliminary observations. Radiology. 1989;170:681–6. doi:10.1148/radiology.170.3.2916021.
118. Kaiser WA. False-positive results in dynamic MR mammography. Causes, frequency, and methods to avoid. Magn Reson Imaging Clin N Am. 1994;2:539–55.
119. Li L, Wang K, Sun X, et al. Parameters of dynamic contrast-enhanced MRI as imaging markers for angiogenesis and proliferation in human breast cancer. Med Sci Monit. 2015;21:376–82. doi:10.12659/MSM.892534.
120. Helbich TH. Contrast-enhanced magnetic resonance imaging of the breast. Eur J Radiol. 2000;34:208–19.
121. Malich A, Fischer DR, Wurdinger S, et al. Potential MRI interpretation model: differentiation of benign from malignant breast masses. AJR Am J Roentgenol. 2005;185:964–70. doi:10.2214/AJR.04.1073.
122. Kaiser WA. Breast magnetic resonance imaging: principles and techniques. Semin Roentgenol. 2007;42:228–35. doi:10.1053/j.ro.2007.07.003.
123. Fischer DR, Wurdinger S, Boettcher J, et al. Further signs in the evaluation of magnetic resonance mammography: a retrospective study. Invest Radiol. 2005;40:430–5.
124. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology. 1999;211:101–10. doi:10.1148/radiology.211.1.r99ap38101.
125. Dietzel M, Kaiser C, Baltzer PAT. Magnetic resonance imaging of intraductal papillomas: typical findings and differential diagnosis. J Comput Assist Tomogr. 2014;39(2):176–84. doi:10.1097/RCT.0000000000000188.
126. Dietzel M, Baltzer PA, Vag T, et al. Magnetic resonance mammography of invasive lobular versus ductal carcinoma: systematic comparison of 811 patients reveals high diagnostic accuracy irrespective of typing. J Comput Assist Tomogr. 2010;34:587–95. doi:10.1097/RCT.0b013e3181db9f0e.
127. Vincensini D, Dedieu V, Eliat PA, et al. Magnetic resonance imaging measurements of vascular permeability and extracellular volume fraction of breast tumors by dynamic Gd-DTPA-enhanced relaxometry. Magn Reson Imaging. 2007;25:293–302. doi:10.1016/j.mri.2006.10.016.
128. Warren RML, Thompson D, Pointon LJ, et al. Evaluation of a prospective scoring system designed for a multicenter breast MR imaging screening study. Radiology. 2006;239:677–85. doi:10.1148/radiol.2393042007.
129. Vag T, Baltzer PAT, Dietzel M, et al. Kinetic characteristics of ductal carcinoma in situ (DCIS) in dynamic breast MRI using computer-assisted analysis. Acta Radiol. 2010;51:955–61. doi:10.3109/02841851.2010.508171.
130. Vag T, Baltzer PAT, Dietzel M, Kaiser WA. Kinetic characterization of non-mass lesions on breast MRI using manual and computer assisted methods. Eur J Radiol. 2012;81 Suppl 1:S177–8. doi:10.1016/S0720-048X(12)70072-4.
131. Baltzer PAT, Vag T, Dietzel M, et al. Computer-aided interpretation of dynamic magnetic resonance imaging reflects histopathology of invasive breast cancer. Eur Radiol. 2010;20:1563–71. doi:10.1007/s00330-010-1722-x.
132. Koo HR, Cho N, Song IC, et al. Correlation of perfusion parameters on dynamic contrast-enhanced MRI with prognostic factors and subtypes of breast cancers. J Magn Reson Imaging. 2012;36:145–51. doi:10.1002/jmri.23635.
133. Baltzer PAT, Zoubi R, Burmeister HP, et al. Computer assisted analysis of MR-mammography reveals association between contrast enhancement and occurrence of distant metastasis. Technol Cancer Res Treat. 2012;11:553–60.
134. Dietzel M, Baltzer PAT, Schön K, Kaiser WA. MR-mammography: high sensitivity but low specificity? New thoughts and fresh data on an old mantra. Eur J Radiol. 2012;81 Suppl 1:S30–2. doi:10.1016/S0720-048X(12)70012-8.
135. Chang Y-W, Kwon KH, Choi DL, et al. Magnetic resonance imaging of breast cancer and correlation with prognostic factors. Acta Radiol. 2009;50:990–8. doi:10.3109/02841850903225180.
136. Szabó BK, Aspelin P, Kristoffersen Wiberg M, et al. Invasive breast cancer: correlation of dynamic MR features with prognostic factors. Eur Radiol. 2003;13:2425–35. doi:10.1007/s00330-003-2000-y.
137. Narisada H, Aoki T, Sasaguri T, et al. Correlation between numeric gadolinium-enhanced dynamic MRI ratios and prognostic factors and histologic type of breast carcinoma. AJR Am J Roentgenol. 2006;187:297–306. doi:10.2214/AJR.05.0696.
138. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy – results from ACRIN 6657/I-SPY TRIAL. Radiology. 2012;263:663–72. doi:10.1148/radiol.12110748.
139. Ah-See M-LW, Makris A, Taylor NJ, et al. Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer. Clin Cancer Res. 2008;14:6580–9. doi:10.1158/1078-0432.CCR-07-4310.
140. Pickles MD, Lowry M, Manton DJ, et al. Role of dynamic contrast enhanced MRI in monitoring early response of locally advanced breast cancer to neoadjuvant chemotherapy. Breast Cancer Res Treat. 2005;91:1–10. doi:10.1007/s10549-004-5819-2.
141. Wasser K, Klein SK, Fink C, et al. Evaluation of neoadjuvant chemotherapeutic response of breast cancer using dynamic MRI with high temporal resolution. Eur Radiol. 2003;13:80–7. doi:10.1007/s00330-002-1654-1.
142. Woolf DK, Padhani AR, Taylor NJ, et al. Assessing response in breast cancer with dynamic contrast-

enhanced magnetic resonance imaging: are signal intensity-time curves adequate? Breast Cancer Res Treat. 2014;147:335–43. doi:10.1007/s10549-014-3072-x.
143. Schipper R-J, Paiman M-L, Beets-Tan RGH, et al. Diagnostic performance of dedicated axillary T2- and diffusion-weighted MR imaging for nodal staging in breast cancer. Radiology. 2015;275:345–55. doi:10.1148/radiol.14141167.
144. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13:2329–34. doi:10.1158/1078-0432.CCR-06-1109.
145. Fernández-Morales LA, Seguí MA, Andreu X, et al. Analysis of the pathologic response to primary chemotherapy in patients with locally advanced breast cancer grouped according to estrogen receptor, progesterone receptor, and HER2 status. Clin Breast Cancer. 2007;7:559–64. doi:10.3816/CBC.2007.n.012.
146. Bufi E, Belli P, Di Matteo M, et al. Effect of breast cancer phenotype on diagnostic performance of MRI in the prediction to response to neoadjuvant treatment. Eur J Radiol. 2014;83:1631–8. doi:10.1016/j.ejrad.2014.05.002.
147. Hayashi Y, Takei H, Nozu S, et al. Analysis of complete response by MRI following neoadjuvant chemotherapy predicts pathological tumor responses differently for molecular subtypes of breast cancer. Oncol Lett. 2013;5:83–9. doi:10.3892/ol.2012.1004.
148. Kuzucan A, Chen J-H, Bahri S, et al. Diagnostic performance of magnetic resonance imaging for assessing tumor response in patients with HER2-negative breast cancer receiving neoadjuvant chemotherapy is associated with molecular biomarker profile. Clin Breast Cancer. 2012;12:110–8. doi:10.1016/j.clbc.2012.01.009.
149. Schrading S, Kuhl CK. Breast cancer: influence of taxanes on response assessment with dynamic contrast-enhanced MR imaging. Radiology. 2015;277(3):687–96. doi:10.1148/radiol.2015150006.
150. Baltzer PA, Renz DM, Herrmann K-H, et al. Diffusion-weighted imaging (DWI) in MR mammography (MRM): clinical comparison of echo planar imaging (EPI) and half-Fourier single-shot turbo spin echo (HASTE) diffusion techniques. Eur Radiol. 2009;19:1612–20. doi:10.1007/s00330-009-1326-5.
151. Marini C, Iacconi C, Giannelli M, et al. Quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesion. Eur Radiol. 2007;17:2646–55. doi:10.1007/s00330-007-0621-2.
152. Guo Y, Cai Y-Q, Cai Z-L, et al. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. J Magn Reson Imaging. 2002;16:172–8. doi:10.1002/jmri.10140.
153. Woodhams R, Kakita S, Hata H, et al. Identification of residual breast carcinoma following neoadjuvant chemotherapy: diffusion-weighted imaging – comparison with contrast-enhanced MR imaging and pathologic findings. Radiology. 2010;254:357–66. doi:10.1148/radiol.2542090405.
154. Pinker K, Stadlbauer A, Bogner W, et al. Molecular imaging of cancer: MR spectroscopy and beyond. Eur J Radiol. 2012;81:566–77. doi:10.1016/j.ejrad.2010.04.028.
155. Sinha S, Sinha U. Recent advances in breast MRI and MRS. NMR Biomed. 2009;22:3–16. doi:10.1002/nbm.1270.
156. Costantini M, Belli P, Rinaldi P, et al. Diffusion-weighted imaging in breast cancer: relationship between apparent diffusion coefficient and tumour aggressiveness. Clin Radiol. 2010;65:1005–12. doi:10.1016/j.crad.2010.07.008.
157. Bogner W, Gruber S, Pinker K, et al. Diffusion-weighted MR for differentiation of breast lesions at 3.0 T: how does selection of diffusion protocols affect diagnosis? Radiology. 2009;253:341–51. doi:10.1148/radiol.2532081718.
158. Dorrius MD, Dijkstra H, Oudkerk M, Sijens PE. Effect of b value and pre-admission of contrast on diagnostic accuracy of 1.5-T breast DWI: a systematic review and meta-analysis. Eur Radiol. 2014;24:2835–47. doi:10.1007/s00330-014-3338-z.
159. Partridge SC, DeMartini WB, Kurland BF, et al. Quantitative diffusion-weighted imaging as an adjunct to conventional breast MRI for improved positive predictive value. AJR Am J Roentgenol. 2009;193:1716–22. doi:10.2214/AJR.08.2139.
160. Kul S, Cansu A, Alhan E, et al. Contribution of diffusion-weighted imaging to dynamic contrast-enhanced MRI in the characterization of breast tumors. AJR Am J Roentgenol. 2011;196:210–7. doi:10.2214/AJR.10.4258.
161. Schmitz AMT, Veldhuis WB, Menke-Pluijmers MBE, et al. Multiparametric MRI with dynamic contrast enhancement, diffusion-weighted imaging, and 31-phosphorus spectroscopy at 7 T for characterization of breast cancer. Invest Radiol. 2015;50:766–71. doi:10.1097/RLI.0000000000000183.
162. Ei Khouli RH, Jacobs MA, Mezban SD, et al. Diffusion-weighted imaging improves the diagnostic accuracy of conventional 3.0-T breast MR imaging. Radiology. 2010;256:64–73. doi:10.1148/radiol.10091367.
163. Yabuuchi H, Matsuo Y, Okafuji T, et al. Enhanced mass on contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. J Magn Reson Imaging. 2008;28:1157–65. doi:10.1002/jmri.21570.
164. Yabuuchi H, Matsuo Y, Kamitani T, et al. Non-mass-like enhancement on contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. Eur J Radiol. 2010;75:e126–32. doi:10.1016/j.ejrad.2009.09.013.
165. Pinker K, Bickel H, Helbich TH, et al. Combined contrast-enhanced magnetic resonance and

diffusion-weighted imaging reading adapted to the "Breast Imaging Reporting and Data System" for multiparametric 3-T imaging of breast lesions. Eur Radiol. 2013;23:1791–802. doi:10.1007/s00330-013-2771-8.
166. Baltzer A, Dietzel M, Kaiser CG, Baltzer PA. Combined reading of Contrast Enhanced and Diffusion Weighted Magnetic Resonance Imaging by using a simple sum score. Eur Radiol. 2015;26:884–91. doi:10.1007/s00330-015-3886-x.
167. Chen X, Li W, Zhang Y, et al. Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. BMC Cancer. 2010;10:693. doi:10.1186/1471-2407-10-693.
168. Tsushima Y, Takahashi-Taketomi A, Endo K. Magnetic resonance (MR) differential diagnosis of breast tumors using apparent diffusion coefficient (ADC) on 1.5-T. J Magn Reson Imaging. 2009;30:249–55. doi:10.1002/jmri.21854.
169. Hatakenaka M, Soeda H, Yabuuchi H, et al. Apparent diffusion coefficients of breast tumors: clinical application. Magn Reson Med Sci. 2008;7:23–9.
170. Woodhams R, Matsunaga K, Iwabuchi K, et al. Diffusion-weighted imaging of malignant breast tumors: the usefulness of apparent diffusion coefficient (ADC) value and ADC map for the detection of malignant breast tumors and evaluation of cancer extension. J Comput Assist Tomogr. 2005;29:644–9.
171. Stadlbauer A, Bernt R, Gruber S, et al. Diffusion-weighted MR imaging with background body signal suppression (DWIBS) for the diagnosis of malignant and benign breast lesions. Eur Radiol. 2009;19:2349–56. doi:10.1007/s00330-009-1426-2.
172. Spick C, Pinker-Domenig K, Rudas M, et al. MRI-only lesions: application of diffusion-weighted imaging obviates unnecessary MR-guided breast biopsies. Eur Radiol. 2014;24:1204–10. doi:10.1007/s00330-014-3153-6.
173. Bickel H, Pinker-Domenig K, Bogner W, et al. Quantitative apparent diffusion coefficient as a noninvasive imaging biomarker for the differentiation of invasive breast cancer and ductal carcinoma in situ. Invest Radiol. 2015;50:95–100. doi:10.1097/RLI.0000000000000104.
174. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. Breast Cancer Res Treat. 2006;97:135–44. doi:10.1007/s10549-005-9101-z.
175. Martincich L, Deantoni V, Bertotto I, et al. Correlations between diffusion-weighted imaging and breast cancer biomarkers. Eur Radiol. 2012;22:1519–28. doi:10.1007/s00330-012-2403-8.
176. Uematsu T, Kasami M, Yuen S. Triple-negative breast cancer: correlation between MR imaging and pathologic findings. Radiology. 2009;250:638–47. doi:10.1148/radiol.2503081054.
177. Dogan BE, Gonzalez-Angulo AM, Gilcrease M, et al. Multimodality imaging of triple receptor-negative tumors with mammography, ultrasound, and MRI. AJR Am J Roentgenol. 2010;194:1160–6. doi:10.2214/AJR.09.2355.
178. Kawashima H. Imaging findings of triple-negative breast cancer. Breast Cancer. 2011;18:145. doi:10.1007/s12282-010-0247-0.
179. Woodhams R, Ramadan S, Stanwell P, et al. Diffusion-weighted imaging of the breast: principles and clinical applications. Radiographics. 2011;31:1059–84. doi:10.1148/rg.314105160.
180. Rahbar H, Partridge SC, Eby PR, et al. Characterization of ductal carcinoma in situ on diffusion weighted breast MRI. Eur Radiol. 2011;21:2011–9. doi:10.1007/s00330-011-2140-4.
181. Rahbar H, Partridge SC, Demartini WB, et al. In vivo assessment of ductal carcinoma in situ grade: a model incorporating dynamic contrast-enhanced and diffusion-weighted breast MR imaging parameters. Radiology. 2012;263:374–82. doi:10.1148/radiol.12111368.
182. Pickles MD, Gibbs P, Lowry M, Turnbull LW. Diffusion changes precede size reduction in neoadjuvant treatment of breast cancer. Magn Reson Imaging. 2006;24:843–7. doi:10.1016/j.mri.2005.11.005.
183. Sharma U, Danishad KKA, Seenu V, Jagannathan NR. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. NMR Biomed. 2009;22:104–13. doi:10.1002/nbm.1245.
184. Park SH, Moon WK, Cho N, et al. Diffusion-weighted MR imaging: pretreatment prediction of response to neoadjuvant chemotherapy in patients with breast cancer. Radiology. 2010;257:56–63. doi:10.1148/radiol.10092021.
185. Richard R, Thomassin I, Chapellier M, et al. Diffusion-weighted MRI in pretreatment prediction of response to neoadjuvant chemotherapy in patients with breast cancer. Eur Radiol. 2013;23:2420–31. doi:10.1007/s00330-013-2850-x.
186. Iacconi C, Giannelli M, Marini C, et al. The role of mean diffusivity (MD) as a predictive index of the response to chemotherapy in locally advanced breast cancer: a preliminary study. Eur Radiol. 2010;20:303–8. doi:10.1007/s00330-009-1550-z.
187. Cakir O, Arslan A, Inan N, et al. Comparison of the diagnostic performances of diffusion parameters in diffusion weighted imaging and diffusion tensor imaging of breast lesions. Eur J Radiol. 2013;82:e801–6. doi:10.1016/j.ejrad.2013.09.001.
188. Eyal E, Shapiro-Feinberg M, Furman-Haran E, et al. Parametric diffusion tensor imaging of the breast. Invest Radiol. 2012;47:284–91. doi:10.1097/RLI.0b013e3182438e5d.
189. Partridge SC, Ziadloo A, Murthy R, et al. Diffusion tensor MRI: preliminary anisotropy measures and mapping of breast tumors. J Magn Reson Imaging. 2010;31:339–47. doi:10.1002/jmri.22045.
190. Jensen JH, Helpern JA, Ramani A, et al. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance

imaging. Magn Reson Med. 2005;53:1432–40. doi:10.1002/mrm.20508.
191. Nogueira L, Brandão S, Matos E, et al. Application of the diffusion kurtosis model for the study of breast lesions. Eur Radiol. 2014;24:1197–203. doi:10.1007/s00330-014-3146-5.
192. Sun K, Chen X, Chai W, et al. Breast cancer: diffusion kurtosis mr imaging-diagnostic accuracy and correlation with clinical-pathologic factors. Radiology. 2015;277:46–55. doi:10.1148/radiol.15141625.
193. Katz-Brull R, Lavin PT, Lenkinski RE. Clinical utility of proton magnetic resonance spectroscopy in characterizing breast lesions. J Natl Cancer Inst. 2002;94:1197–203.
194. Begley JK, Redpath TW, Bolan PJ, Gilbert FJ. In vivo proton magnetic resonance spectroscopy of breast cancer: a review of the literature. Breast Cancer Res. 2012;14:207. doi:10.1186/bcr3132.
195. Gruber S, Debski B-K, Pinker K, et al. Three-dimensional proton MR spectroscopic imaging at 3 T for the differentiation of benign and malignant breast lesions. Radiology. 2011;261:752–61. doi:10.1148/radiol.11102096.
196. Meisamy S, Bolan PJ, Baker EH, et al. Adding in vivo quantitative 1H MR spectroscopy to improve diagnostic accuracy of breast MR imaging: preliminary results of observer performance study at 4.0 T. Radiology. 2005;236:465–75. doi:10.1148/radiol.2362040836.
197. Bartella L, Morris EA, Dershaw DD, et al. Proton MR spectroscopy with choline peak as malignancy marker improves positive predictive value for breast cancer diagnosis: preliminary study. Radiology. 2006;239:686–92. doi:10.1148/radiol.2393051046.
198. Bartella L, Thakur SB, Morris EA, et al. Enhancing nonmass lesions in the breast: evaluation with proton (1H) MR spectroscopy. Radiology. 2007;245:80–7. doi:10.1148/radiol.2451061639.
199. Ackerstaff E, Glunde K, Bhujwalla ZM. Choline phospholipid metabolism: a target in cancer cells? J Cell Biochem. 2003;90:525–33. doi:10.1002/jcb.10659.
200. Glunde K, Bhujwalla ZM, Ronen SM. Choline metabolism in malignant transformation. Nat Rev Cancer. 2011;11:835–48. doi:10.1038/nrc3162.
201. Podo F. Tumour phospholipid metabolism. NMR Biomed. 1999;12:413–39.
202. Signs in MR-mammography.pdf.
203. Mountford C, Ramadan S, Stanwell P, Malycha P. Proton MRS of the breast in the clinical setting. NMR Biomed. 2009;22:54–64. doi:10.1002/nbm.1301.
204. Haddadin IS, McIntosh A, Meisamy S, et al. Metabolite quantification and high-field MRS in breast cancer. NMR Biomed. 2009;22:65–76. doi:10.1002/nbm.1217.
205. Magometschnigg HF, Helbich T, Brader P, et al. Molecular imaging for the characterization of breast tumors. Expert Rev Anticancer Ther. 2014;14:711–22. doi:10.1586/14737140.2014.885383.
206. Baltzer PAT, Dietzel M, Kaiser WA. MR-spectroscopy at 1.5 tesla and 3 tesla. Useful? A systematic review and meta-analysis. Eur J Radiol. 2012;81 Suppl 1:S6–9. doi:10.1016/S0720-048X(12)70003-7.
207. Stanwell P, Mountford C. In vivo proton MR spectroscopy of the breast. Radiographics. 2007;27 Suppl 1:S253–66. doi:10.1148/rg.27si075519.
208. Pinker K, Helbich TH, Magometschnigg H, et al. Molecular breast imaging. An update. Radiologe. 2014;54:241–53. doi:10.1007/s00117-013-2580-8.
209. Baltzer PAT, Dietzel M. Breast lesions: diagnosis by using proton MR spectroscopy at 1.5 and 3.0 T – systematic review and meta-analysis. Radiology. 2013;267:735–46. doi:10.1148/radiol.13121856.
210. Shin HJ, Baek H-M, Cha JH, Kim HH. Evaluation of breast cancer using proton MR spectroscopy: total choline peak integral and signal-to-noise ratio as prognostic indicators. AJR Am J Roentgenol. 2012;198:W488–97. doi:10.2214/AJR.11.7292.
211. Aboagye EO, Bhujwalla ZM. Malignant transformation alters membrane choline phospholipid metabolism of human mammary epithelial cells. Cancer Res. 1999;59:80–4.
212. for the European Society of Breast Imaging (EUSOBI), Sardanelli F, Helbich TH. Mammography: EUSOBI recommendations for women's information. Insights Imaging. 2012;3:7–10. doi:10.1007/s13244-011-0127-y.
213. Sardanelli F, Boetes C, Borisch B, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. Eur J Cancer. 2010;46:1296–316. doi:10.1016/j.ejca.2010.02.015.
214. Ramadan S, Arm J, Silcock J, et al. Lipid and metabolite deregulation in the breast tissue of women carrying BRCA1 and BRCA2 genetic mutations. Radiology. 2015;275:675–82. doi:10.1148/radiol.15140967.
215. Jagannathan NR, Kumar M, Seenu V, et al. Evaluation of total choline from in-vivo volume localized proton MR spectroscopy and its response to neoadjuvant chemotherapy in locally advanced breast cancer. Br J Cancer. 2001;84:1016–22. doi:10.1054/bjoc.2000.1711.
216. Danishad KKA, Sharma U, Sah RG, et al. Assessment of therapeutic response of locally advanced breast cancer (LABC) patients undergoing neoadjuvant chemotherapy (NACT) monitored using sequential magnetic resonance spectroscopic imaging (MRSI). NMR Biomed. 2010;23:233–41. doi:10.1002/nbm.1436.
217. Sharma U, Baek HM, Su MY, Jagannathan NR. In vivo 1H MRS in the assessment of the therapeutic response of breast cancer patients. NMR Biomed. 2011;24:700–11. doi:10.1002/nbm.1654.
218. Mercado CL, Moy L. Breast MRI. Preface. Magn Reson Imaging Clin N Am. 2010;18:xiii. doi:10.1016/j.mric.2010.02.014.
219. Watermann DO, Földi M, Hanjalic-Beck A, et al. Three-dimensional ultrasound for the assessment of breast lesions. Ultrasound Obstet Gynecol. 2005;25:592–8. doi:10.1002/uog.1909.

220. Gruber R, Jaromi S, Rudas M, et al. Histologic work-up of non-palpable breast lesions classified as probably benign at initial mammography and/or ultrasound (BI-RADS category 3). Eur J Radiol. 2013;82:398–403. doi:10.1016/j.ejrad.2012.02.004.
221. Gartlehner G, Thaler KJ, Chapman A, et al. Adjunct ultrasonography for breast cancer screening in women at average risk: a systematic review. Int J Evid Based Healthc. 2013;11:87–93. doi:10.1111/1744-1609.12022.
222. Reisegger M, Schueller G, Gruber R, et al. Imaging of molecular structures of breasts with new sonography techniques. Radiologe. 2010;50:1014–21. doi:10.1007/s00117-010-2013-x.
223. Scaperrotta G, Ferranti C, Costa C, et al. Role of sonoelastography in non-palpable breast lesions. Eur Radiol. 2008;18:2381–9. doi:10.1007/s00330-008-1032-8.
224. Thomas A, Kümmel S, Fritzsche F, et al. Real-time sonoelastography performed in addition to B-mode ultrasound and mammography: improved differentiation of breast lesions? Acad Radiol. 2006;13:1496–504. doi:10.1016/j.acra.2006.08.012.
225. Berg WA, Cosgrove DO, Doré CJ, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. Radiology. 2012;262:435–49. doi:10.1148/radiol.11110640.
226. Cho N, Jang M, Lyou CY, et al. Distinguishing benign from malignant masses at breast US: combined US elastography and color doppler US--influence on radiologist accuracy. Radiology. 2012;262:80–90. doi:10.1148/radiol.11110886.
227. Calas MJG, Almeida RMVR, Gutfilen B, Pereira WCA. Interobserver concordance in the BI-RADS classification of breast ultrasound exams. Clinics (Sao Paulo). 2012;67:185–9.
228. Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. Radiology. 2006;241:355–65. doi:10.1148/radiol.2412051710.
229. Weismann C, Mayr C, Egger H, Auer A. Breast sonography – 2D, 3D, 4D ultrasound or elastography? Breast Care (Basel). 2011;6:98–103. doi:10.1159/000327504.
230. Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. Breast J. 2013;19:64–70. doi:10.1111/tbj.12053.
231. Chang JM, Cha JH, Park JS, et al. Automated breast ultrasound system (ABUS): reproducibility of mass localization, size measurement, and characterization on serial examinations. Acta Radiol. 2015;56:1163–70. doi:10.1177/0284185114551565.
232. Brem RF, Tabár L, Duffy SW, et al. Assessing improvement in detection of breast cancer with three-dimensional automated breast US in women with dense breast tissue: the SomoInsight Study. Radiology. 2015;274:663–73. doi:10.1148/radiol.14132832.
233. Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. AJR Am J Roentgenol. 2015;204:234–40. doi:10.2214/AJR.13.12072.
234. Kettenbach J, Helbich TH, Huber S, et al. Computer-assisted quantitative assessment of power Doppler US: effects of microbubble contrast agent in the differentiation of breast tumors. Eur J Radiol. 2005;53:238–44. doi:10.1016/j.ejrad.2004.04.017.
235. Krestan CR, Riedl C, Memarsadeghi M, et al. 3D-power Doppler ultrasound of breast lesions with a microbubble contrast agent. Acad Radiol. 2002;9 Suppl 2:S384–5.
236. Lindner JR. Microbubbles in medical imaging: current applications and future directions. Nat Rev Drug Discov. 2004;3:527–32. doi:10.1038/nrd1417.
237. Willmann JK, van Bruggen N, Dinkelborg LM, Gambhir SS. Molecular imaging in drug development. Nat Rev Drug Discov. 2008;7:591–607. doi:10.1038/nrd2290.
238. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. Radiology. 2006;239:341–50. doi:10.1148/radiol.2391041676.
239. Insana MF, Pellot-Barakat C, Sridhar M, Lindfors KK. Viscoelastic imaging of breast tumor microenvironment with ultrasound. J Mammary Gland Biol Neoplasia. 2004;9:393–404. doi:10.1007/s10911-004-1409-5.
240. Zhi H, Xiao X-Y, Yang H-Y, et al. Ultrasonic elastography in breast cancer diagnosis: strain ratio vs 5-point scale. Acad Radiol. 2010;17:1227–33. doi:10.1016/j.acra.2010.05.004.
241. Ginat DT, Destounis SV, Barr RG, et al. US elastography of breast and prostate lesions. Radiographics. 2009;29:2007–16. doi:10.1148/rg.297095058.
242. Evans A, Whelehan P, Thomson K, et al. Quantitative shear wave ultrasound elastography: initial experience in solid breast masses. Breast Cancer Res. 2010;12:R104. doi:10.1186/bcr2787.
243. Nightingale K, Bentley R, Trahey G. Observations of tissue response to acoustic radiation force: opportunities for imaging. Ultrason Imaging. 2002;24:129–38.
244. Hoyt K, Parker KJ, Rubens DJ. Real-time shear velocity imaging using sonoelastographic techniques. Ultrasound Med Biol. 2007;33:1086–97. doi:10.1016/j.ultrasmedbio.2007.01.009.
245. Nightingale K, Nightingale R, Stutz D, Trahey G. Acoustic radiation force impulse imaging of in vivo vastus medialis muscle under varying isometric load. Ultrason Imaging. 2002;24:100–8.
246. Ophir J, Alam SK, Garra B, et al. Elastography: ultrasonic estimation and imaging of the elastic properties of tissues. Proc Inst Mech Eng H. 1999;213:203–33.
247. Garra BS. Elastography: current status, future prospects, and making it work for you. Ultrasound Q. 2011;27:177–86. doi:10.1097/RUQ.0b013e31822a2138.
248. Bercoff J, Tanter M, Fink M. Supersonic shear imaging: a new technique for soft tissue elasticity map-

ping. IEEE Trans Ultrason Ferroelectr Freq Control. 2004;51:396–409.
249. Maeda M. New angiographic knowledge regarding the female breast (author's transl). RoFo. 1979;130:711–5. doi:10.1055/s-0029-1231359.
250. Ianculescu V, Ciolovan LM, Dunant A, et al. Added value of Virtual Touch IQ shear wave elastography in the ultrasound assessment of breast lesions. Eur J Radiol. 2014;83:773–7. doi:10.1016/j.ejrad.2014.01.021.
251. Kapetas P, Woitek R, Clauser P, Marino MA, Mahajan MD, Pinker-Domenig K, Bernathova M, Helbich TH, Baltzer PA. Diagnostic performance and reproducibility of breast acoustic radiation force impulse (ARFI) imaging in the clinical setting-BR266-SD-WEB7.
252. Hayashi M, Yamamoto Y, Ibusuki M, et al. Evaluation of tumor stiffness by elastography is predictive for pathologic complete response to neoadjuvant chemotherapy in patients with breast cancer. Ann Surg Oncol. 2012;19:3042–9. doi:10.1245/s10434-012-2343-1.
253. Falou O, Sadeghi-Naini A, Prematilake S, et al. Evaluation of neoadjuvant chemotherapy response in women with locally advanced breast cancer using ultrasound elastography. Transl Oncol. 2013;6:17–24.
254. Athanasiou A, Latorre-Ossa H, Criton A, et al. Feasibility of imaging and treatment monitoring of breast lesions with three-dimensional shear wave elastography. Ultraschall Med. 2015. doi:10.1055/s-0034-1398980.
255. del Cura JL, Elizagaray E, Zabala R, et al. The use of unenhanced Doppler sonography in the evaluation of solid breast lesions. AJR Am J Roentgenol. 2005;184:1788–94. doi:10.2214/ajr.184.6.01841788.
256. Svensson WE, Pandian AJ, Hashimoto H. The use of breast ultrasound color Doppler vascular pattern morphology improves diagnostic sensitivity with minimal change in specificity. Ultraschall Med. 2010;31:466–74. doi:10.1055/s-0028-1109478.
257. Svensson WE, Pandian AJ, Hashimoto H. The use of breast ultrasound color Doppler vascular pattern morphology improves diagnostic sensitivity with minimal change in specificity. Ultraschall Med. 2010;31:466–74. doi:10.1055/s-0028-1109478.
258. Delorme S, Zuna I, Huber S, et al. Colour Doppler sonography in breast tumours: an update. Eur Radiol. 1998;8:189–93. doi:10.1007/s003300050361.
259. Delorme S, Weisser G, Zuna I, et al. Quantitative characterization of color Doppler images: reproducibility, accuracy, and limitations. J Clin Ultrasound. 1995;23:537–50.
260. Bruening W, Uhl S, Fontanarosa J, et al. Noninvasive diagnostic tests for breast abnormalities: update of a 2006 review. Rockville: Agency for Healthcare Research and Quality; 2012.
261. Ozdemir A, Ozdemir H, Maral I, et al. Differential diagnosis of solid breast lesions: contribution of Doppler studies to mammography and gray scale imaging. J Ultrasound Med. 2001;20:1091–101; quiz 1102.
262. Moon WK, Im JG, Noh DY, Han MC. Nonpalpable breast lesions: evaluation with power Doppler US and a microbubble contrast agent-initial experience. Radiology. 2000;217:240–6. doi:10.1148/radiology.217.1.r00oc03240.
263. Huber S, Helbich T, Kettenbach J, et al. Effects of a microbubble contrast agent on breast tumors: computer-assisted quantitative assessment with color Doppler US – early experience. Radiology. 1998;208:485–9. doi:10.1148/radiology.208.2.9680580.
264. Kedar RP, Cosgrove D, McCready VR, et al. Microbubble contrast agent for color Doppler US: effect on breast masses. Work in progress. Radiology. 1996;198:679–86. doi:10.1148/radiology.198.3.8628854.
265. Kumar A, Singh S, Pradhan S, et al. Doppler ultrasound scoring to predict chemotherapeutic response in advanced breast cancer. World J Surg Oncol. 2007;5:99. doi:10.1186/1477-7819-5-99.
266. Huber S, Medl M, Helbich T, et al. Locally advanced breast carcinoma: computer assisted semiquantitative analysis of color Doppler ultrasonography in the evaluation of tumor response to neoadjuvant chemotherapy (work in progress). J Ultrasound Med. 2000;19:601–7.
267. Catalano O, Nunziata A, Lobianco R, Siani A. Real-time harmonic contrast material-specific US of focal liver lesions. Radiographics. 2005;25:333–49. doi:10.1148/rg.252045066.
268. Balleyguier C, Opolon P, Mathieu MC, et al. New potential and applications of contrast-enhanced ultrasound of the breast: own investigations and review of the literature. Eur J Radiol. 2009;69:14–23. doi:10.1016/j.ejrad.2008.07.037.
269. Wan C, Du J, Fang H, et al. Evaluation of breast lesions by contrast enhanced ultrasound: qualitative and quantitative analysis. Eur J Radiol. 2012;81:e444–50. doi:10.1016/j.ejrad.2011.03.094.
270. Dayton PA, Rychak JJ. Molecular ultrasound imaging using microbubble contrast agents. Front Biosci. 2007;12:5124–42.
271. Sorace AG, Saini R, Mahoney M, Hoyt K. Molecular ultrasound imaging using a targeted contrast agent for assessing early tumor response to antiangiogenic therapy. J Ultrasound Med. 2012;31:1543–50.
272. Pochon S, Tardy I, Bussat P, et al. BR55: a lipopeptide-based VEGFR2-targeted ultrasound contrast agent for molecular imaging of angiogenesis. Invest Radiol. 2010;45:89–95. doi:10.1097/RLI.0b013e3181c5927c.
273. Anderson CR, Hu X, Zhang H, et al. Ultrasound molecular imaging of tumor angiogenesis with an integrin targeted microbubble contrast agent. Invest Radiol. 2011;46:215–24. doi:10.1097/RLI.0b013e3182034fed.
274. Warram JM, Sorace AG, Saini R, et al. A triple-targeted ultrasound contrast agent provides improved localization to tumor vasculature. J Ultrasound Med. 2011;30:921–31.

275. Rix A, Palmowski M, Kiessling F. Molecular ultrasound imaging: clinical applications. Radiologe. 2015;55:956–63. doi:10.1007/s00117-015-0023-4.
276. Wan CF, Du J, Fang H, et al. Enhancement patterns and parameters of breast cancers at contrast-enhanced US: correlation with prognostic factors. Radiology. 2012;262:450–9. doi:10.1148/radiol.11110789.
277. Caproni N, Marchisio F, Pecchi A, et al. Contrast-enhanced ultrasound in the characterisation of breast masses: utility of quantitative analysis in comparison with MRI. Eur Radiol. 2010;20:1384–95. doi:10.1007/s00330-009-1690-1.
278. Willmann JK, Lutz AM, Paulmurugan R, et al. Dual-targeted contrast agent for US assessment of tumor angiogenesis in vivo. Radiology. 2008;248:936–44. doi:10.1148/radiol.2483072231.
279. Bachawal SV, Jensen KC, Lutz AM, et al. Earlier detection of breast cancer with ultrasound molecular imaging in a transgenic mouse model. Cancer Res. 2013;73:1689–98. doi:10.1158/0008-5472.CAN-12-3391.
280. Bachawal SV, Jensen KC, Wilson KE, et al. Breast cancer detection by B7-H3-targeted ultrasound molecular imaging. Cancer Res. 2015;75:2501–9. doi:10.1158/0008-5472.CAN-14-3361.
281. Jung EM, Jungius K-P, Rupp N, et al. Contrast enhanced harmonic ultrasound for differentiating breast tumors – first results. Clin Hemorheol Microcirc. 2005;33:109–20.
282. Hoyt K, Warram JM, Umphrey H, et al. Determination of breast cancer response to bevacizumab therapy using contrast-enhanced ultrasound and artificial neural networks. J Ultrasound Med. 2010;29:577–85.
283. Wang J-W, Zheng W, Liu J-B, et al. Assessment of early tumor response to cytotoxic chemotherapy with dynamic contrast-enhanced ultrasound in human breast cancer xenografts. PLoS One. 2013;8:e58274. doi:10.1371/journal.pone.0058274.
284. van de Bank BL, Voogt IJ, Italiaander M, et al. Ultra high spatial and temporal resolution breast imaging at 7T. NMR Biomed. 2013;26:367–75. doi:10.1002/nbm.2868.
285. Klomp DWJ, van de Bank BL, Raaijmakers A, et al. 31P MRSI and 1H MRS at 7 T: initial results in human breast cancer. NMR Biomed. 2011;24:1337–42. doi:10.1002/nbm.1696.
286. Stehouwer BL, Klomp DWJ, Korteweg MA, et al. 7 T versus 3T contrast-enhanced breast magnetic resonance imaging of invasive ductulolobular carcinoma: first clinical experience. Magn Reson Imaging. 2013;31:613–7. doi:10.1016/j.mri.2012.09.005.
287. Brown R, Storey P, Geppert C, et al. Breast MRI at 7 Tesla with a bilateral coil and robust fat suppression. J Magn Reson Imaging. 2014;39:540–9. doi:10.1002/jmri.24205.
288. Brown R, Storey P, Geppert C, et al. Breast MRI at 7 Tesla with a bilateral coil and T1-weighted acquisition with robust fat suppression: image evaluation and comparison with 3 Tesla. Eur Radiol. 2013;23:2969–78. doi:10.1007/s00330-013-2972-1.
289. Pinker K, Bogner W, Baltzer P, et al. Clinical application of bilateral high temporal and spatial resolution dynamic contrast-enhanced magnetic resonance imaging of the breast at 7 T. Eur Radiol. 2014;24:913–20. doi:10.1007/s00330-013-3075-8.
290. Pinker K, Grabner G, Bogner W, et al. A combined high temporal and high spatial resolution 3 Tesla MR imaging protocol for the assessment of breast lesions: initial results. Invest Radiol. 2009;44:553–8. doi:10.1097/RLI.0b013e3181b4c127.
291. Bogner W, Pinker K, Zaric O, et al. Bilateral diffusion-weighted MR imaging of breast tumors with submillimeter resolution using readout-segmented echo-planar imaging at 7 T. Radiology. 2015;274:74–84. doi:10.1148/radiol.14132340.
292. Madelin G, Regatte RR. Biomedical applications of sodium MRI in vivo. J Magn Reson Imaging. 2013;38:511–29. doi:10.1002/jmri.24168.
293. Ouwerkerk R. Sodium MRI. Methods Mol Biol. 2011;711:175–201. doi:10.1007/978-1-61737-992-5_8.
294. Ouwerkerk R, Jacobs MA, Macura KJ, et al. Elevated tissue sodium concentration in malignant breast lesions detected with non-invasive 23Na MRI. Breast Cancer Res Treat. 2007;106:151–60. doi:10.1007/s10549-006-9485-4.
295. Zaric O, Pinker K, Zbyn S, Strasser S, Robinson S, Minarikova L, Gruber S, Farr A, Singer C, Helbich TH, Trattnig S, Bogner W. Quantitative sodium MR imaging at 7 Tesla – initial results and comparison with diffusion-weighted imaging in patients with breast tumors. Radiology. 2015;280(1):39–48.
296. Arias-Mendoza F, Payne GS, Zakian KL, et al. In vivo 31P MR spectral patterns and reproducibility in cancer patients studied in a multi-institutional trial. NMR Biomed. 2006;19:504–12. doi:10.1002/nbm.1057.
297. Barzilai A, Horowitz A, Geier A, Degani H. Phosphate metabolites and steroid hormone receptors of benign and malignant breast tumors. A Nuclear Magnetic Resonance study. Cancer. 1991;67:2919–25.
298. Wijnen JP, van der Kemp WJM, Luttje MP, et al. Quantitative 31P magnetic resonance spectroscopy of the human breast at 7 T. Magn Reson Med. 2012;68:339–48. doi:10.1002/mrm.23249.
299. Ward KM, Aletras AH, Balaban RS. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). J Magn Reson. 2000;143:79–87. doi:10.1006/jmre.1999.1956.
300. Schmitt B, Trattnig S, Schlemmer H-P. CEST-imaging: a new contrast in MR-mammography by means of chemical exchange saturation transfer. Eur J Radiol. 2012;81 Suppl 1:S144–6. doi:10.1016/S0720-048X(12)70060-8.
301. Desmond KL, Moosvi F, Stanisz GJ. Mapping of amide, amine, and aliphatic peaks in the CEST spectra of murine xenografts at 7 T. Magn Reson Med. 2014;71:1841–53. doi:10.1002/mrm.24822.

302. Rivlin M, Horev J, Tsarfaty I, Navon G. Molecular imaging of tumors and metastases using chemical exchange saturation transfer (CEST) MRI. Sci Rep. 2013;3:3045. doi:10.1038/srep03045.
303. Chan KWY, McMahon MT, Kato Y, et al. Natural D-glucose as a biodegradable MRI contrast agent for detecting cancer. Magn Reson Med. 2012;68: 1764–73. doi:10.1002/mrm.24520.
304. Walker-Samuel S, Ramasawmy R, Torrealdea F, et al. In vivo imaging of glucose uptake and metabolism in tumors. Nat Med. 2013;19:1067–72. doi:10.1038/nm.3252.
305. Nasrallah FA, Pagès G, Kuchel PW, et al. Imaging brain deoxyglucose uptake and metabolism by glucoCEST MRI. J Cereb Blood Flow Metab. 2013;33:1270–8. doi:10.1038/jcbfm.2013.79.
306. Ardenkjaer-Larsen JH, Fridlund B, Gram A, et al. Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. Proc Natl Acad Sci U S A. 2003;100:10158–63. doi:10.1073/pnas.1733835100.
307. Brindle KM, Bohndiek SE, Gallagher FA, Kettunen MI. Tumor imaging using hyperpolarized 13C magnetic resonance spectroscopy. Magn Reson Med. 2011;66:505–19. doi:10.1002/mrm.22999.
308. Jóhannesson H, Macholl S, Ardenkjaer-Larsen JH. Dynamic Nuclear Polarization of [1-13C]pyruvic acid at 4.6 tesla. J Magn Reson. 2009;197:167–75. doi:10.1016/j.jmr.2008.12.016.
309. Golman K, Ardenkjaer-Larsen JH, Petersson JS, et al. Molecular imaging with endogenous substances. Proc Natl Acad Sci U S A. 2003;100:10435–9. doi:10.1073/pnas.1733836100.
310. Golman K, Zandt RI, Lerche M, et al. Metabolic imaging by hyperpolarized 13C magnetic resonance imaging for in vivo tumor diagnosis. Cancer Res. 2006;66:10855–60. doi:10.1158/0008-5472.CAN-06-2564.
311. Park I, Larson PEZ, Zierhut ML, et al. Hyperpolarized 13C magnetic resonance metabolic imaging: application to brain tumors. Neuro Oncol. 2010;12:133–44. doi:10.1093/neuonc/nop043.
312. Chen AP, Albers MJ, Cunningham CH, et al. Hyperpolarized C-13 spectroscopic imaging of the TRAMP mouse at 3T-initial experience. Magn Reson Med. 2007;58:1099–106. doi:10.1002/mrm.21256.
313. Kurhanewicz J, Bok R, Nelson SJ, Vigneron DB. Current and potential applications of clinical 13C MR spectroscopy. J Nucl Med. 2008;49:341–4. doi:10.2967/jnumed.107.045112.
314. Albers MJ, Bok R, Chen AP, et al. Hyperpolarized 13C lactate, pyruvate, and alanine: noninvasive biomarkers for prostate cancer detection and grading. Cancer Res. 2008;68:8607–15. doi:10.1158/0008-5472.CAN-08-0749.
315. Zierhut ML, Yen Y-F, Chen AP, et al. Kinetic modeling of hyperpolarized 13C1-pyruvate metabolism in normal rats and TRAMP mice. J Magn Reson. 2010;202:85–92. doi:10.1016/j.jmr.2009.10.003.
316. Keshari KR, Sai V, Wang ZJ, et al. Hyperpolarized [1-13C]dehydroascorbate MR spectroscopy in a murine model of prostate cancer: comparison with 18F-FDG PET. J Nucl Med. 2013;54:922–8. doi:10.2967/jnumed.112.115402.
317. Asghar Butt S, Søgaard LV, Ardenkjaer-Larsen JH, et al. Monitoring mammary tumor progression and effect of tamoxifen treatment in MMTV-PymT using MRI and magnetic resonance spectroscopy with hyperpolarized [1-(13) C]pyruvate. Magn Reson Med. 2014;73(1):51–8. doi:10.1002/mrm.25095.
318. Taillefer R. Clinical applications of 99mTc-sestamibi scintimammography. Semin Nucl Med. 2005;35:100–15. doi:10.1053/j.semnuclmed.2004.11.002.
319. Helbich TH, Becherer A, Trattnig S, et al. Differentiation of benign and malignant breast lesions: MR imaging versus Tc-99m sestamibi scintimammography. Radiology. 1997;202:421–9. doi:10.1148/radiology.202.2.9015068.
320. Scopinaro F, Schillaci O, Ussof W, et al. A three center study on the diagnostic accuracy of 99mTc-MIBI scintimammography. Anticancer Res. 1997;17:1631–4.
321. Arslan N, Oztürk E, Ilgan S, et al. 99Tcm-MIBI scintimammography in the evaluation of breast lesions and axillary involvement: a comparison with mammography and histopathological diagnosis. Nucl Med Commun. 1999;20:317–25.
322. Maffioli L, Agresti R, Chiti A, et al. Prone scintimammography in patients with non-palpable breast lesions. Anticancer Res. 1996;16:1269–73.
323. Brem RF, Rapelyea JA, Zisman G, et al. Occult breast cancer: scintimammography with high-resolution breast-specific gamma camera in women at high risk for breast cancer. Radiology. 2005;237:274–80. doi:10.1148/radiol.2371040758.
324. Brem RF, Shahan C, Rapleyea JA, et al. Detection of occult foci of breast cancer using breast-specific gamma imaging in women with one mammographic or clinically suspicious breast lesion. Acad Radiol. 2010;17:735–43. doi:10.1016/j.acra.2010.01.017.
325. Coover LR, Caravaglia G, Kuhn P. Scintimammography with dedicated breast camera detects and localizes occult carcinoma. J Nucl Med. 2004;45:553–8.
326. Rhodes DJ, O'Connor MK, Phillips SW, et al. Molecular breast imaging: a new technique using technetium Tc 99m scintimammography to detect small tumors of the breast. Mayo Clin Proc. 2005;80:24–30. doi:10.1016/S0025-6196(11)62953-4.
327. Brem RF, Floerke AC, Rapelyea JA, et al. Breast-specific gamma imaging as an adjunct imaging modality for the diagnosis of breast cancer. Radiology. 2008;247:651–7. doi:10.1148/radiol.2473061678.
328. Brem RF, Fishman M, Rapelyea JA. Detection of ductal carcinoma in situ with mammography, breast specific gamma imaging, and magnetic resonance imaging: a comparative study. Acad Radiol. 2007;14:945–50. doi:10.1016/j.acra.2007.04.004.
329. Sun Y, Wei W, Yang H-W, Liu J-L. Clinical usefulness of breast-specific gamma imaging as an adjunct modality to mammography for diagnosis of

breast cancer: a systemic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2013;40:450–63. doi:10.1007/s00259-012-2279-5.
330. Kalles V, Zografos GC, Provatopoulou X, et al. The current status of positron emission mammography in breast cancer diagnosis. Breast Cancer. 2013;20:123–30. doi:10.1007/s12282-012-0433-3.
331. Avril N, Adler LP. F-18 fluorodeoxyglucose-positron emission tomography imaging for primary breast cancer and loco-regional staging. Radiol Clin North Am. 2007;45:645–57. doi:10.1016/j.rcl.2007.05.004, vi.
332. Quon A, Gambhir SS. FDG-PET and beyond: molecular breast cancer imaging. J Clin Oncol. 2005;23:1664–73. doi:10.1200/JCO.2005.11.024.
333. Avril N, Rosé CA, Schelling M, et al. Breast imaging with positron emission tomography and fluorine-18 fluorodeoxyglucose: use and limitations. J Clin Oncol. 2000;18:3495–502.
334. Rosen EL, Eubank WB, Mankoff DA. FDG PET, PET/CT, and breast cancer imaging. Radiographics. 2007;27 Suppl 1:S215–29. doi:10.1148/rg.27si075517.
335. Samson DJ, Flamm CR, Pisano ED, Aronson N. Should FDG PET be used to decide whether a patient with an abnormal mammogram or breast finding at physical examination should undergo biopsy? Acad Radiol. 2002;9:773–83.
336. Magometschnigg HF, Baltzer PA, Fueger B, et al. Diagnostic accuracy of (18)F-FDG PET/CT compared with that of contrast-enhanced MRI of the breast at 3 T. Eur J Nucl Med Mol Imaging. 2015;42:1656–65. doi:10.1007/s00259-015-3099-1.
337. Kumar R, Lal N, Alavi A. 18F-FDG PET in detecting primary breast cancer. J Nucl Med. 2007;48:1751. doi:10.2967/jnumed.107.043265; author reply 1752.
338. Tatsumi M, Cohade C, Mourtzikos KA, et al. Initial experience with FDG-PET/CT in the evaluation of breast cancer. Eur J Nucl Med Mol Imaging. 2006;33:254–62. doi:10.1007/s00259-005-1835-7.
339. Escalona S, Blasco JA, Reza MM, et al. A systematic review of FDG-PET in breast cancer. Med Oncol. 2010;27:114–29. doi:10.1007/s12032-009-9182-3.
340. Groheux D, Espié M, Giacchetti S, Hindié E. Performance of FDG PET/CT in the clinical management of breast cancer. Radiology. 2013;266:388–405. doi:10.1148/radiol.12110853.
341. Smith-Jones PM, Solit D, Afroze F, et al. Early tumor response to Hsp90 therapy using HER2 PET: comparison with 18F-FDG PET. J Nucl Med. 2006;47:793–6.
342. Pinker K, Bickel H, Magometschnigg H, et al. Molecular imaging of breast tumours with PET-MRI – proof of concept. In: ISMRM 2011 (Eds). European Journal of Cancer, Montreal.
343. Fueger BJ, Weber WA, Quon A, et al. Performance of 2-deoxy-2-[F-18]fluoro-D-glucose positron emission tomography and integrated PET/CT in restaged breast cancer patients. Mol Imaging Biol. 2005;7:369–76. doi:10.1007/s11307-005-0013-4.
344. Radan L, Ben-Haim S, Bar-Shalom R, et al. The role of FDG-PET/CT in suspected recurrence of breast cancer. Cancer. 2006;107:2545–51. doi:10.1002/cncr.22292.
345. Kalinyak JE, Berg WA, Schilling K, et al. Breast cancer detection using high-resolution breast PET compared to whole-body PET or PET/CT. Eur J Nucl Med Mol Imaging. 2014;41:260–75. doi:10.1007/s00259-013-2553-1.
346. Moy L, Ponzo F, Noz ME, et al. Improving specificity of breast MRI using prone PET and fused MRI and PET 3D volume datasets. J Nucl Med. 2007;48:528–37.
347. Domingues RC, Carneiro MP, Lopes FCR, et al. Whole-body MRI and FDG PET fused images for evaluation of patients with cancer. AJR Am J Roentgenol. 2009;192:1012–20. doi:10.2214/AJR.08.1498.
348. Cheng J, Lei L, Xu J, et al. 18F-fluoromisonidazole PET/CT: a potential tool for predicting primary endocrine therapy resistance in breast cancer. J Nucl Med. 2013;54:333–40. doi:10.2967/jnumed.112.111963.
349. Blankenberg FG. In vivo detection of apoptosis. J Nucl Med. 2008;49 Suppl 2:81S–95. doi:10.2967/jnumed.107.045898.
350. Whisenant JG, Peterson TE, Fluckiger JU, et al. Reproducibility of static and dynamic (18)F-FDG, (18)F-FLT, and (18)F-FMISO MicroPET studies in a murine model of HER2+ breast cancer. Mol Imaging Biol. 2013;15:87–96. doi:10.1007/s11307-012-0564-0.
351. Sharma R, Tripathi M, Panwar P, et al. 99mTc-methionine scintimammography in the evaluation of breast cancer. Nucl Med Commun. 2009;30:338–42. doi:10.1097/MNM.0b013e32832999dc.
352. Lindholm P, Lapela M, Någren K, et al. Preliminary study of carbon-11 methionine PET in the evaluation of early response to therapy in advanced breast cancer. Nucl Med Commun. 2009;30:30–6.
353. Yang Z, Sun Y, Zhang Y, et al. Can fluorine-18 fluoroestradiol positron emission tomography-computed tomography demonstrate the heterogeneity of breast cancer in vivo? Clin Breast Cancer. 2013;13:359–63. doi:10.1016/j.clbc.2013.02.012.
354. van Kruchten M, de Vries EGE, Brown M, et al. PET imaging of oestrogen receptors in patients with breast cancer. Lancet Oncol. 2013;14:e465–75. doi:10.1016/S1470-2045(13)70292-4.
355. Smith-Jones PM, Solit DB, Akhurst T, et al. Imaging the pharmacodynamics of HER2 degradation in response to Hsp90 inhibitors. Nat Biotechnol. 2004;22:701–6. doi:10.1038/nbt968.
356. Tamura K, Kurihara H, Yonemori K, et al. 64Cu-DOTA-trastuzumab PET imaging in patients with HER2-positive breast cancer. J Nucl Med. 2013;54:1869–75. doi:10.2967/jnumed.112.118612.
357. Mortimer JE, Bading JR, Colcher DM, et al. Functional imaging of human epidermal growth factor receptor 2-positive metastatic breast cancer using (64)Cu-DOTA-trastuzumab PET. J Nucl Med. 2014;55:23–9. doi:10.2967/jnumed.113.122630.
358. Takada Y, Ye X, Simon S. The integrins. Genome Biol. 2007;8:215. doi:10.1186/gb-2007-8-5-215.

359. Bach-Gansmo T, Danielsson R, Saracco A, et al. Integrin receptor imaging of breast cancer: a proof-of-concept study to evaluate 99mTc-NC100692. J Nucl Med. 2006;47:1434–9.
360. Haubner R, Wester HJ, Weber WA, et al. Noninvasive imaging of alpha(v)beta3 integrin expression using 18F-labeled RGD-containing glycopeptide and positron emission tomography. Cancer Res. 2001;61:1781–5.
361. Gaertner FC, Kessler H, Wester H-J, et al. Radiolabelled RGD peptides for imaging and therapy. Eur J Nucl Med Mol Imaging. 2012;39 Suppl 1:S126–38. doi:10.1007/s00259-011-2028-1.
362. Orel SG, Schnall MD. MR imaging of the breast for the detection, diagnosis, and staging of breast cancer. Radiology. 2001;220:13–30. doi:10.1148/radiology.220.1.r01jl3113.
363. Davis PL, McCarty KS. Sensitivity of enhanced MRI for the detection of breast cancer: new, multicentric, residual, and recurrent. Eur Radiol. 1997;7 Suppl 5:289–98.
364. Heywang-Köbrunner SH, Viehweg P, Heinig A, Küchler C. Contrast-enhanced MRI of the breast: accuracy, value, controversies, solutions. Eur J Radiol. 1997;24:94–108.

Use Case IV: Imaging Biomarkers in Thorax and Heart

18

Jean-Paul Vallée and David Carballo

18.1 Introduction

As discussed previously in this book, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" [3]. Cardiovascular biomarkers are usually classified in three main categories according to their potential use:

- Disease assessment
- Risk stratification and prognostic evaluation
- Clinical surrogate or response to treatment

Regarding risk stratification, well-validated cardiovascular disease risk assessment algorithms such as the Framingham Heart Study risk score in the United States [35] or the European risk prediction system SCORE [5] exist. These scoring systems are routinely used in clinical practice to assess the 10-year risk of occurrence of major coronary heart disease (CHD) events like coronary death or myocardial infarction. Individuals, for example, with a 10-year risk superior to 20 %, benefit from aggressive risk-reduction measures [14]. With this in mind, it is the incremental value of cardiovascular biomarkers that needs to be considered. Risk stratification derived uniquely from imaging biomarkers is therefore at present not clinically significant. Cardiovascular biomarkers do however allow in certain cases the reclassification of an individual cardiovascular risk assessment [10].

Cardiac imaging has traditionally provided results mainly as numerical indices. The list of validated or potentially useful cardiovascular imaging biomarkers is long and includes such values as the left ventricular volume and function assessed by magnetic resonance imaging (MRI) recognized as a gold standard and valuable clinical surrogate. The image acquisition and analysis protocols for cardiac function evaluation are now well established [18, 31]. The reproducibility of MRI for cardiac function assessment is high by comparison to other methods [11]. Proof of efficacy and effectiveness has been largely recognized [26], and function assessment by MRI fulfilled all the criteria of an imaging biomarker. It has become part of the clinical routine and provided by almost all cardiac MR exams. Myocardial perfusion assessed by CMR may also be an interesting potential biomarker. Quantitative analysis of myocardial perfusion by CMR is superior to qualitative analysis [24]. Discordance

J.-P. Vallée, MD, PhD (✉)
Cardiovascular Imaging Unit, Radiology Department, Geneva University Hospital, Genève, Switzerland
e-mail: Jean-Paul.Vallee@unige.ch

D. Carballo, MD, MPH
Cardiology Department, Geneva University Hospital, Genève, Switzerland

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_18

does exist however between the best models to use [32, 25, 13] which has thus far not allowed for the establishment of myocardial perfusion by CMR as a robust biomarker.

There are also new potential cardiovascular imaging biomarkers such as the longitudinal and transverse decays (T1 and T2) mapping which may be able to measure myocardial fibrosis or oedema [30]. These new biomarkers are very promising but are at an early stage of validation, mainly at the stage of proof of principle, whilst proof of efficacy and effectiveness are still lacking.

We will now focus our discussion on the coronary artery calcium score (CAC) which has so far been the most studied and validated cardiovascular imaging biomarkers.

18.2 Coronary Calcium

18.2.1 Clinical Question and Biomarker Concept

The current consensus for the prevention of coronary artery disease (CAD) is that it should be tailored to the level of risk of the individual [21]. Therefore, an accurate estimate of CAD risk is of paramount importance in clinical practice. As atherosclerotic plaques are responsible for CAD, an imaging biomarker of the arterial wall could add relevant information for this risk assessment. Direct imaging of the lipid core of a plaque could ideally be used as biomarker as these plaques are the most unstable and prone to rupture and therefore responsible for acute cardiovascular events. However, imaging the lipid core is difficult and remains largely the realm of invasive procedures with techniques such as IVUS or less validated techniques using coronary angio-CT. CAC is an easy and robust parameter which is strongly linked to the anatomy of atherosclerotic plaques. However, calcifications may appear at various stages of plaque development, sometimes arising on unstable plaques, or after plaque rupture as part of the healing process, but other times related to more stable plaques. Therefore, CAC is not useful for the localization of stenotic lesions and neither a gatekeeper of coronary angio-CT [6] nor as a surrogate endpoint but has demonstrated value as a CAD risk factor [14].

18.2.2 Proof of Mechanism

The exact pathophysiology explaining the efficiency of CAC as a CAD biomarker is still debated. Based on angiographic data, and as discussed above, calcifications are sometimes more related to stable angina than acute coronary events [2]. From histopathologic study, CAC is strongly correlated to the total plaque surface area and therefore to the total coronary atherosclerotic disease burden [29]. The co-occurrence of calcified and non-calcified plaque explains the efficacy of CAC to assess CHD event risk [10].

18.2.3 Image Acquisition and Analysis

In 1990, Agatston introduced a coronary calcium scoring method based on a non-enhanced cardiac CT using a tube voltage of 120 kV [1]. As modification of this value in multi-detector CT (MDCT) changes the CT density measured by the Hounsfield unit (HU), it is recommended to maintain a tube voltage of 120 kV and to adapt the tube current to the patient morphology and to the protocol used [7, 22]. Slices through the heart are usually reconstructed by standard filtered back projection with a thickness of 3 mm as iterative reconstruction algorithms may also modify the calcium score [19, 34]. Coronary calcifications are defined as contiguous pixels with an intensity superior to 130 HU and an area ≥ 1 mm^2. The Agatston score is obtained by multiplying the coronary calcification by a weighting factor depending on the Hounsfield unit density of the calcium deposits [1]. The reproducibility of this biomarker has been well documented in phantom [15] and in clinical studies [8, 16, 20].

18.2.4 Proof of Efficacy and Effectiveness

When defining the characteristics of an intervention, a distinction must be made between *efficacy* and *effectiveness*. Efficacy can be defined as the performance of an intervention to produce the expected result under ideal circumstances, whereas effectiveness measures the performance effect under "real-world" clinical settings [33].

Several prospective studies have demonstrated the efficacy of CAC to predict the risk of future cardiovascular events and to reclassify patients' risk stratification [4, 9, 12, 27, 28]. A novel risk score to estimate 10-year CHD risk using CAC and traditional risk factors was recently developed using the MESA study and validated using two external studies [23].

CAC is considered to have a possible role in reclassifying symptomatic patients with intermediate CAD risk into a higher risk yielding a modification of patient management; there appears to be no role in low- or high-risk patients as no modification of management is expected [10]. Recently, it has been shown that a CAC of zero can be used as an aid in decision-making to avoid initiation of statin therapy in high-risk patients [17]. However, the effectiveness of CAC for reclassification remains debated [14], and the need for an additional prospective randomized trial is needed [21].

Use Case 1

A 50-year-old woman, known only for mild dyslipidaemia, presented to her general practitioner with atypical versus non-anginal chest pain. Physical examination, ECG and myocardial injury blood biomarkers were all unremarkable. Her pretest probability of obstructive coronary artery disease was deemed as low, and no further invasive work-up was carried out. In the context of her dyslipidaemia, however, a CAC score was obtained showing a value of zero, enabling initial pharmacologic therapeutic initiation to be withheld (Fig. 18.1).

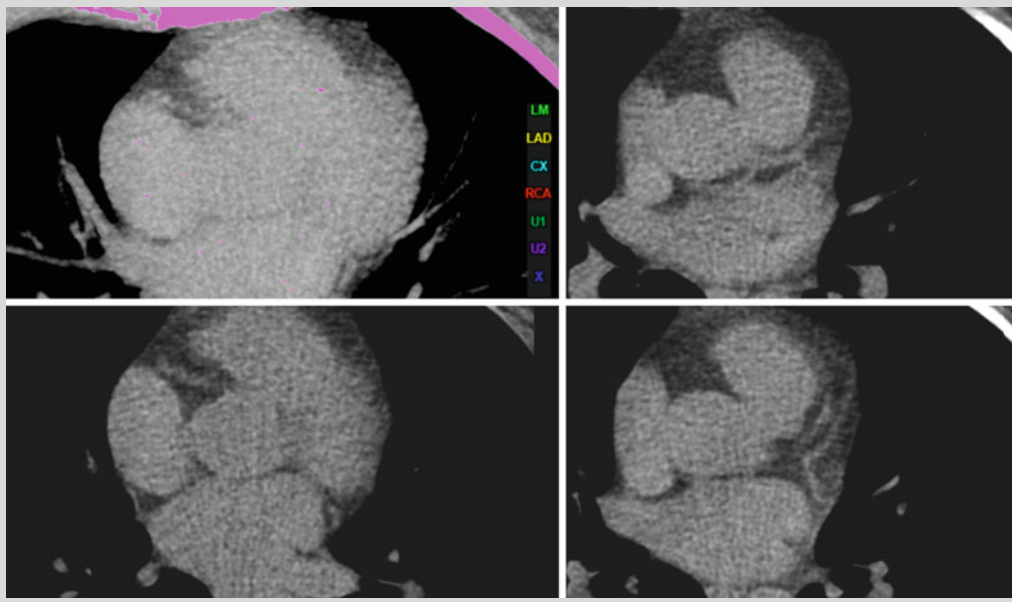


Fig. 18.1 Selective views of an enhanced ECG-gated cardiac CT in a 56-year-old man smoker and familial history with CAC equals zero in a 50-year-old woman with atypical chest pain and hyperlipemia

Use Case 2

A 56-year-old man known with a positive family history and an active smoker status, with a normal stress test, was evaluated as having an intermediate cardiovascular risk. He was addressed for further risk stratification; a CAC of 33 was obtained enabling a reclassification as low risk of 10-year cardiovascular risk (Fig. 18.2).

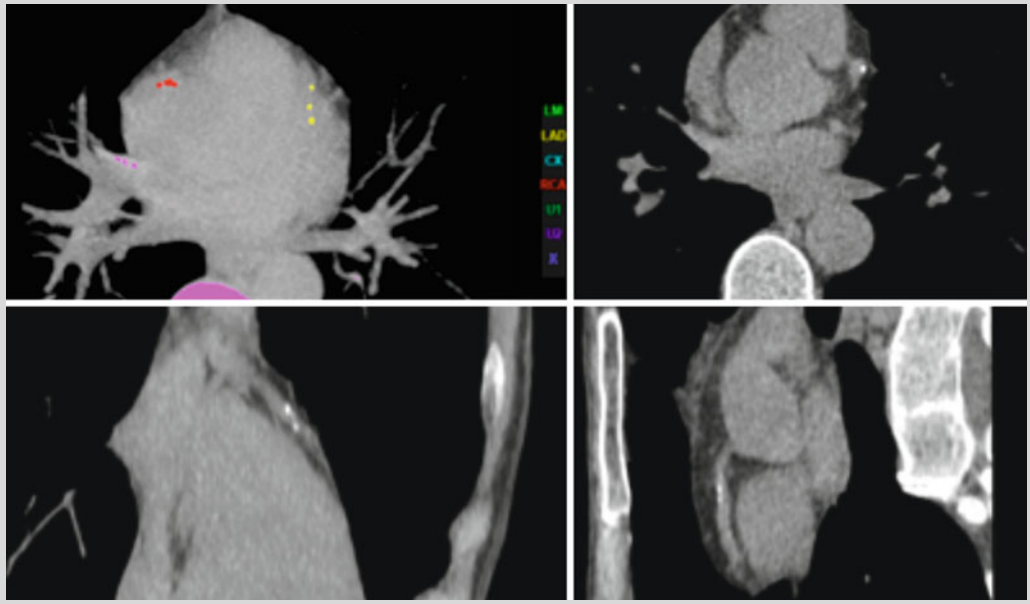


Fig. 18.1 Selective views of an enhanced ECG-gated cardiac CT. In this 56-year-old man smoker and familial history, the CAC was 33. There were few calcifications in the left anterior descending coronary artery and the right coronary artery

Use Case 3

This 69 year old man known for a two vessel coronary artery disease with a history of an infero-lateral transmural myocardial infarction, angioplasty of a circumflex artery chronic occlusion and an angioplasty of the Left anterior descending artery was addressed for reevaluation by coronary angio-CT of a left main coronary artery (LCA) stenosis. A 50% stenotic mixed plaque was found on the LCA as well as multiple diffuse calcified plaques. The CAC was measured at 2676, although in this case, did not modify patient management but illustrates however the strong correlation between the CAC and the severity of coronary disease. (Fig. 18.3).

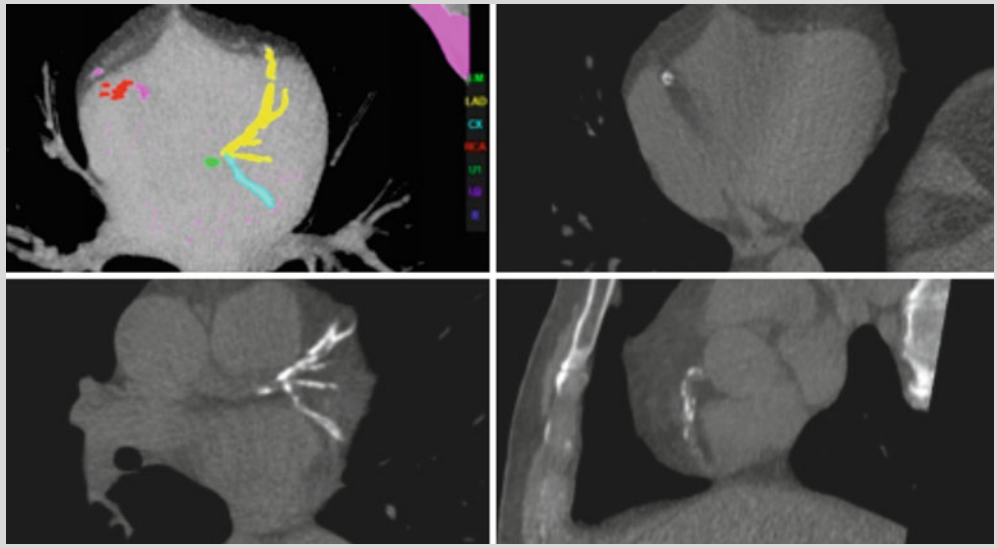


Fig. 18.3 Selective views of an enhanced ECG-gated cardiac CT. In this 56-year-old man smoker and familial history, the CAC was 33. There were few calcifications in the left anterior descending coronary artery and the right coronary artery

Conclusions

CAC is a cardiovascular imaging biomarker derived from the volume and density of coronary artery plaques detected on CT. CAC carries an important potential to reclassify patients' risk stratification especially in the intermediate risk population. The interest of this imaging biomarker, which has been extensively validated through very large clinical cohorts, resides not in its relatively simple calculation method but more on its recognized robustness.

References

1. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol. 1990;15(4):827–32.
2. Beckman JA, Ganz J, Creager MA, et al. Relationship of clinical presentation and calcification of culprit coronary artery stenoses. Arterioscler Thromb Vasc Biol. 2001;21(10):1618–22.
3. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89–95.
4. Budoff MJ, Shaw LJ, Liu ST, et al. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. J Am Coll Cardiol. 2007;49(18):1860–70.
5. Conroy RM, Pyorala K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24(11):987–1003.
6. den Dekker MA, de Smet K, de Bock GH, et al. Diagnostic performance of coronary CT angiography for stenosis detection according to calcium score: systematic review and meta-analysis. Eur Radiol. 2012;22:2688–98.
7. Deprez FC, Vlassenbroek A, Ghaye B, et al. Controversies about effects of low-kilovoltage MDCT acquisition on Agatston calcium scoring. J Cardiovasc Comput Tomogr. 2013;7(1):58–61.
8. Detrano RC, Anderson M, Nelson J, et al. Coronary calcium measurements: effect of CT scanner type and calcium measure on rescan reproducibility – MESA study. Radiology. 2005;236(2):477–84.
9. Durhan G, Hazirolan T, Sunman H, et al. Does coronary calcium scoring with a SCORE better predict significant coronary artery stenosis than without? Correlation with computed tomography coronary angiography. Eur Radiol. 2015;25(3):776–84.
10. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation

Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). Circulation. 2007;115(3):402–26.
11. Grothues F, Smith GC, Moon JC, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol. 2002;90(1):29–34.
12. Han D, Ó Hartaigh B, Gransar H, et al. Incremental Benefit of Coronary Artery Calcium Score Above Traditional Risk Factors for All-Cause Mortality in Asymptomatic Korean Adults. Circ J. 2015;79(11):2445–51.
13. Handayani A, Triadyaksa P, Dijkstra H, et al. Intermodel agreement of myocardial blood flow estimation from stress-rest myocardial perfusion magnetic resonance imaging in patients with coronary artery disease. Invest Radiol. 2015;50(4):275–82.
14. Helfand M, Buckley DI, Freeman M, et al. Emerging risk factors for coronary heart disease: a summary of systematic reviews conducted for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151(7):496–507.
15. Horiguchi J, Kiguchi M, Fujioka C, et al. Variability of repeated coronary artery calcium scoring and radiation Dose on 64- and 16-slice computed tomography by prospective electrocardiographically-triggered axial and retrospective electrocardiographically-gated spiral computed tomography: a phantom study. Acad Radiol. 2008;15(8):958–65.
16. Horiguchi J, Matsuura N, Yamamoto H, et al. Variability of repeated coronary artery calcium measurements by 1.25-mm- and 2.5-mm-thickness images on prospective electrocardiograph-triggered 64-slice CT. Eur Radiol. 2008;18(2):209–16.
17. Joshi PH, Nasir K. Discordance between risk factors and coronary artery calcium: implications for guiding treatment strategies in primary prevention settings. Prog Cardiovasc Dis. 2015;58(1):10–8.
18. Kramer CM, Barkhausen J, Flamm SD, et al. Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update. J Cardiovasc Magn Reson. 2013;15:91.
19. Kurata A, Dharampal A, Dedic A, et al. Impact of iterative reconstruction on CT coronary calcium quantification. Eur Radiol. 2013;23(12):3246–52.
20. Leipsic J, Abbara S, Achenbach S, et al. SCCT guidelines for the interpretation and reporting of coronary CT angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee. J Cardiovasc Comput Tomogr. 2014;8(5):342–58.
21. Lloyd-Jones DM. Coronary artery calcium scoring: Are We there Yet? J Am Coll Cardiol. 2015;66(15):1654–6.
22. Marwan M, Mettin C, Pflederer T, et al. Very low-dose coronary artery calcium scanning with high-pitch spiral acquisition mode: comparison between 120-kV and 100-kV tube voltage protocols. J Cardiovasc Comput Tomogr. 2013;7(1):32–8.
23. McClelland RL, Jorgensen NW, Budoff M, et al. 10-year coronary heart disease risk prediction using coronary artery calcium and traditional risk factors: derivation in the MESA (Multi-Ethnic Study of Atherosclerosis) with validation in the HNR (Heinz Nixdorf Recall) study and the DHS (Dallas Heart Study). J Am Coll Cardiol. 2015;66(15):1643–53.
24. Pan J, Huang S, Lu Z, et al. Comparison of myocardial transmural perfusion gradient by magnetic resonance imaging to fractional flow reserve in patients with suspected coronary artery disease. Am J Cardiol. 2015;115(10):1333–40.
25. Papanastasiou G, Williams MC, Kershaw LE, et al. Measurement of myocardial blood flow by cardiovascular magnetic resonance perfusion: comparison of distributed parameter and Fermi models with single and dual bolus. J Cardiovasc Magn Reson. 2015;17:17.
26. Pitcher A, Ashby D, Elliott P, et al. Cardiovascular MRI in clinical trials: expanded applications through novel surrogate endpoints. Heart. 2011;97(16):1286–92.
27. Polonsky TS, McClelland RL, Jorgensen NW, et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA. 2010;303(16):1610–6.
28. Raggi P, Gongora MC, Gopal A, et al. Coronary artery calcium to predict all-cause mortality in elderly men and women. J Am Coll Cardiol. 2008;52(1):17–23.
29. Rumberger JA, Simons DB, Fitzpatrick LA, et al. Coronary artery calcium area by electron-beam computed tomography and coronary atherosclerotic plaque area. A histopathologic correlative study. Circulation. 1995;92(8):2157–62.
30. Salerno M, Kramer CM. Advances in parametric mapping with CMR imaging. JACC Cardiovasc Imaging. 2013;6(7):806–22.
31. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. J Cardiovasc Magn Reson. 2013;15:35.
32. Schwab F, Ingrisch M, Marcus R, et al. Tracer kinetic modeling in myocardial perfusion quantification using MRI. Magn Reson Med. 2015;73(3):1206–15.
33. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol. 2014;5:e45.
34. van Osch JA, Mouden M, van Dalen JA, et al. Influence of iterative image reconstruction on CT-based calcium score measurements. Int J Cardiovasc Imaging. 2014;30(5):961–7.
35. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837–47.

Use Case V: Imaging Biomarkers in Musculoskeletal Disorders

19

Julio Carballido-Gamio

19.1 Preview

Structural, compositional, and functional changes are common manifestations of musculoskeletal (MSK) disorders. Therefore, with the advent of advanced medical image acquisition and image analysis techniques, there has been a demand and an increased interest toward image-based noninvasive quantification of different pathophysiological elements associated with MSK disorders. Osteoporosis and osteoarthritis of the knee are two MSK disorders with considerable personal and economic burdens. Therefore, these two pathologies have received substantial attention in the medical imaging community, with dual X-ray absorptiometry (DXA) and radiography playing key roles in their clinical diagnosis, respectively. In terms of research, quantitative computed tomography (QCT) has become the leading in vivo imaging modality for the study of osteoporosis enabling the assessment of the most relevant tissues involved in this pathology: bone and muscle [1, 2]. However, high-resolution peripheral quantitative computed tomography (HR-pQCT) is a relatively novel imaging modality with great potential to uncover new clinical findings in the study of osteoporosis [3, 4] and, with the new generation of scanners, also in the study of knee osteoarthritis. Regarding knee osteoarthritis research, magnetic resonance imaging (MRI) is the leading in vivo imaging modality enabling the assessment of the most relevant tissues involved in this pathology: cartilage, bone, muscle, and meniscus [5].

In this chapter, we present some of the imaging biomarkers most studied in MSK disorders. The chapter is divided into four sections. In order to better understand the rationale behind each imaging biomarker, the first section "Relevant MSK Tissues" provides a brief overview of the most relevant MSK tissues currently investigated with medical imaging. While medical imaging has certainly been applied to the study of multiple MSK disorders including those of the spine or those resulting as medical complications of other diseases such as rheumatoid arthritis or diabetes, the second and third sections focus on "osteoporosis" and "knee osteoarthritis," respectively. The reasons behind this decision are space limitations and the fact that many of the imaging biomarkers used in these two pathologies are commonly adapted for the study of other MSK disorders. The chapter concludes with a "summary" where future directions in MSK imaging biomarkers are presented.

QCT imaging biomarkers for the study of osteoporosis that are included in this chapter include those quantifying volumetric bone

J. Carballido-Gamio, PhD
Department of Radiology,
University of Colorado Denver,
Denver, CO, USA
e-mail: Julio.Carballido-Gamio@ucdenver.edu

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_19

mineral density (vBMD), bone geometry, bone strength, and muscle cross-sectional area and composition. Regarding HR-pQCT, this chapter includes imaging biomarkers for the study of osteoporosis that quantify vBMD, trabecular bone microstructure, cortical structure, and bone strength. Magnetic resonance imaging biomarkers for the study of knee osteoarthritis that are covered in this chapter include those describing cartilage and meniscus morphology, cartilage and meniscus composition, trabecular bone microstructure, as well as those investigating muscle composition. For each MSK disorder, imaging modality and imaging biomarker, image processing and analysis techniques currently used for their quantification are also presented. Clinical applications of the imaging biomarkers here presented have been omitted due to space limitations since the number of studies aiming to validate their accuracy, precision, and performance in the assessment of different aspects of osteoporosis and knee osteoarthritis is vast. Readers interested in these aspects are referred to excellent reviews in the literature [1–4, 6–9].

19.2 Relevant MSK Tissues

In this section, we provide a brief overview of the structure and composition of relevant tissues in studies of osteoporosis and knee osteoarthritis with medical imaging. In particular, characteristics of the bone, cartilage, meniscus, and muscle are provided.

19.2.1 Bone

The bone is a living tissue consisting of three main components: (1) a mineral phase of hydroxyapatite; (2) an organic phase of type I collagen, noncollagenous proteins, and lipids; and (3) water [10]. There are three main types of bone tissue: (1) compact or cortical bone, (2) cancellous or trabecular bone, and (3) subchondral bone. Cortical bone is composed of dense bone tissue and forms the outer shell of bones surrounding the marrow space. The diaphysis is primarily composed of cortical bone. The trabecular bone is a honeycomb-like network of trabecular plates and rods in the bone marrow compartment. The metaphysis and epiphysis are rich in trabecular bone. The subchondral bone is the smooth tissue located at the ends of bones that is covered with cartilage. The outer surface of the cortical bone is covered by the periosteum, while its inner surface and the surfaces of the trabeculae are covered by the endosteum. The trabecular bone pores are filled with bone marrow.

19.2.2 Cartilage

There are three bones in the knee joint, the distal part of the femur, the proximal part of the tibia, and the patella, in which articular surfaces are covered by a thick layer of articular cartilage. The articular cartilage covers the articular surfaces of bones in the diarthrodial joints of the human body. The articular cartilage plays key mechanical roles such as low friction movement, shock absorption, and load distribution. The hyaline cartilage, the type of cartilage found in articular surfaces, is composed of cells called chondrocytes, which are surrounded by an extracellular matrix. Approximately 80 % of the extracellular matrix is composed of water with the other major components being collagen and proteoglycans. The articular cartilage exhibits an organized laminar organization composed of the calcified zone, the deep zone, the middle zone, and the superficial zone. The articular cartilage has no blood vessels, nerves, or lymphatics [11].

19.2.3 Meniscus

In the knee joint, stability is provided by ligaments inside and outside the joint capsule; by anterior, posterior, and medial muscles; by lateral dense connective tissue; and by the menisci. The menisci—the crescent-shaped structures located in the medial and lateral aspects of the knee joint sitting on the tibial plateaus—also reduce friction

between articular surfaces, contribute with load distribution, and play a key role on shock absorption. In order to accommodate these functions, the superior surfaces of the menisci are concave providing joint congruity, while its inferior surfaces are flat to fit in the tibial plateaus [12]. The extracellular matrix of menisci and its integrity largely determine the mechanical properties of the tissue. The extracellular matrix is composed of water and a network of collagen fibers (type I and smaller amounts of type II). Proteoglycan molecules, at lower concentrations than those found in the articular cartilage, are entrapped in this network [13].

19.2.4 Muscle

Muscles play central roles in motion and metabolism. Muscle multinucleated cells are called fibers, which incorporate filament-like contractile myosin and actin proteins that are longitudinally organized into sarcomeres, which in turn form myofibrils which are tubelike structures, which number is proportional to the strength of muscle fibers [14].

The quadriceps, composed of the vastus medialis, vastus intermedius, vastus lateralis, and the rectus femoris, is the most relevant anterior muscle group in the knee joint providing extension functions. The amstrings, composed of the biceps femoris, semimembranosus, and semitendinosus, is the most relevant posterior muscle group in the knee joint providing flexion functions. Medially, the pes anserinus muscle group, composed of the sartorius, the gracilis, and the semitendinosus muscles, provides flexion and medial stability functions [15].

There are multiple muscles providing stability and motion to the hip joint with four groups lying within the field of view of QCT scans of the proximal femur: (1) hip extensors (gluteus maximus), (2) hip flexors (sartorius and rectus femoris), (3) hip abductors (gluteus minimus, gluteus medius, and piriformis), and (4) hip adductors (adductor longus, adductor magnus, adductor brevis, quadratus femoris, quadratus externus, and pectineus) [16].

19.3 Osteoporosis

Osteoporosis is a skeletal disorder characterized by low bone mass and structural deterioration of bone tissue, leading to bone fragility and an increased risk of fractures [17]. Osteoporosis is currently diagnosed based on DXA areal bone mineral density (aBMD) using T-scores. T-scores are the result of comparing aBMD values against the mean of a population of healthy 30-year-old adults. A normal T-score is within one standard deviation of the young adult norm. A person with a T-score between 1 and 2.5 standard deviations below the young adult mean is considered osteopenic. A person with a T-score 2.5 standard deviations or less below the young adult mean is diagnosed with osteoporosis, and a person with the same T-score and with osteoporotic fracture (s) is considered as having severe osteoporosis. However, bone strength depends not only on BMD, which according to previous studies explains about 70–75 % of its variance [18]. Bone strength also depends on bone structure, geometry, tissue composition, micro-damage, and biomechanical factors; all of these characteristics are commonly referred as measures of bone quality [19, 20]. Medical imaging allows the quantification of some of these bone quality factors including structure, geometry, and biomechanical parameters.

19.3.1 QCT

QCT has become the leading in vivo imaging modality in osteoporosis research studies. The three-dimensional (3D) capabilities of QCT along with its capability to distinguish between the cortical and trabecular bone enable studies of etiology and drug effects that cannot be performed with DXA, which provides projected measurements that are confounded by bone size [21]. The spine and the proximal femur are the main anatomical sites where QCT has been applied since these represent the most relevant anatomical sites in osteoporosis. The last generation of QCT scanners can provide images with

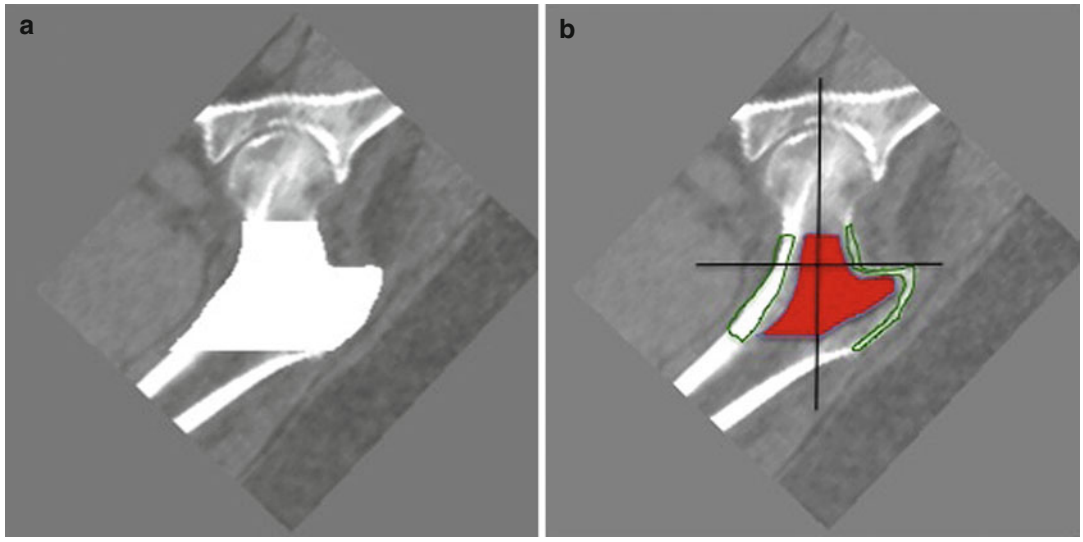


Fig. 19.1 Density analysis of the proximal femur with QCT. (**a**) Coronal cross section where neck and trochanteric voxels have been highlighted in white. (**b**) Cortical (*green contours*) and trabecular (*red voxels*) compartments. The *black* vertical line indicates the femoral neck axis, while the *black* horizontal line divides the femoral neck and trochanteric subvolumes [1]

high-spatial resolution; however, there is a trade-off between high-spatial resolution and radiation dose [22].

19.3.1.1 vBMD

In QCT studies, a calibration phantom with elements of different known concentration values of bone mineral density is usually placed under the scanner table and scanned simultaneously with the patient. The purpose of this phantom is to convert voxel intensities to equivalent values of mass of bone per unit tissue volume, thus enabling vBMD assessments (mg/cm^3). Current clinical QCT acquisitions are performed with an in-plane voxel size ~0.8 and ~1 mm slice thickness. However, the mean trabecular bone thickness in the vertebral bodies and femoral head is ~122 and ~194 μm [23], respectively, thus causing partial volume effects. The limited spatial resolution, and presence of red marrow and marrow fat in the trabecular compartment, means that vBMD measurements are contaminated. The fact that most of the QCT scanners are single-energy devices implies that the different voxel constituents cannot be resolved. However, for clinical applications, the associated errors have not been considered relevant [24]. Figure 19.1 depicts the most common region analyzed in QCT density studies of the proximal femur with its corresponding cortical and trabecular bone compartments.

19.3.1.2 Bone Geometry

Bones break when applied forces overcome their strength. Bone resistance to fracture is dictated by its geometry, distribution of material properties, and magnitude and direction of the applied forces. Measures of bone geometry in the proximal femur include minimum femoral neck cross-sectional area, estimates of cortical bone volume, moments of inertia for strength estimation, and cortical bone thickness (Fig. 19.2) [1].

19.3.1.3 Finite Element Modeling (FEM)

FEM is a numerical engineering technique where the performance of a structure is studied under different loading conditions. In FEM, the complex 3D geometry and distribution of material properties play essential roles in the mathematical model. For this purpose, FEM subdivides the structure into many small elements that are

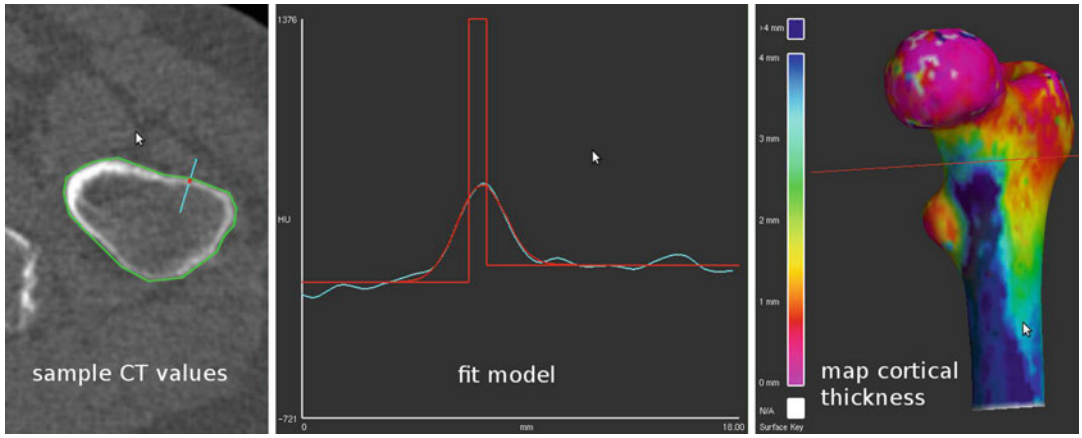


Fig. 19.2 Estimation of cortical bone thickness from QCTs of the proximal femur is performed in four steps: (1) bone segmentation (*green contour*), (2) sampling of CT values along lines perpendicular to the cortical bone (cyan), (3) model fitting, and (4) mapping of estimated thickness values to the femoral surfaces [25]

referred as finite elements. Then material properties are assigned to each finite element and loading conditions are mathematically applied in the form of displacements or forces. The strength of the structure under specific loading conditions is then estimated based on material failure criteria, stress, and strain. Stress refers to an applied force or a system of forces that tends to strain or deform an object, while strain refers to a deformation produced by stress.

Studies have demonstrated that bone strength as measured on cadaveric bones can be best estimated by combining QCT imaging and subject-specific FEM [26–28]. Medical imaging has therefore recently incorporated FEM as a valuable in vivo tool to estimate material strength. Since QCT enables an accurate 3D representation of bone geometry as well as the derivation of material properties based on vBMD values, the use of QCT images of the proximal femur and spine has dramatically increased in recent years to estimate bone strength using FEM [29]. In studies involving the spine, uniform compression loading is the most common clinical loading condition [30], while in studies of the proximal femur, single-limb stance and a fall to the side with an impact on the posterolateral or lateral aspect of the greater trochanter are the most common loading conditions [31]. The most common parameter reported in FEM analyses is failure load (N). Figure 19.3 illustrates a subject-specific FEM example of the proximal femur derived from QCT.

19.3.1.4 Muscle

While the majority of osteoporosis studies have been focused on bone assessment, there has been a recent interest toward surrogate imaging biomarkers of muscle strength. Smaller cross-sectional thigh muscle area—a measure of muscle size—and greater fat infiltration have been associated with increased risk of mobility loss in older men and women, suggesting that the association between low muscle mass and functional decline could be a function of muscle strength [33]. In QCT, inter- and intracellular adipose content of the muscles is primarily measured based on Hounsfield units. Lower Hounsfield units are indicative of higher fat infiltration [16]. Four muscle groups that are within the FOV of QCT acquisitions of the proximal femur are illustrated in Fig. 19.4.

Image Analysis

Conventional vBMD analyses in both the spine and the proximal femur are based on summary statistics. In the spine, the lumbar level is the most studied in osteoporosis research. For most of the QCT analyses, the periosteal contours of the vertebral bodies and the proximal femora

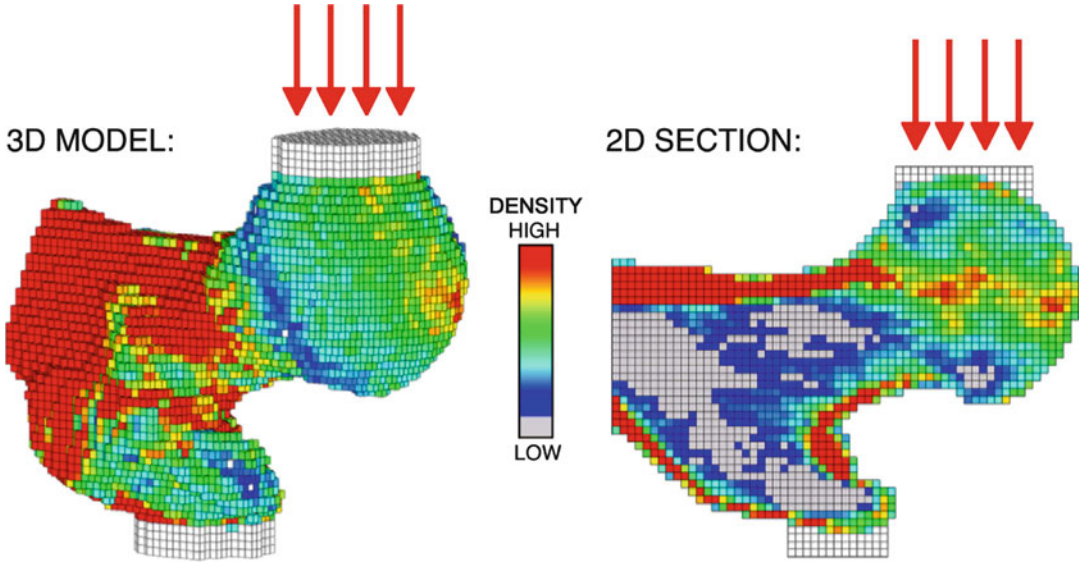


Fig. 19.3 Full 3D and coronal cross section of a FEM model of the proximal femur simulating a sideways fall. Each cube represents a single finite element with isotropic size 1.5 mm^3 [32]

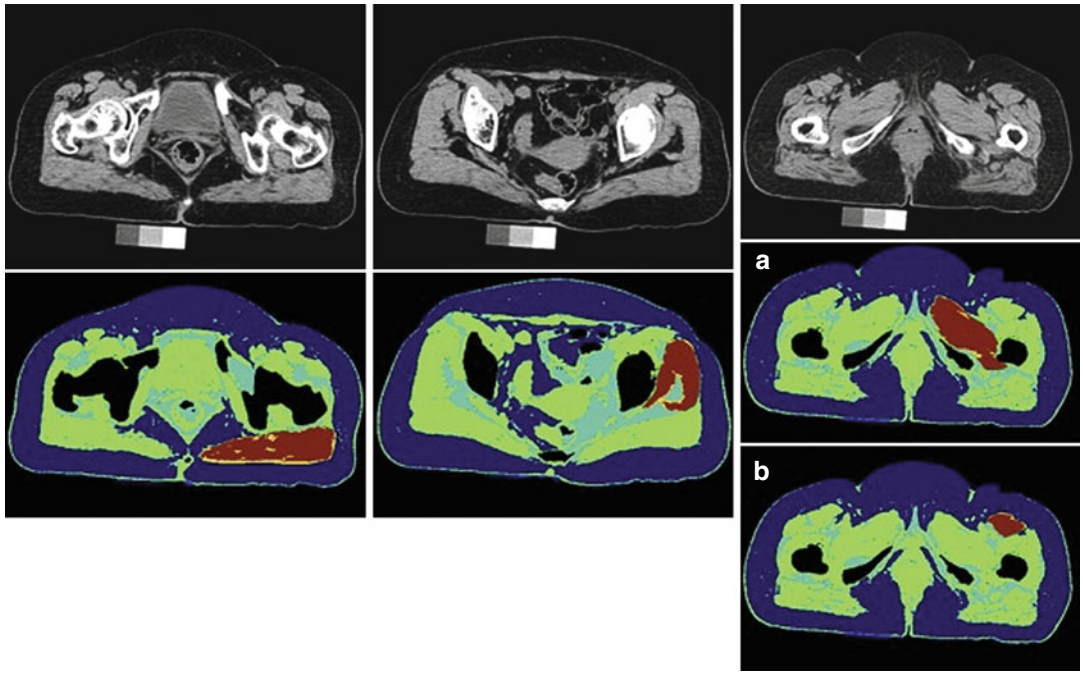


Fig. 19.4 Muscle groups commonly studied with QCTs of the proximal femur are highlighted in *red*. *Left*: hip extensors; *middle*: hip adductors; *right-a*: hip adductors; and *right-b*: hip flexors [1]

must be delineated. This process has been commonly accomplished automatically for the lumbar spine [34] and semi-automatically for the proximal femur [35]. However, recent advances in medical image segmentation algorithms have now enabled the automatic segmentation and multiparametric assessments of proximal femora with QCT [21]. In density analyses, once the outer bone surfaces have been identified, the cortical and trabecular bone compartments are separated and subdivided into subregions from which mean and standard deviation values of cortical and trabecular vBMD are computed. Because the femoral neck is a common fracture site, this anatomic region has been further subdivided into smaller regions with the aim of providing a better understanding of the association of the spatial distribution of vBMD and hip fracture [36, 37].

In conventional vBMD analyses, regions of interest are prescribed based on prior assumptions or hypotheses, and summary statistics discard relevant spatial information. For this reason, new population-based image analyses techniques previously developed for neuroimaging studies have been adapted to musculoskeletal applications. In particular, voxel-based morphometry (VBM) is a technique that enables the identification of bone regions where vBMD is significantly associated with a variable of interest [38, 39]. VBM is a data-driven technique that does not require specification of regions of interest and performs an automatic unbiased assessment of the whole structure. Another technique that makes no assumptions about the spatial association of vBMD with a variable of interest is statistical shape and intensity modeling. This technique is based on principal component analysis to describe each anatomical structure in terms of its associations with principal modes of shape and intensity [40].

Most of the studies have estimated cortical bone thickness in femoral neck cross sections or in the entire femoral neck region using thresholding or the 50 % relative threshold method and repored summary statistics, i.e., mean and standard deviation values. Since fractures usually initiate in the cortical bone, which has been observed to become thinner with aging, the assessment of cortical bone thickness has become quite relevant in recent years. Therefore, new methods have been developed toward not only the accurate [41, 42] and reproducible [21] quantification of cortical bone thickness but also toward the quantification of its spatial distribution across the entire proximal femur [21, 41, 42]. In addition, statistical parametric mapping (SPM) techniques similar to VBM now enable the spatial comparisons of cortical bone thickness and other cortical bone properties at corresponding anatomical locations between populations [21, 25, 41]. SPM thus allows the automatic identification of regions in the bone surface where cortical bone parameters, e.g., thickness or cortical vBMD, are significantly associated with a variable of interest.

Brain image analysis techniques investigating local differences in shape between two populations such as tensor-based morphometry (TBM) have also been adapted to musculoskeletal applications in order to study internal and external structural differences [43].

QCT-based subject-specific FEM is a complex engineering technique, which limits its widespread use even across research institutions. Most of the current FEM approaches are nonparametric and require significant amount of time. For this reason, a parametric FEM approach based on statistical shape and intensity modeling derived from clinical QCT data of proximal femora was recently proposed. From this model a 3D FEM of a new QCT scan can then be derived with high levels of accuracy [44].

Currently, image analyses of muscle size and composition in QCT studies of the proximal femur have not gone beyond summary statistics [16].

19.3.2 HR-pQCT

HR-pQCT is a relatively novel medical imaging modality aimed for the in vivo imaging of bone microstructure in the extremities, particularly the distal radius and distal tibia. Although the manufacturing company producing these devices (XtremeCT; Scanco Medical, Brüttisellen, Switzerland) has recently released the second generation of these scanners, which allow faster

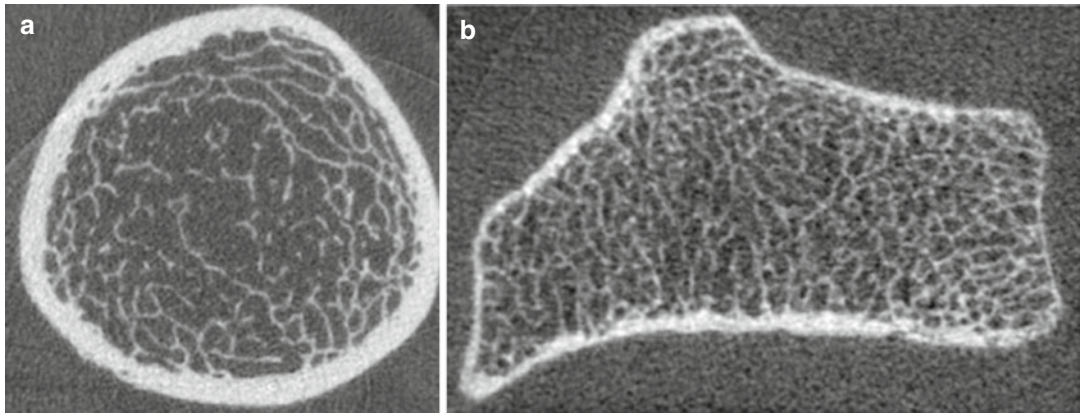


Fig. 19.5 Representative cross sections of HR-pQCT scans of the distal tibia (**a**) and distal radius (**b**) of postmenopausal women [47]

acquisition times, smaller voxel sizes [45] and enable imaging of the knee joint in certain populations, most of the studies in the literature have used first generation scanners and were based on the standard acquisition protocol provided by the manufacturer.

In the standard acquisition protocol, the extremity of a subject is immobilized using a carbon fiber cast that is fixed within the gantry of the scanner. Then an anterior–posterior projection is acquired which is used to plan the 3D acquisition. In the scout view, the operator manually places a reference line at the radial or tibial endplate, and the scanner acquires a slab of 9.02 mm consisting of 110 slices with isotropic voxel sizes of 82 μm. The slab is acquired at 9.5 and 22.5 mm proximal from the reference line and extends proximally for the radius and tibia, respectively. Scans take in the order of 3 min with an effective patient dose per scan in the order of 3 μSv. The standard protocol provided by the manufacturer also incorporates techniques to identify the periosteal and endosteal boundaries, thus enabling compartmental analyses [4].

19.3.2.1 vBMD

Similar to QCT acquisitions, a calibration phantom is scanned simultaneously with the patient in order to convert gray-scale attenuation values to equivalent concentrations of hydroxyapatite (HA). The phantom consists of five cylinders of HA-resin mixtures with the following concentrations: 0, 100, 200, 400, and 800 mgHA/cm^3 [46]. Representative axial cross sections of acquisitions of the distal tibia and distal radius are shown in Fig. 19.5.

19.3.2.2 Trabecular Bone Microstructure

Trabecular bone structure is assessed based on binary images representing the trabecular bone network and with parameters equivalent to those of histomorphometry studies. The binary images are usually obtained with the software provided by the manufacturer using a Laplace–Hamming filter and a fixed threshold, although alternative approaches have also been developed. Common trabecular bone parameters include trabecular bone volume fraction (BV/TV), trabecular bone number (Tb.N; 1/mm), trabecular bone spacing (Tb.Sp; mm), and trabecular bone thickness (Tb.Th; mm). Trabecular bone number (Tb.N) is computed in three steps. In the first step, a mid-axis transformation is applied to extract trabecular bone ridges. In the second step, a distance transform method is used to create a map of distances between ridges. In the final step, the average of the distances over all interridge voxels is calculated and its inverse defined as Tb.N. Bone volume fraction is also computed directly; however, it is based on vBMD and the assumption that fully mineralized bone has a density of 1200 mgHA/cm^3. Tb.Sp and Tb.Th are then derived from BV/TV and Tb.N using plate model assumptions [48].

19.3.2.3 Cortical Bone Structure

The most common cortical bone parameter assessed in HR-pQCT acquisitions is cortical bone thickness, which can be derived or measured directly. The derived measure of cortical bone thickness is based on the ratio of cortical bone volume and the outer bone surface, while the direct assessment is based on the distance transform of the cortical bone region.

A relatively new assessment of cortical bone structure is cortical porosity (Fig. 19.6). The algorithm provided by the manufacturer is based on the segmentation of cortical bone and the identification of cortical voids. Cortical porosity is then defined as the ratio of the volume of the cortical voids, i.e., intracortical pore volume (Ct.Po.V) and the sum of the mineralized and intracortical pore volume. Additional pore parameters include a cortical pore diameter and the distribution of cortical pore diameters [49].

19.3.2.4 Micro-finite Element Modeling (μFEM)

HR-pQCT also enables the estimation of bone strength through micro-finite element analysis. In these simulations, binary images representing the bone structure are converted into meshes of isotropic bone elements using voxel conversion techniques. Although material properties can be applied according to X-ray attenuation values in a similar way as in QCT, bone elements are commonly assigned homogeneous elastic moduli. An advantage of the small voxel sizes is that the cortical and trabecular bone compartments can be labeled as two different materials with identical material properties, thus facilitating compartmental analyses. The most common μFEM configuration for both anatomical sites is a uniaxial compression to 1 % strain. The most common imaging biomarkers of bone strength that are reported are bone stiffness (N/mm) and failure load (N) [50]. Figure 19.7 illustrates the loading conditions as well as representative boundary conditions and distributions of von Mises stresses for the tibia.

Image Analysis

Most of the HR-pQCT studies have been based on summary statistics of cortical and trabecular bone parameters. However, some studies have subdivided the distal radius and distal tibia into subregions with the aim of incorporating some spatial information into the statistics to increase sensitivity in the analyses [52, 53]. Recent

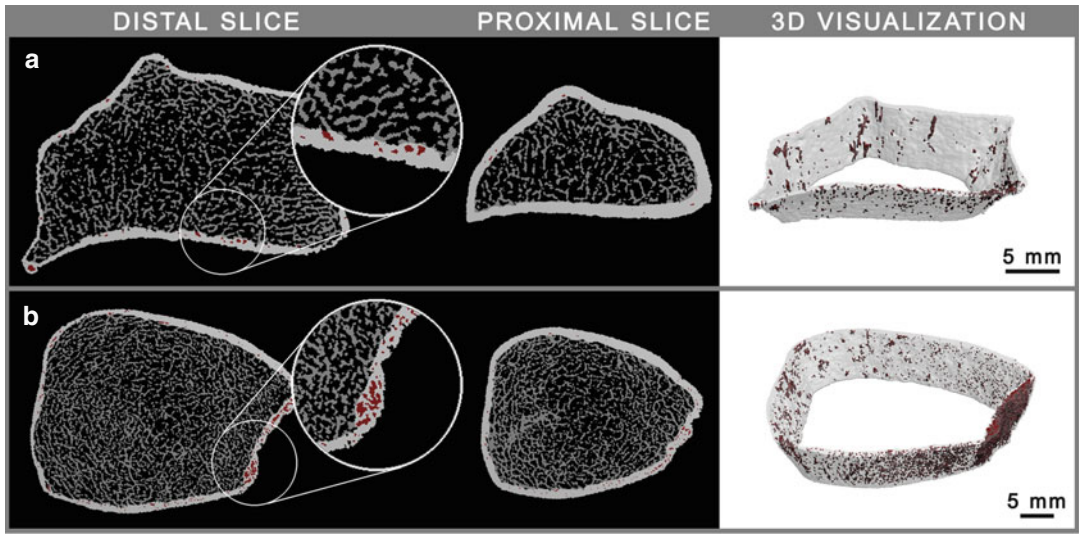


Fig. 19.6 Cross sections of a radius (**a**) and tibia (**b**) where the cortical bone was color-coded in *white*, trabecular bone in *gray*, and intracortical porosity in *red*. On the right side, 3D visualizations of the intracortical porosity (*red*) are shown with a semitransparent cortical bone [49]

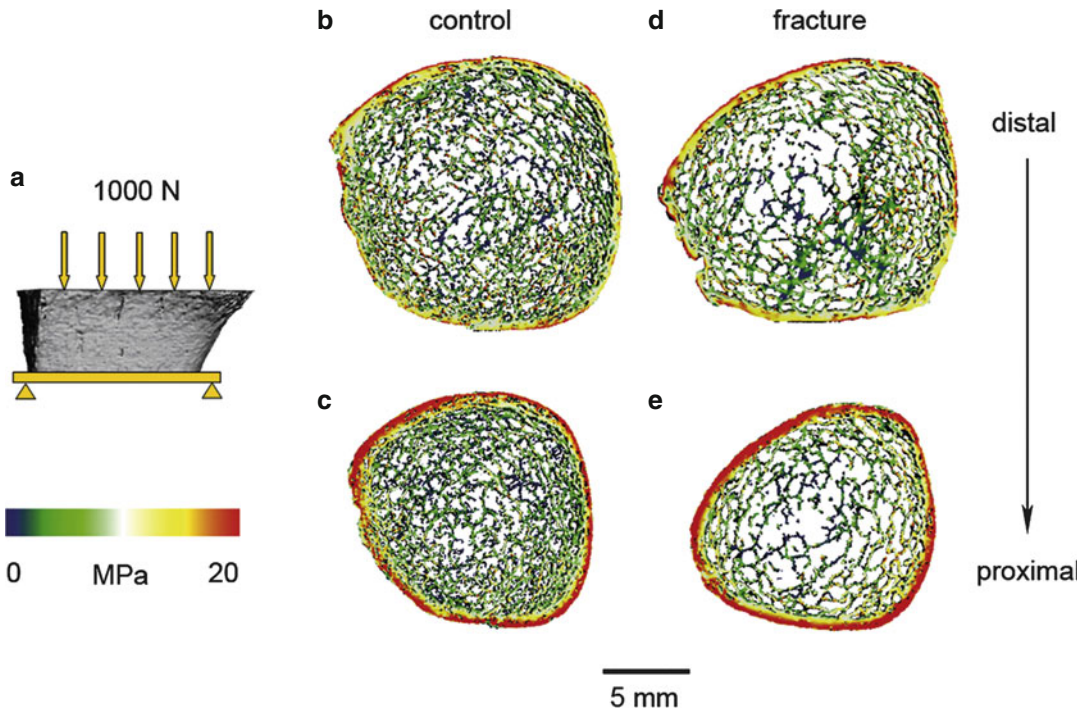


Fig. 19.7 Micro-finite element modeling of the distal tibia. (**a**) Axial compression with a 1000 N load applied to the first distal slice while the proximal surface was fully constrained. The boundary conditions and distributions of von Mises stresses for a control (**b**–**c**), and a fracture case (**d**–**e**) are shown where the level of stress has been color-coded, with red representing high stresses and blue low stresses [51]

algorithms have also been developed to investigate the different cortical porosity phenotypes using 2D laminar analyses of cortical bone [54].

Initial indications of the adaptation of 3D data-driven techniques such as VBM, TBM and SPM to HR-pQCT have been observed in the literature [55]. These techniques will enable population-based comparisons to automatically identify regions where bone parameters are significantly associated with variables of interest.

19.4 Knee Osteoarthritis

Knee osteoarthritis is a condition characterized by pain, functional impairment, and biochemical and structural changes in articular tissue. The increasing elderly population and sports-related injuries have made knee osteoarthritis a leading cause of chronic disability. Despite this, knee osteoarthritis is still poorly understood and few therapeutic options are available. The current surrogate marker for demonstrating structural changes by regulatory agencies is joint space narrowing on weight-bearing radiographs. However, knee osteoarthritis is a disease of the whole joint, and in order to investigate the etiology of knee osteoarthritis, and consequently improve the development of disease modifying osteoarthritis drugs, investigation of all the relevant tissues involved in this disease should be performed.

19.4.1 MRI

MRI is the medical imaging modality of choice to perform studies of knee osteoarthritis [6]. MRI offers in vivo imaging characteristics that cannot be matched with current medical imaging technology such as non-ionizing radiation, tomographic capabilities along practically any orientation, high

spatial resolution, high soft-tissue contrast, and unique capabilities to perform quantitative compositional and functional imaging. Furthermore, MRI is perhaps the medical imaging modality that moves faster technologically speaking, thus constantly offering new imaging and quantification approaches. MRI has therefore been extensively used in studies of knee osteoarthritis. While most of the studies have been focused on the morphology and composition of the knee cartilage, current studies are also looking into other joint tissues such as the bone, meniscus, and muscle, as well as into the hip joint.

19.4.1.1 Cartilage

Cartilage thinning and loss are common manifestations of knee osteoarthritis; therefore there has been substantial interest into the 3D morphological quantification of knee cartilage. Cartilage morphology has been primarily assessed in the form of volume and thickness using high-spatial-resolution images where cartilage signal is bright and bone signal is low [56]. This type of contrast can now be achieved with a variety of pulse sequences. While the majority of previous studies were done in the sagittal plane, current advances in MRI hardware and pulse sequence development allow the acquisition of high-spatial-resolution images with (quasi)-isotropic voxels thus enabling reformations into other planes without compromising image characteristics [57].

Unfortunately, morphological changes of cartilage are small and slow. In addition, recent studies have suggested that cartilage undergoes compositional changes prior to morphological events [5]. MRI has then been used to study cartilage composition based on the quantification of T_2 (Fig. 19.8), $T_{1\rho}$, and T_{1Gd} relaxation times [7], which have been associated with different cartilage properties, therefore emerging as potential in vivo noninvasive imaging biomarkers for early detection of knee osteoarthritis. T_2 relaxation times have been associated with tissue hydration and biochemical composition, specifically the integrity of the collagen matrix. $T_{1\rho}$ relaxation times describe the spin–lattice relaxation in the rotating frame and have been associated with changes in proteoglycan loss—a manifestation of early knee osteoarthritis. The rationale is that pulse sequences for $T_{1\rho}$ quantification were designed with the idea of providing biochemical information in the low-frequency regime (few hundred Hz to few kHz), thus interrogating the slow interactions in the extracellular matrix between motion-restricted water molecules, proteoglycan, and collagen. T_{1Gd} refers to the T_1 relaxation time of the tissue after intravenous injection of Gd-DTPA^{2-}, which is a charged contrast agent that is absorbed in articular cartilage. This technique is known as delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) and has been used to investigate the distribution and changes in glycosaminoglycan

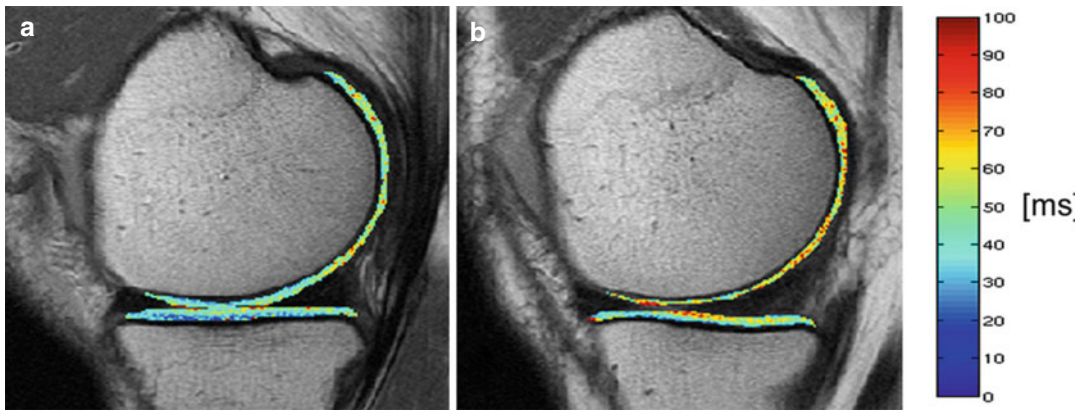


Fig. 19.8 Representative T_2 relaxation time maps of the medial femur and medial tibia of a subject without knee pain (**a**) and a subject with knee pain (**b**) [60]

concentrations that occur early in knee osteoarthritis [58]. In order to calculate relaxation times, multiple MRI acquisitions are performed at different echo times (T_2), spin-lock times ($T_{1\rho}$), and inversion delay times (T_{1Gd}). Then voxel-wise exponential fittings are performed according to the relaxation time in question, thus yielding maps depicting the spatial distribution of T_2, $T_{1\rho}$, or T_{1Gd}. Other approaches to quantify cartilage composition and structure include gagCEST, sodium MRI, and diffusion tensor imaging [9, 59].

19.4.1.2 Meniscus

The relevance of the menisci in the biomechanics of the knee joint is clearly acknowledged. Therefore, it has been suggested that meniscal damage may be associated with onset and progression of knee osteoarthritis [61–63]. Because cartilage has been the dominant tissue investigated in knee osteoarthritis using MRI, studies focused on meniscal integrity took longer to materialize. Initial MRI assessments of meniscal integrity were morphometric in nature characterizing meniscal volume, regional thickness, and position [64]. However, the meniscal fibrocartilage is primarily composed of type I collagen and a small percentage of proteoglycans embedded in a dense collagen matrix. These characteristics have consequently stimulated studies of meniscal compositional integrity with MR relaxation times including T_2 (Fig. 19.9), $T_{1\rho}$ [65], and T_{1Gd} [66].

19.4.1.3 Trabecular Bone

In terms of bone quantification, MRI has been primarily used for trabecular bone assessment [68]. This is, however, a challenging task due to a trade-off between signal-to-noise ratio (SNR) and spatial resolution. The small dimensions of trabecular bone push the limits of clinical MRI scanners in order to obtain high-spatial-resolution images with high SNR at clinically feasible scan times.

Trabecular bone assessment with MRI has been primarily done in the distal radius, distal

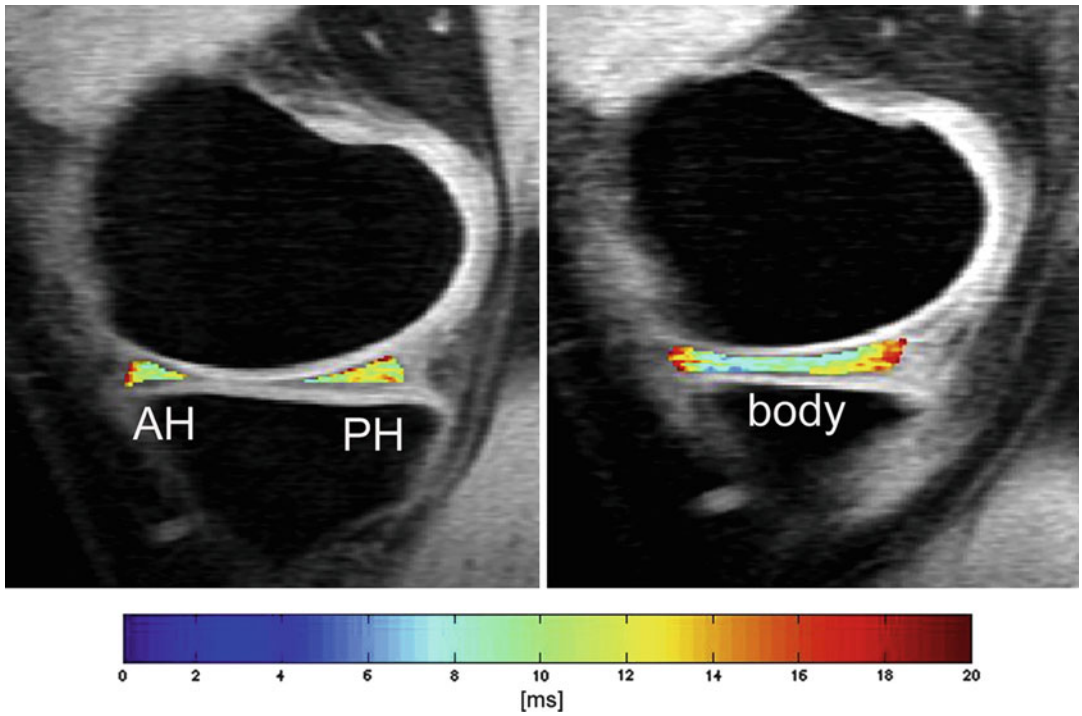


Fig. 19.9 Representative meniscal T_2 relaxation time maps: (**a**) anterior horn (AH) and posterior horn (PH); (**b**) body of medial meniscus [67]

tibia, and calcaneus. The main reasons were not only due to their anatomical relevance in osteoporosis but also due to the trade-off mentioned above. These anatomical sites do not impose the same level of SNR limitations as deep body locations like the proximal femur, where radio frequency signals are attenuated by surrounding tissues such as fat and muscle. However, recent advances in MRI hardware, coil design, pulse sequences, and image analysis techniques have enabled the acquisition and analyses of high-spatial-resolution images depicting the trabecular bone microstructure of the proximal femur (Fig. 19.10) [69, 70], which is a site of utmost importance in studies of osteoporosis.

Since clinical MRI is based on mobile protons and the bone is solid, then the bone exhibits very short T_1 and T_2 relaxation times thus yielding no signal. On the other hand, mobile protons in the fatty and water content of bone marrow yield a relatively strong signal. The result is therefore an "inverse" image depicting the trabecular bone micro-architecture. Several parameters have been developed with the aim of quantifying the trabecular bone micro-architecture using MRI. These parameters have been primarily divided into three classes according to the main aspect that they quantify: (1) scale, (2) topology, and (3) orientation also known as anisotropy.

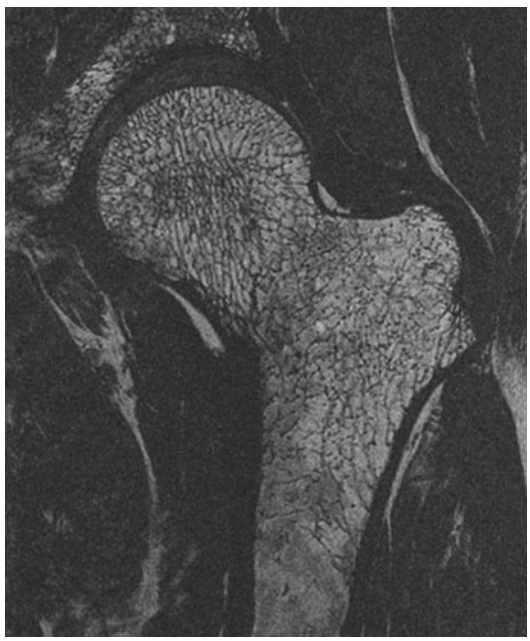


Fig. 19.10 Coronal cross section of an in vivo high-spatial-resolution MR image of the proximal femur with a spatial resolution of $234 \times 234 \times 500\ \mu m^3$ [71]

19.4.1.4 Muscle

Similar to research involving menisci, muscle was largely ignored giving preference to studies of cartilage. However, it has been recently acknowledged that knee osteoarthritis is a pathology that goes beyond cartilage; it is a pathology that involves all the different tissues of the joint. For this reason, investigators have started to investigate the associations of muscle characteristics with knee osteoarthritis. In particular, investigators have looked into muscle cross-sectional area as well as into surrogate measures of fat infiltration [72], since fat infiltration has been observed in multiple chronic conditions such as diabetes and sarcopenia due to aging. For this purpose, investigators have primarily used T_1-weighted images. However, T_1-weighted images can only provide qualitative measures of fat content based on pure signal contrast. On the other hand, more advanced imaging techniques such as proton magnetic resonance spectroscopic imaging (1H MRSI) and chemical shift-based water-fat separation approaches can provide accurate measures of fat fraction content. 1H MRSI provides detailed spectral information at low spatial resolutions, while chemical shift-based water-fat separation approaches provide high-spatial-resolution fat fraction maps enabling voxel-wise analyses [73–75]. Figure 19.11 illustrates a representative fat infiltration muscle analysis in the thigh using chemical shift-based water-fat separation.

Image Analysis

Initial studies of knee osteoarthritis were primarily based on summary statistics of femoral, tibial, and patellar morphological and compositional cartilage parameters. The femur was then subdivided into medial and lateral compartments as well as into a trochlear region and the tibia into medial and lateral compartments. Increased interest into the spatial assessment of imaging

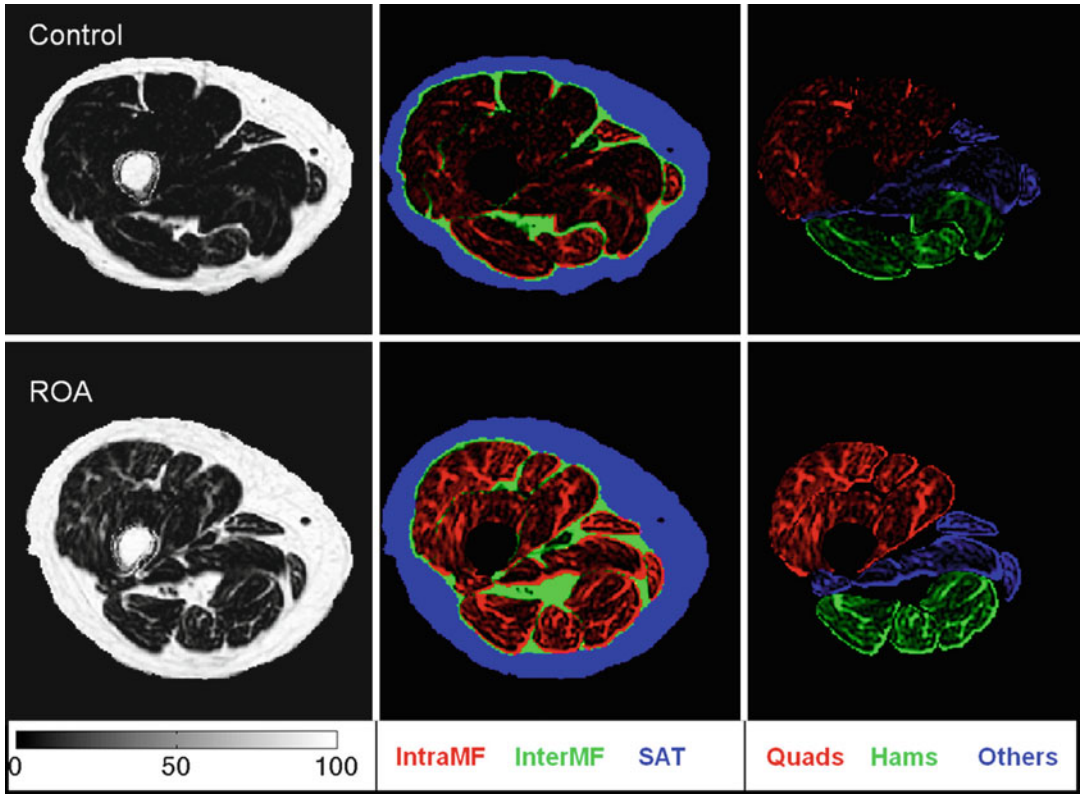


Fig. 19.11 Fat distribution and fat content of a control subject and a subject with radiographic knee osteoarthritis using chemical shift-based water-fat separation. Subjects are of the same age, gender, and similar total muscle lean anatomical cross-sectional area. First column: fat fraction maps (color bar units are %). Second column: color-coded, fat fraction maps highlighting the intramuscular fat fraction (IntraMF; *red*), intermuscular fat fraction (InterMF; *green*), and subcutaneous adipose tissue (SAT; *blue*) regions. Third column: color-coded fat fraction maps highlighting the quadriceps (Quads; *red*), hamstrings (Hams; *green*), and other muscle (Others; *blue*) regions. The color-coded maps were weighted by the fat fraction map values [75]

biomarkers motivated further subdivisions based on weight-bearing regions. Similarly, meniscus analyses went from global to regional assessments of morphology and composition. However, the feasibility of performing local assessments of cartilage properties was recently demonstrated as well as the adaptation of VBM-like approaches to relaxation time values of the knee cartilage [76, 77].

The majority of the clinical MRI scanners acquire images of subjects lying in a supine position. This position therefore provides images under non-weight-bearing conditions. However, assessment of joint structures in weight-bearing conditions could provide further insight into tissue health. For this reason, research groups have also investigated the morphology and composition of the knee cartilage [78, 79] and menisci [13, 80] under weight-bearing conditions where specialized MRI-compatible devices were built in order to simulate weight-bearing with different amounts of loading.

Trabecular bone microstructural scale parameters are based on the principles of stereology, and they are commonly computed using the mean intercept length (MIL) approach. There are four scale parameters that have been used in multiple studies: (1) apparent trabecular bone volume fraction (app.BV/TV), (2) apparent trabecular bone number (app.Tb.N; 1/mm), (3) apparent trabecular bone spacing (app.Tb.Sp; mm), and (4) apparent trabecular bone thickness (app.Tb.Th;

mm) [81]. They are called apparent due to the fact that the spatial resolution is not enough to completely discern the individual trabeculae. Early work on trabecular bone topological analyses decomposed the trabecular bone into a network of connected curves and surfaces thus enabling the quantification of different properties of this skeleton, e.g., (1) surface-to-curve ratio; (2) number of curve–curve junctions; (3) number of surface–surface junctions; and (4) number of surface–curve junctions [82, 83]. The rationale behind these analyses is that the trabecular bone network can be assumed to be primarily composed of trabecular bone plates and rods and that fenestration of trabecular plates and eventual loss of connectivity is a common pathophysiological mechanism in osteoporosis. A new topological analysis that quantifies the full 3D trabecular bone network and not a skeletonized version of it was recently developed [84]. Regarding trabecular bone anisotropy, MIL [85], topological-based orientation analyses of trabecular bone networks [83] and the 3D spatial autocorrelation function [86] have been used to quantify bone orientation. Techniques encompassing measures of scale, topology and anisotropy have also been developed such as the geodesic topological analysis (GTA) approach [87]. Since junctions play a central role in trabecular bone connectivity, GTA quantifies the trabecular bone network in terms of its junctions and geodesic distances, which is the shortest path between two points.

The analysis of trabecular bone using MRI is however not straightforward. Before scale, topology, and anisotropy can be quantified, there are a series of image processing steps that have to be performed. In order to obtain high SNR, especially in the deep-seated locations like the proximal femur, surface coils are commonly used. As a consequence, images show nonuniform intensity across the field of view with tissue structures located closer to the coil being brighter than those that are far away. Therefore, specialized image processing techniques like the nonparametric nonuniform intensity normalization (N3) [88] have to be applied to remove the bias field prior to image analysis [89]. In addition, regions of interest have to be prescribed within the trabecular bone compartments. While this step might look straightforward, previous studies have shown variations in trabecular bone parameters within the same anatomic structures [90]. Therefore, atlas-based approaches have been developed to ensure the consistent placement of corresponding anatomic regions of interest in either longitudinal or cross-sectional studies. In this approaches, standard regions of interest are prescribed in a template (also known as atlas) and mapped to the rest of the scans in a study using image registration incorporating affine and nonlinear transformations [70]. Once the intensity has been normalized and regions of interest have been defined, the trabecular bone network has to be segmented, i.e., distinguished from the marrow spaces, so it can be quantified. Different segmentation algorithms have been proposed in the literature with initial approaches using basic histogram-based thresholding techniques [81]. However, current methods aim to provide soft segmentations incorporating partial volume effects [91]. Once the trabecular bone has been segmented then it can be analyzed using the different techniques mentioned above.

Similar to QCT studies of muscle composition, analyses of muscle composition in MRI have not gone beyond summary statistics such as mean and standard deviation values.

19.5 Summary

In this chapter, we have presented some of the imaging biomarkers most investigated in osteoporosis and knee osteoarthritis using QCT, HR-pQCT, and MRI. In order to understand the rationale behind each parameter, a brief overview of the most relevant tissues involved in these pathologies was also provided. It is important to emphasize that medical imaging is currently used to investigate several MSK disorders beyond osteoporosis and knee osteoarthritis and that the imaging biomarkers presented here only represent a small subset of a number of parameters that are constantly increasing. We have also presented a brief overview of different image analysis approaches which aim is to go beyond

summary statistics to extract relevant clinical information. The current and future trend in the study of MSK disorders with medical imaging is to perform multiparametric and multimodality analyses with the inclusion of other disciplines such as machine learning and genetics with the long-term goal of personalized medicine.

References

1. Lang TF. Quantitative computed tomography. Radiol Clin North Am. 2010;48(3):589–600.
2. Engelke K, Libanati C, Fuerst T, Zysset P, Genant HK. Advanced CT based in vivo methods for the assessment of bone density, structure, and strength. Curr Osteoporos Rep. 2013;11(3):246–55.
3. Cheung AM, Adachi JD, Hanley DA, Kendler DL, Davison KS, Josse R, Brown JP, Ste-Marie LG, Kremer R, Erlandson MC, Dian L, Burghardt AJ, Boyd SK. High-resolution peripheral quantitative computed tomography for the assessment of bone strength and structure: a review by the Canadian Bone Strength Working Group. Curr Osteoporos Rep. 2013;11(2):136–46. PMCID: 3641288.
4. Nishiyama KK, Shane E. Clinical imaging of bone microarchitecture with HR-pQCT. Curr Osteoporos Rep. 2013;11(2):147–55. PMCID: 4102136.
5. Roemer FW, Eckstein F, Hayashi D, Guermazi A. The role of imaging in osteoarthritis. Best Pract Res Clin Rheumatol. 2014;28(1):31–60.
6. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed. 2006;19(7):822–54.
7. Burstein D, Gray M, Mosher T, Dardzinski B. Measures of molecular composition and structure in osteoarthritis. Radiol Clin North Am. 2009;47(4):675–86.
8. Carballido-Gamio J, Nicolella DP. Computational anatomy in the study of bone structure. Curr Osteoporos Rep. 2013;11(3):237–45.
9. Guermazi A, Alizai H, Crema MD, Trattnig S, Regatte RR, Roemer FW. Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. Osteoarthritis Cartilage OARS Osteoarthritis Res Soc. 2015;23(10):1639–53.
10. Boskey AL. Bone composition: relationship to bone fragility and antiosteoporotic drug effects. Bonekey Rep. 2013;2:447. PMCID: 3909232.
11. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. Sports Health. 2009;1(6):461–8. PMCID: 3445147.
12. Fox AJ, Bedi A, Rodeo SA. The basic science of human knee menisci: structure, composition, and function. Sports Health. 2012;4(4):340–51. PMCID: 3435920.
13. Subburaj K, Souza RB, Wyman BT, Le Graverand-Gastineau MP, Li X, Link TM, Majumdar S. Changes in MR relaxation times of the meniscus with acute loading: an in vivo pilot study in knee osteoarthritis. J Magn Reson Imaging JMRI. 2015;41(2):536–43. PMCID: 4062611.
14. Brooks SV. Current topics for teaching skeletal muscle physiology. Adv Physiol Educ. 2003;27(1–4):171–82.
15. Souza RB, Doan R. Anatomy and physiology of the knee. In: Majumdar S, editor. Advances in MRI of the knee for. Osteoarthritis: World Scientific; 2010.
16. Lang T, Koyama A, Li C, Li J, Lu Y, Saeed I, Gazze E, Keyak J, Harris T, Cheng X. Pelvic body composition measurements by quantitative computed tomography: association with recent hip fracture. Bone. 2008;42(4):798–805.
17. NIH Consensus Development Panel on Osteoporosis Prevention Diagnosis and Therapy. Osteoporosis prevention, diagnosis, and therapy. JAMA. 2001;285(6):785–95.
18. Ammann P, Rizzoli R. Bone strength and its determinants. Osteoporos Int J Established Result Cooperation Between Eur Found Osteoporos Nat Osteoporos Found USA. 2003;14 Suppl 3:S13–8.
19. Seeman E, Delmas PD. Bone quality--the material and structural basis of bone strength and fragility. N Engl J Med. 2006;354(21):2250–61.
20. Chapurlat RD, Delmas PD. Bone microdamage: a clinical perspective. Osteoporos Int J Established Result Cooperation Between Eur Found Osteoporos Nat Osteoporos Found USA. 2009;20(8):1299–308.
21. Carballido-Gamio J, Bonaretti S, Saeed I, Harnish R, Recker R, Burghardt AJ, Keyak JH, Harris T, Khosla S, Lang T. Automatic multi-parametric quantification of the proximal femur with quantitative computed tomography. Quant Imaging Med Surg. 2015;5(4):552–68.
22. Carballido-Gamio J, Majumdar S. Clinical utility of microarchitecture measurements of trabecular bone. Curr Osteoporos Rep. 2006;4(2):64–70.
23. Hildebrand T, Laib A, Muller R, Dequeker J, Ruegsegger P. Direct three-dimensional morphometric analysis of human cancellous bone: microstructural data from spine, femur, iliac crest, and calcaneus. J Bone Miner Res Off J Am Soc Bone Miner Res. 1999;14(7):1167–74.
24. Lang TF. Essentials of bone biology: assessment of bone architecture. In: Orwoll ES, Bilezikian JP, Vanderschueren D, editors. Osteoporosis in men: the effects of gender on skeletal health: Academic Press; San Diego, CA, U.S.A., 1999.
25. Treece GM, Poole KE, Gee AH. Imaging the femoral cortex: thickness, density and mass from clinical CT. Med Image Anal. 2012;16(5):952–65. PMCID: 3417239.
26. Keyak JH, Rossi SA, Jones KA, Skinner HB. Prediction of femoral fracture load using automated finite element modeling. J Biomech. 1998;31(2):125–33.
27. Cody DD, Gross GJ, Hou FJ, Spencer HJ, Goldstein SA, Fyhrie DP. Femoral strength is better predicted

by finite element models than QCT and DXA. J Biomech. 1999;32(10):1013–20.
28. Bessho M, Ohnishi I, Matsuyama J, Matsumoto T, Imai K, Nakamura K. Prediction of strength and strain of the proximal femur by a CT-based finite element method. J Biomech. 2007;40(8):1745–53.
29. Keaveny TM. Biomechanical computed tomography-noninvasive bone strength analysis using clinical computed tomography scans. Ann N Y Acad Sci. 2010;1192:57–65.
30. Kopperdahl DL, Aspelund T, Hoffmann PF, Sigurdsson S, Siggeirsdottir K, Harris TB, Gudnason V, Keaveny TM. Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. J Bone Miner Res Off J Am Soc Bone Miner Res. 2014;29(3):570–80. PMCID: 3925753.
31. Keyak JH, Sigurdsson S, Karlsdottir G, Oskarsdottir D, Sigmarsdottir A, Zhao S, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. Male–female differences in the association between incident hip fracture and proximal femoral strength: a finite element analysis study. Bone. 2011;48(6):1239–45.
32. Keaveny TM, McClung MR, Wan X, Kopperdahl DL, Mitlak BH, Krohn K. Femoral strength in osteoporotic women treated with teriparatide or alendronate. Bone. 2012;50(1):165–70.
33. Visser M, Goodpaster BH, Kritchevsky SB, Newman AB, Nevitt M, Rubin SM, Simonsick EM, Harris TB, Study HA. Muscle mass, muscle strength, and muscle fat infiltration as predictors of incident mobility limitations in well-functioning older persons. J Gerontol A Biol Sci Med Sci. 2005;60(3):324–33.
34. Engelke K, Mastmeyer A, Bousson V, Fuerst T, Laredo JD, Kalender WA. Reanalysis precision of 3D quantitative computed tomography (QCT) of the spine. Bone. 2009;44(4):566–72.
35. Li W, Sode M, Saeed I, Lang T. Automated registration of hip and spine for longitudinal QCT studies: integration with 3D densitometric and structural analysis. Bone. 2006;38(2):273–9. PMCID: 1550383.
36. Johannesdottir F, Poole KE, Reeve J, Siggeirsdottir K, Aspelund T, Mogensen B, Jonsson BY, Sigurdsson S, Harris TB, Gudnason VG, Sigurdsson G. Distribution of cortical bone in the femoral neck and hip fracture: a prospective case–control analysis of 143 incident hip fractures; the AGES-REYKJAVIK Study. Bone. 2011;48(6):1268–76. PMCID: 3129599.
37. Johannesdottir F, Turmezei T, Poole KE. Cortical bone assessed with clinical computed tomography at the proximal femur. J Bone Miner Res Off J Am Soc Bone Miner Res. 2014;29(4):771–83.
38. Li W, Kezele I, Collins DL, Zijdenbos A, Keyak J, Kornak J, Koyama A, Saeed I, Leblanc A, Harris T, Lu Y, Lang T. Voxel-based modeling and quantification of the proximal femur using inter-subject registration of quantitative CT images. Bone. 2007;41(5):888–95. PMCID: 2080679.
39. Carballido-Gamio J, Harnish R, Saeed I, Streeper T, Sigurdsson S, Amin S, Atkinson EJ, Therneau TM, Siggeirsdottir K, Cheng X, Melton 3rd LJ, Keyak J, Gudnason V, Khosla S, Harris TB, Lang TF. Proximal femoral density distribution and structure in relation to age and hip fracture risk in women. J Bone Miner Res Off J Am Soc Bone Miner Res. 2013;28(3):537–46. PMCID: 3578081.
40. Bredbenner TL, Mason RL, Havill LM, Orwoll ES, Nicolella DP, The Osteoporotic Fractures in Men S. Fracture risk predictions based on statistical shape and density modeling of the proximal femur. J Bone Miner Res Off J Am Soc Bone Miner Res. 2014;29:2090–100.
41. Treece GM, Gee AH, Mayhew PM, Poole KE. High resolution cortical bone thickness measurement from clinical CT data. Med Image Anal. 2010;14(3):276–90. PMCID: 2868358.
42. Treece GM, Gee AH. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. Med Image Anal. 2015;20(1):249–64.
43. Carballido-Gamio J, Harnish R, Saeed I, Streeper T, Sigurdsson S, Amin S, Atkinson EJ, Therneau TM, Siggeirsdottir K, Cheng X, Melton 3rd LJ, Keyak JH, Gudnason V, Khosla S, Harris TB, Lang TF. Structural patterns of the proximal femur in relation to age and hip fracture risk in women. Bone. 2013;57(1):290–9. PMCID: 3809121.
44. Nicolella DP, Bredbenner TL. Development of a parametric finite element model of the proximal femur using statistical shape and density modelling. Comput Methods Biomech Biomed Engin. 2012;15(2):101–10.
45. Manske SL, Zhu Y, Sandino C, Boyd SK. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. Bone. 2015;79:213–21.
46. Sekhon K, Kazakia GJ, Burghardt AJ, Hermannsson B, Majumdar S. Accuracy of volumetric bone mineral density measurement in high-resolution peripheral quantitative computed tomography. Bone. 2009;45(3):473–9. PMCID: 4454742.
47. Link TM. The Founder's Lecture 2009: advances in imaging of osteoporosis and osteoarthritis. Skeletal Radiol. 2010;39(10):943–55. PMCID: 2920421.
48. Kazakia GJ, Hyun B, Burghardt AJ, Krug R, Newitt DC, de Papp AE, Link TM, Majumdar S. In vivo determination of bone structure in postmenopausal women: a comparison of HR-pQCT and high-field MR imaging. J Bone Miner Res. 2008;23(4):463–74.
49. Burghardt AJ, Buie HR, Laib A, Majumdar S, Boyd SK. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. Bone. 2010;47(3):519–28. PMCID: 2926164.
50. van Rietbergen B, Ito K. A survey of micro-finite element analysis for clinical assessment of bone strength: the first decade. J Biomech. 2015;48(5):832–41.
51. Vilayphiou N, Boutroy S, Sornay-Rendu E, Van Rietbergen B, Munoz F, Delmas PD, Chapurlat R. Finite element analysis performed on radius

and tibia HR-pQCT images and fragility fractures at all sites in postmenopausal women. Bone. 2010;46(4):1030–7.
52. Sode M, Burghardt AJ, Kazakia GJ, Link TM, Majumdar S. Regional variations of gender-specific and age-related differences in trabecular bone structure of the distal radius and tibia. Bone. 2010;46(6):1652–60. PMCID: 2875319.
53. Kazakia GJ, Nirody JA, Bernstein G, Sode M, Burghardt AJ, Majumdar S. Age- and gender-related differences in cortical geometry and microstructure: improved sensitivity by regional analysis. Bone. 2013;52(2):623–31. PMCID: 3564644.
54. Nirody JA, Cheng KP, Parrish RM, Burghardt AJ, Majumdar S, Link TM, Kazakia GJ. Spatial distribution of intracortical porosity varies across age and sex. Bone. 2015;75:88–95.
55. Carballido-Gamio J, Bonaretti S, Kazakia GJ, Khosla S, Lang TF, Burghardt AJ. Population-based local multi-parametric comparisons of HR-pQCT studies. J Orthop Translat. 2014;4(2):234–5.
56. Carballido-Gamio J, Eckstein F. Quantitative morphological imaging of the knee joint. In: Majumdar S, editor. Advances in MRI of the knee for Osteoarthritis. Hackensack: World Scientific; 2010.
57. Souza RB, Kumar D, Calixto N, Singh J, Schooler J, Subburaj K, Li X, Link TM, Majumdar S. Response of knee cartilage T1rho and T2 relaxation times to in vivo mechanical loading in individuals with and without knee osteoarthritis. Osteoarthritis Cartilage OARS Osteoarthritis Res Soc. 2014;22(10):1367–76. PMCID: 4184934.
58. Blumenkrantz G, Li X, Regatte RR, Jerschow A, Majumdar S. Functional imaging of the knee joint. In: Majumdar S, editor. Advances in MRI of the knee for osteoarthritis. Hackensack: World Scientific; 2010.
59. Matzat SJ, van Tiel J, Gold GE, Oei EH. Quantitative MRI techniques of cartilage composition. Quant Imaging Med Surg. 2013;3(3):162–74. PMCID: 3701096.
60. Baum T, Joseph GB, Arulanandan A, Nardo L, Virayavanich W, Carballido-Gamio J, Nevitt MC, Lynch J, McCulloch CE, Link TM. Association of magnetic resonance imaging-based knee cartilage T2 measurements and focal knee lesions with knee pain: data from the Osteoarthritis Initiative. Arthritis Care Res. 2012;64(2):248–55. PMCID: 3267009.
61. Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, Einhorn TA, Felson DT. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. J Bone Joint Surg Am. 2003;85-A(1):4–9.
62. Englund M, Niu J, Guermazi A, Roemer FW, Hunter DJ, Lynch JA, Lewis CE, Torner J, Nevitt MC, Zhang YQ, Felson DT. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis Rheum. 2007;56(12):4048–54.
63. Hunter DJ, Zhang YQ, Niu JB, Tu X, Amin S, Clancy M, Guermazi A, Grigorian M, Gale D, Felson DT. The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. Arthritis Rheum. 2006;54(3):795–801.
64. Wirth W, Frobell RB, Souza RB, Li XJ, Wyman BT, Le Grayerand MPH, Link TM, Majumdar S, Eckstein F. A three-dimensional quantitative method to measure meniscus shape, position, and signal intensity using MR images: a pilot study and preliminary results in knee osteoarthritis. Magn Reson Med. 2010;63(5):1162–71.
65. Calixto NE, Kumar D, Subburaj K, Singh J, Schooler J, Nardo L, Li X, Souza RB, Link TM, Majumdar S. Zonal differences in meniscus MR relaxation times in response to in vivo static loading in knee osteoarthritis. J Orthop Res Off Publ Orthop Res Soc. 2015;34:249–61.
66. Krishnan N, Shetty SK, Williams A, Mikulis B, McKenzie C, Burstein D. Delayed gadolinium-enhanced magnetic resonance imaging of the meniscus: an index of meniscal tissue degeneration? Arthritis Rheum. 2007;56(5):1507–11.
67. Baum T, Joseph GB, Karampinos DC, Jungmann PM, Link TM, Bauer JS. Cartilage and meniscal T2 relaxation time as non-invasive biomarker for knee osteoarthritis and cartilage repair procedures. Osteoarthritis Cartilage. 2013;21(10):1474–84.
68. Wehrli FW. Structural and functional assessment of trabecular and cortical bone by micro magnetic resonance imaging. J Magn Reson Imag JMRI. 2007;25(2):390–409.
69. Krug R, Banerjee S, Han ET, Newitt DC, Link TM, Majumdar S. Feasibility of in vivo structural analysis of high-resolution magnetic resonance images of the proximal femur. Osteoporos Int J Established Result Cooperation Between Eur Found Osteoporos Nat Osteoporos Found USA. 2005;16(11):1307–14.
70. Carballido-Gamio J, Folkesson J, Karampinos DC, Baum T, Link TM, Majumdar S, Krug R. Generation of an atlas of the proximal femur and its application to trabecular bone analysis. Magn Reson Med Off J Soc Magn Reson Med/Soc Magn Reson Med. 2011;66(4):1181–91. PMCID: 3596104.
71. Krug R, Burghardt AJ, Majumdar S, Link TM. High-resolution imaging techniques for the assessment of osteoporosis. Radiol Clin North Am. 2010;48(3):601–21. PMCID: 2901255.
72. Raynauld JP, Pelletier JP, Roubille C, Dorais M, Abram F, Li W, Wang Y, Fairley J, Cicuttini FM, Martel-Pelletier J. Magnetic resonance imaging-assessed vastus medialis muscle Fat content and risk for knee osteoarthritis progression: relevance from a clinical trial. Arthritis Care Res. 2015;67(10):1406–15.
73. Karampinos DC, Baum T, Nardo L, Alizai H, Yu HZ, Carballido-Gamio J, Yap P, Shimakawa A, Link TM, Majumdar S. Characterization of the regional distribution of skeletal muscle adipose tissue in type 2 diabetes using chemical shift-based water/fat separation. J Magn Reson Imaging. 2012;35(4):899–907.
74. Fischer MA, Nanz D, Shimakawa A, Schirmer T, Guggenberger R, Chhabra A, Carrino JA, Andreisek G. Quantification of muscle fat in patients with low

back pain: comparison of multi-echo MR imaging with single-voxel MR spectroscopy. Radiology. 2013;266(2):555–63.
75. Kumar D, Karampinos DC, MacLeod TD, Lin W, Nardo L, Li X, Link TM, Majumdar S, Souza RB. Quadriceps intramuscular fat fraction rather than muscle size is associated with knee osteoarthritis. Osteoarthritis Cartilage. 2014;22(2):226–34.
76. Carballido-Gamio J, Majumdar S. Atlas-based knee cartilage assessment. Magn Reson Med Off J Soc Magn Reson Med/Soc Magn Reson Med. 2011;66(2):574–83. PMCID: 3346276.
77. Pedoia V, Li X, Su F, Calixto N, Majumdar S. Fully automatic analysis of the knee articular cartilage T relaxation time using voxel-based relaxometry. J Magn Reson Imaging JMRI. 2015;43:970–80.
78. Souza RB, Stehling C, Wyman BT, Hellio Le Graverand MP, Li X, Link TM, Majumdar S. The effects of acute loading on T1rho and T2 relaxation times of tibiofemoral articular cartilage. Osteoarthritis Cartilage OARS Osteoarthritis Res Soc. 2010;18(12):1557–63.
79. Subburaj K, Souza RB, Stehling C, Wyman BT, Le Graverand-Gastineau MP, Link TM, Li X, Majumdar S. Association of MR relaxation and cartilage deformation in knee osteoarthritis. J Orthop Res Off Publ Orthop Res Soc. 2012;30(6):919–26. PMCID: 3652010.
80. Stehling C, Souza RB, Le Graverandc MPH, Wyman BT, Li XJ, Majumdara S, Link TM. Loading of the knee during 3.0 T MRI is associated with significantly increased medial meniscus extrusion in mild and moderate osteoarthritis. Eur J Radiol. 2012;81(8):1839–45.
81. Majumdar S, Genant HK, Grampp S, Newitt DC, Truong VH, Lin JC, Mathur A. Correlation of trabecular bone structure with age, bone mineral density, and osteoporotic status: In vivo studies in the distal radius using high resolution magnetic resonance imaging. J Bone Miner Res. 1997;12(1):111–8.
82. Saha PK, Chaudhuri BB. 3D digital topology under binary transformation with applications. Comput Vis Image Und. 1996;63(3):418–29.
83. Gomberg BR, Saha PK, Song HK, Hwang SN, Wehrli FW. Topological analysis of trabecular bone MR images. IEEE Trans Med Imaging. 2000;19(3):166–74.
84. Saha PK, Xu Y, Duan H, Heiner A, Liang GY. Volumetric topological analysis: a novel approach for trabecular bone classification on the continuum between plates and rods. IEEE Trans Med Imaging. 2010;29(11):1821–38.
85. Whitehouse WJ. The quantitative morphology of anisotropic trabecular bone. J Microsc. 1974;101(Pt 2):153–68.
86. Rotter M, Berg A, Langenberger H, Grampp S, Imhof H, Moser E. Autocorrelation analysis of bone structure. J Magn Reson Imaging JMRI. 2001;14(1):87–93.
87. Carballido-Gamio J, Krug R, Huber MB, Hyun B, Eckstein F, Majumdar S, Link TM. Geodesic topological analysis of trabecular bone microarchitecture from high-spatial resolution magnetic resonance images. Magn Reson Med Off J Soc Magn Reson Med/Soc Magn Reson Med. 2009;61(2):448–56.
88. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998;17(1):87–97.
89. Folkesson J, Krug R, Goldenstein J, Issever AS, Fang C, Link TM, Majumdar S. Evaluation of correction methods for coil-induced intensity inhomogeneities and their influence on trabecular bone structure parameters from MR images. Med Phys. 2009;36(4):1267–74.
90. Newitt DC, van Rietbergen B, Majumdar S. Processing and analysis of in vivo high-resolution MR images of trabecular bone for longitudinal studies: reproducibility of structural measures and micro-finite element analysis derived mechanical properties. Osteoporos Int J Established Result Cooperation Between Eur Found Osteoporos Nat Osteoporos Found USA. 2002;13(4):278–87.
91. Folkesson J, Carballido-Gamio J, Eckstein F, Link TM, Majumdar S. Local bone enhancement fuzzy clustering for segmentation of MR trabecular bone images. Med Phys. 2010;37(1):295–302. PMCID: 2804381.

Use Case VI: Imaging Biomarkers in Diffuse Liver Disease. Quantification of Fat and Iron

20

Manuela França, Ángel Alberich-Bayarri, and Luis Martí-Bonmatí

20.1 Introduction

Diffuse liver diseases result from a wide spectrum of etiologies, and, since they can lead to cirrhosis, end-stage liver disease, and hepatocellular carcinoma, they represent an important cause of morbidity, mortality, and healthcare costs worldwide. Regardless of the etiology, liver fat, iron, and combined overload are common pathological features of different diffuse liver diseases. Fat and iron frequently coexist and they act as cofactors in disease progression. Because liver biopsy has several limitations, quantitative imaging biomarkers have been developed for liver fat and iron quantification.

In this chapter, we will focus on MR imaging techniques that provide quantitative imaging biomarkers of fat and iron overload.

M. França (✉)
Imaging Department, Centro Hospitalar do Porto, Porto, Portugal
e-mail: mariamanuela.franca@gmail.com

Á. Alberich-Bayarri
Medical Imaging Department and Imaging Research Group (GIBI230), Hospital Universitario y Politécnico La Fe and Research Institute; QUIBIM SL. Valencia, Spain
e-mail: angel@quibim.com

L. Martí-Bonmatí
Medical Imaging Department and Imaging Research Group (GIBI230), Hospital Universitario y Politécnico La Fe and Research Institute; QUIBIM SL. Hospital Quiron Valencia, Valencia, Spain
e-mail: marti_lui@gva.es

20.1.1 The Proof of Concept

Liver fat accumulation is the histological hallmark of nonalcoholic fatty liver disease (NAFLD), which is the most prevalent chronic hepatic disease in the Western countries [1]. About one third of patients with NAFLD may progress to nonalcoholic steatohepatitis (NASH) [2], with increased risk of cirrhosis and hepatocellular carcinoma. Nevertheless, steatosis itself is not a diagnosis but, instead, a common feature to different diffuse liver disorders, such as NAFLD, alcoholic liver disease, viral hepatitis, drug toxicity, and even hemochromatosis [3]. Systemically, liver steatosis is considered a determinant of insulin resistance and is associated with metabolic syndrome, being an independent risk factor for cardiovascular mortality [3] [1].

Hepatic iron overload is found in genetic hemochromatosis and transfusional hemosiderosis though it may also occur in chronic hepatopathies. In chronic liver diseases, hepatic iron can act as a comorbid factor in the progression of liver fibrosis and cirrhosis and in the development of hepatocarcinoma [4].

Both fat and iron liver deposits impose an oxidative stress on hepatocytes, interfering with each other in a cross-dependent fashion [5, 6] and acting as cofactors in disease progression. For example, in patients with hemochromatosis, coexisting steatosis acts as a cofactor in the development of liver fibrosis and cirrhosis [7].

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_20

On the other hand, in patients with NAFLD and NASH, iron overload appears to be related with disease severity and development of liver fibrosis [4, 6]. Furthermore, both steatosis and iron overload seem to reduce the response to antiviral interferon therapy in patients with HCV infection [8].

Liver core biopsy has been routinely considered the gold standard for detection and quantification of fat and iron deposits in the liver parenchyma. At histopathologic analysis, the presence of steatosis is usually graded on a standard visual scale, from 0 (none) to 3 (severe) grades, based on the proportion of hepatocytes containing intracellular vacuoles of fat (≤5 %, 6–33 %, 34–66 %, ≥67 %) [9]. Histologic assessment of liver iron overload is based on a visual scoring system graded from 0 to 4, after Perls' Prussian blue staining of iron granules [10]. Furthermore, biochemical quantification of liver iron concentration (LIC) can be determined from biopsy samples (fresh or paraffin embedded) by atomic absorption spectrophotometry. In primary and secondary hemochromatosis, iron overload is defined as a LIC superior to 36 μmol Fe/g dry weight. However, patients with chronic liver diseases usually have an iron burden below this threshold [4].

Liver iron quantification is important not only for diagnosis of iron overload. Biochemically determined LIC has been used as the surrogate of total body iron burden, and, therefore, LIC measurements are important to select iron-overloaded patients who will benefit from iron-reducing treatments and to accurately monitor these therapies.

Although comprehensive, liver core biopsy has several limitations. It is expensive and invasive, having a potential for adverse complications. Histologic evaluation is based on estimations of percentages of hepatocytes with fat or iron deposits (visual scores), and thus it is operator dependent and subject to high interobserver and intraobserver variability. Nevertheless, the greatest limitation of liver biopsy is its high sampling variability that results from the spatial heterogeneous distribution of disease throughout the liver parenchyma and the small sample size [11]. Therefore, there is a need for noninvasive biomarkers able to detect, quantify, and monitor steatosis and iron deposits in the whole liver, obviating the need for a liver biopsy.

20.1.2 The Proof of Mechanism

The ideal biomarkers for liver fat and iron should demonstrate a strong correlation with the accepted reference standard (liver biopsy), having a high intra- and inter-examination repeatability, and should be robust. Besides providing clinically important thresholds to diagnose liver steatosis or iron overload, fat and iron imaging biomarkers should be able to measure slight changes in those hepatic deposits during patients' follow-up, allowing for precise treatment monitoring [12].

Ideally, noninvasive imaging biomarkers would be considered as hepatic virtual biopsies, and liver core biopsy would be reserved for cases in which the histologic diagnosis of underlying liver disease would be needed.

Among imaging techniques, MR imaging has a unique ability to detect fat and iron. Intracellular fat and iron can be depicted due to the characteristic changes in the tissue magnetic properties, which are related to the amount of intracellular deposits. MR quantification of liver fat and iron, using dedicated MR sequences and post-processing algorithms, provides numerical and continuous measurements, rather than categorical or semiquantitative grades as does liver biopsy.

In the last decade, fat and iron MR imaging biomarkers have been developed and are being increasingly translated into clinical practice.

20.2 Imaging Biomarkers of Liver Fat: MR Sequences, Measurements, and Biases

20.2.1 MR Spectroscopy

Single-voxel 1H-MR spectroscopy (MRS) depicts proton signals as a function of their resonance frequency. Fat and water concentrations are quantified directly from spectral signal,

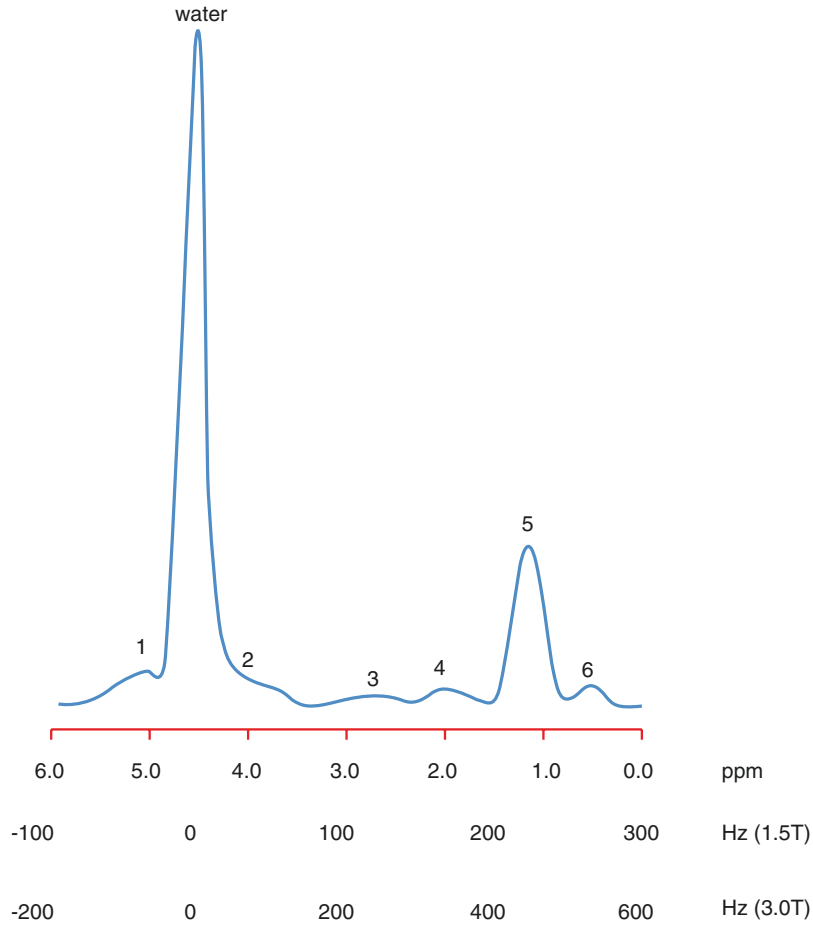


Fig. 20.1 Illustrative example of a synthetic spectrum signal showing the water (highest peak, on the left) and the multiple fat peaks. On a 3 Tesla MR, the main fat peak is located at a frequency shift of 420 Hz (1.46 ppm) relative to water peak. Fat proton density results from the sum of multiple fat peaks of the diverse chemical moieties of triglycerides. These multiple fat peaks should be included in the signal fitting algorithm

based on the prior knowledge of their resonance frequencies [13] (Fig. 20.1). Proton density fat fraction (PDFF) can be calculated as a ratio:

$$\mathrm{PDFF} = \mathrm{PD}_{\mathrm{fat}} / \left(\mathrm{PD}_{\mathrm{fat}} + \mathrm{PD}_{\mathrm{water}}\right)$$

where $\mathrm{PD}_{\mathrm{fat}}$ is the proton density of all the triglyceride peaks and the $\mathrm{PD}_{\mathrm{water}}$ is the proton density of water protons. Tissue spectra are usually recorded from volumes ranging from 1 to 27 cm^3 of the liver parenchyma. Although MRS presents an excellent correlation with hepatic lipid content, being accurate, reproducible, and sensitive to small variations in liver fat deposits [13], unfortunately, MRS has a high cost, is time-consuming, and needs dedicated analysis tools, being only available in specialized hospital centers. Moreover, it suffers from similar limitations as liver biopsy regarding the limited sample size of analysis [12].

20.2.2 Chemical-Shift-Based MR Sequences

For several years, *dual-echo chemical-shift (ChSh) gradient-echo (GRE) sequences* have been widely used in clinical practice for the visual assessment of liver steatosis. For a given magnetic field strength, fat protons precess

slower than water protons. Images are acquired with a first echo, when water and fat peaks are "out of phase" (OP), and then with a second echo, when the two peaks are "in phase" (IP). The echo time (TE) corresponding to the IP and OP images is dependent on the field strength: fat and water peaks are "in phase" every 4.6 ms (1.5 T MR) or 2.3 ms (3 T MR) and are "out of phase" at 2.3 ms (1.5 T MR) or 1.15 ms (3 T MR) and subsequent multiples [13]. On IP images, the liver signal intensity results from the water plus fat signals, while on OP images, the liver signal results from the difference between the water and fat signals. Therefore, steatosis can be recognized as a liver signal dropout on the OP image in comparison to the IP image (Fig. 20.2). Theoretically, the liver fat content could be estimated as:

$$\text{Fat fraction} = \left(S_{IP} - S_{OP}\right) / 2S_{IP},$$

where S_{IP} corresponds to liver/spleen signal intensity measured on the IP image and S_{OP} represents the relative signal measured on the OP image [14]. Unfortunately, this easy approach cannot be used for precise fat quantification, because this estimated fat fraction is based on signal intensities of water and fat, which are subject to several biases, rather than on proton densities. For example, in patients with iron overload, due to the iron T2*-shortening effects, there is a decrease in liver signal intensity on second-echo IP images relative to the first-echo OP images (Fig. 20.2). Therefore, when fat and iron coexist, the decrease of liver signal in the second-echo (IP) image due to increased iron-related T2* decay results in less or no apparent liver signal change between the first- (OP) and second-echo (IP) images [15]. Consequently, liver fat content will be underestimated by the coexistence of iron, if the T2* decay is not taken into account. Since these chemical-shift sequences are acquired with at least two different echoes, some signal T2* decay occurs between different echoes, even in normal livers (Fig. 20.2).

Besides the T2* decay effect, the "signal" fat fraction is also influenced by other factors, such as the T1 relaxation, noise, eddy currents, and the spectral complexity of the fat spectrum [13].

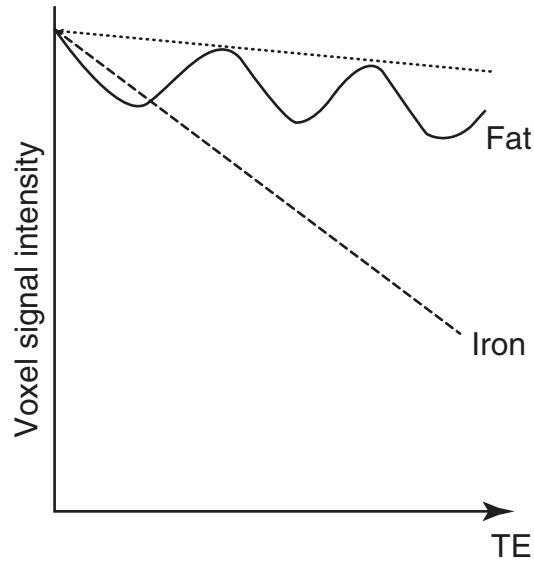


Fig. 20.2 Scheme illustrating the T2* decay in liver parenchyma without fat or iron overload (*dot line*), in livers with fat (solid line) and in livers with iron overload (*dashed line*)

Nevertheless, if all these confounding factors are corrected or minimized, the "signal" fat fraction becomes equivalent to the "proton density" fat fraction (PDFF). In the last decade, the development of multi-echo chemical-shift-encoded MR (MECSE-MR) sequences has provided an accurate tool for PDFF quantification. Nowadays, MR-estimated PDFF is being regarded as the best imaging biomarker for liver steatosis.

20.2.3 Multi-echo Chemical-Shift-Encoded MR (MECSE-MR) Sequences

20.2.3.1 Acquisition: MR Sequence, Biases of T2* effect and T1 (Flip angle)

MECSE-MR sequences are fast, spoiled gradient-recalled echo (SPGR) sequences performed with more than three echoes (usually between 6 and 12), acquiring images of the whole liver within one or few breath holds. The acquisition of multiple echoes allows to separate water and fat signals while simultaneously estimating the T2* decay.

For accurate liver fat quantification, the influence of T2* relaxation and T1 relaxation in the measurements of fat and water signals must be corrected or minimized [13]. If not taken into account, the effect of T2* decay confounds PDFF quantification, either in iron-overloaded or normal livers. Different approaches can be used to correct for the effect of T2* decay by incorporating the T2* in the post-processing fitting model, thereby correcting the T2* decay as part of the fitting, or by measuring the T2* of water and fat separately and then correcting for the effects of T2* [13]. Quite relevant, this T2* estimation also allows for the simultaneous quantification of iron deposits [14, 16–19], which is particularly important as iron overload and steatosis frequently coexist. These methods can assume a single T2* decay for water and fat, or they can perform independent T2* estimations of water and fat. Published reports are controversial about which of these strategies gives improved PDFF quantification [20, 21].

The different T1 relaxation times between water and fat (fat has a shorter T1 than water) may introduce a significant bias in fat fraction estimation. If the acquisition is T1 weighted, fat will have its signal relatively amplified as compared to water. The T1 bias can be reduced either by image acquisition with minimum T1 weighting or postacquisition processing to estimate and correct T1 effect. The T1 weighting can be minimized if a low flip angle is chosen in the gradient-echo sequence [22]. If using a 2D sequence with TR greater than 100 ms, the flip angle should be less than 10°, whereas using a 3D sequence, a flip angle of 2–5° should be preferred. The combination of low flip angle and TR must be chosen to optimize signal-to-noise ratio (close to Ernst angle) and contrast, because the signal-to-noise ratio (SNR) of an SPGR acquisition decreases rapidly at small flip angles [22]. For postacquisition correction, two or more acquisitions should be performed with different T1-weighting parameters, and the estimated T1 bias is then computationally corrected in the final PDFF quantification [23].

20.2.3.2 Imaging Processing and Analysis: Noise, Eddy Currents, and Fat Spectral Complexity

Two major approaches can be used in the imaging processing analysis, both being robust and accurate: the magnitude-based and the complex-based techniques [13]. The firsts discard phase information and only use the magnitude images. Thus, they cannot differentiate which is the dominant tissue component (fat or water) and are unable to quantify fat fraction beyond the 0–50 % spectrum. The complex-based methods use both magnitude and phase information from the different TEs and, in contrast to magnitude-based methods, allow for fat quantification between the range of 0 and 100 %.

Noise is particularly relevant for magnitude-based post-processing methods: after performing the magnitude operation, when the phase information is discarded, the signal intensities of areas with lower fat content will be increased by noise, introducing a bias in PDFF calculation. This effect is more significant at low fat fractions, and it can be avoided using phase constrained or magnitude discrimination methods [22].

Eddy currents may result into phase errors on images acquired at different echo times, such as complex images acquired with these MECSE-MR sequences, leading to bias in PDFF quantification. Eddy currents affect post-processing methods that use phase information (complex methods), whereas magnitude-based methods (which discard phase information) are relatively insensitive to its effects. Hybrid and mixed magnitude/complex fitting methods have been proposed to correct for eddy currents, particularly at 1.5 T acquisitions [22, 24, 25].

Whether magnitude or complex data are used, the acquired liver MR signal is modeled for fat-water separation. The fat signal has at least six distinct frequency components, with different amplitudes (Fig. 20.1). In contrast to dual-echo chemical-shift GRE sequences, which only take into account the frequency of the main fat spectral peak, multipeak reconstruction models consider the complexity of fat spectrum by modeling the fat signal as a weighted sum of all fat frequency com-

ponents. Multipeak methods provide higher accuracy and should be performed for precise measuring of liver fat [18, 26–29]. Because it is impractical to determine every fat peak for each individual liver, pre-calibrated multipeak fat spectral methods have been proposed, in which the resonance frequencies and relative amplitudes of each fat peak are known a priori and based on MR spectroscopy-derived measurements [30–32]. Although no specific multipeak spectral model was shown to be superior to the rest [32], further research is needed to define the influence of the number, location, and fitting of the fat peaks in PDFF quantification with different multipeak reconstruction models.

Parametric maps can be obtained if the post-processing models are applied in a voxel-wise approach, demonstrating the quantity and distribution of PDFF data throughout the liver parenchyma. These parametric maps are particularly useful because of the nonhomogeneous distribution of fat deposits throughout the liver parenchyma (Fig. 20.3, Fig. 20.4). There are

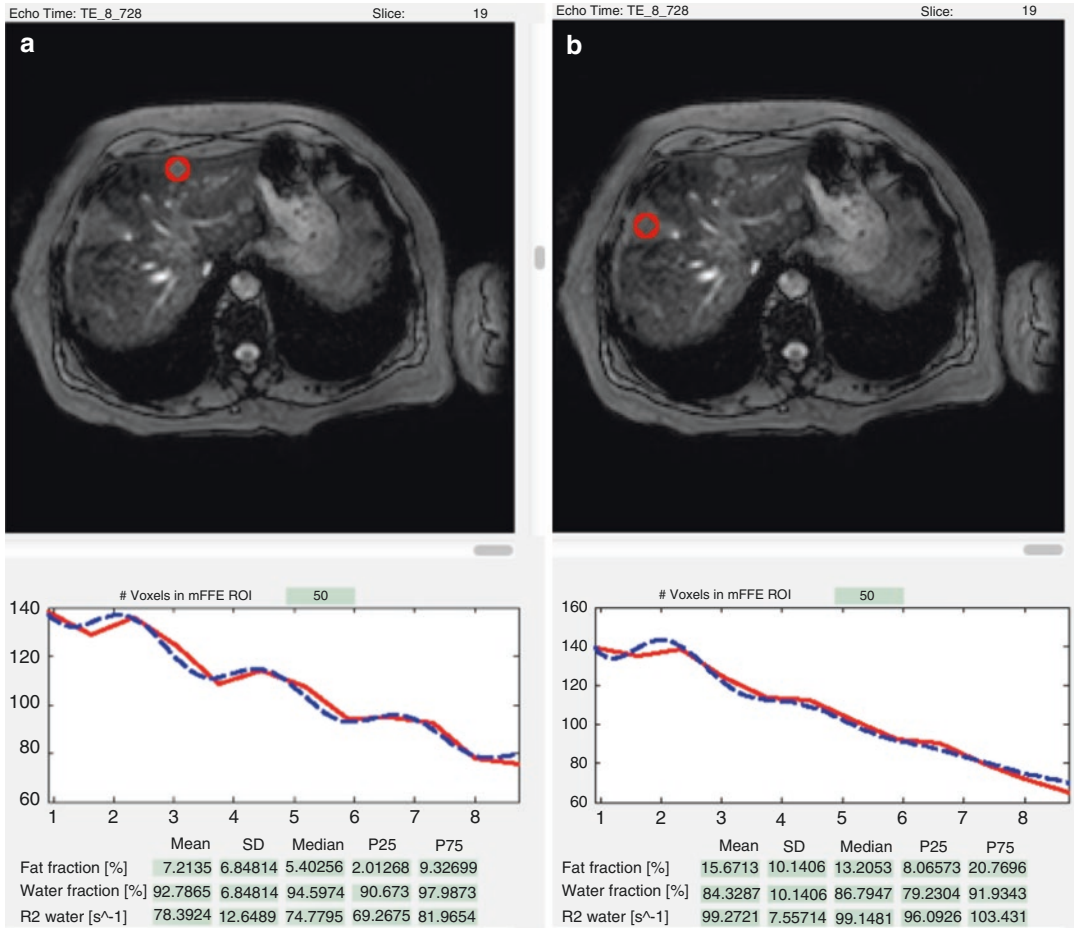


Fig. 20.3 PDFF and R2* quantification with a MECSE-MR sequence. (a) and (b) represent magnitude images used for fat and iron quantification, and the respective signal decay curve calculated with QLiver® software. Images show a heterogeneous and patchy distribution of fat and iron deposits throughout the liver parenchyma. (a) A ROI area in IVa liver segment determined a PDFF=5.4% and R2*=74 s^{-1}. (b) Another ROI placed over the VIII liver segment measured a PDFF=13.2% and R2*=99 s^{-1}. Parametric maps of PDFF (c) and iron related-R2* (d) measurements are particularly useful to demonstrate the heterogeneous distribution of fat and iron deposits throughout the liver parenchyma. (e) Liver biopsy (after hematoxylin-eosin and Perls' Prussian staining, 400x magnification) demonstrating coexistence of steatosis and iron deposits in liver parenchyma

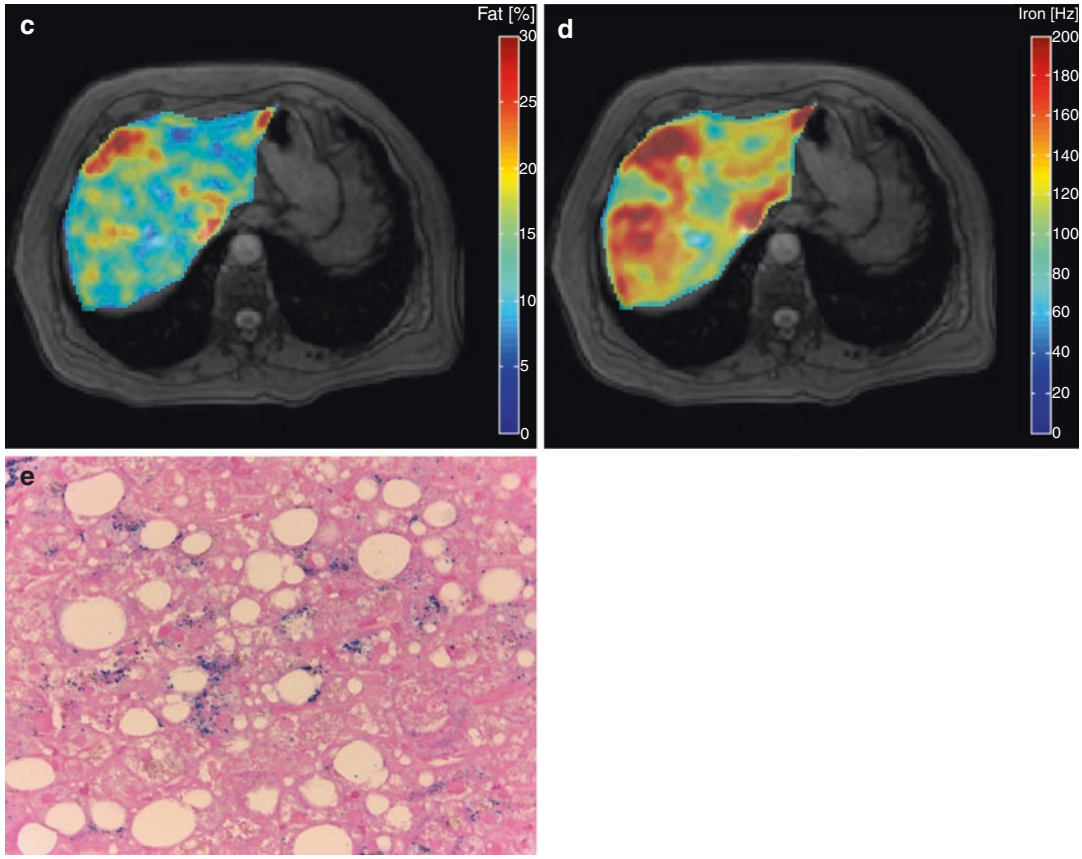


Fig. 20.3 (continued)

some commercial products available to measure PDFF with T1-independent and T2*-corrected MECSE-MR, like IDEAL IQ (General Electric®) and mDIXON QUANT (Philips®).

At their own institutions, the authors are using a 2D MECSE-MR sequence in a 3-T MR scanner, with 12 echoes (TEs = 0.99, 1.69, 2.39, 3.09, 3.79, 4.49, 5.19, 5.89, 6.59, 7.29, 7.99, 8.69; TR = 10 ms, echo spacing = 0.7 ms) and a flip angle of 10° [14]. The whole liver is covered under end-expiratory phase single breath-hold acquisition with 34 axial slices (voxel dimensions, 3 × 3 mm; slice thickness, 7 mm; 0.3 mm gap; reconstruction voxel size, 2 × 2 mm; field of view, 375 × 302 mm; parallel imaging effective acceleration factor, 1.8; bandwidth, 2433 Hz per pixel). Acquired images are exported as raw data to PDFF and R2*(=1/T2*) quantification, using QLiver software (QUIBIM®, Valencia, Spain) (Fig. 20.3, Fig. 20.4).

The complex phase information estimates the resonance peak of the larger component (water or fat), to generate frequency distribution maps. Then, a joint fit between water signal, fat signal, R2* of water, and R2* of fat is performed. A multipeak reconstruction model for fat quantification is used as proposed by Yu et al. (75 %, 420 Hz; 17 %, 318 Hz; 8 %, −94 Hz) [30]. Finally, the pixel PDFF is calculated as the ratio between the normalized fat proton density and the total (fat and water) proton density ($PDFF = PD_{fat}/[PD_{fat} + PD_{water}]$). In addition to T2* correction, the R2* (=1/T2*) of water is used to estimate the iron content.

20.2.4 Proof of Principle

These MECSE-MR imaging sequences were shown to be robust and accurate for PDFF quantification in several studies, either using MRS measurements [17, 29, 33, 34] or liver biopsy [18, 19, 26, 27, 35, 36] as reference standards. Also, PDFF measurements were shown to be repeatable [37–39, 41] and reproducible across different MR scanner platforms at 1.5- and 3-T scanners [42]. Longitudinal hepatic PDFF changes greater than 1.6–1.8 % are likely to represent real changes rather than noise or measurement imprecision [39, 40].

When interpreting PDFF measurements as a biomarker of steatosis, it should be remembered that although PDFF and histological estimated steatosis percentages are highly correlated, they are not equivalent: histologic evaluation measures the percentage of hepatocytes with macrovesicules of fat, whereas PDFF measures the amount of fat protons within the hepatocytes. Recently, PDFF threshold values have been proposed either to diagnose liver steatosis or to distinguish between different histologic grades. PDFF threshold values to diagnose hepatic steatosis range from 2.9 to 7.53 % [18, 26–28, 43, 44].

20.2.5 Proof of Efficacy and Effectiveness

PDFF quantification corrected for the main confounding factors (T1 bias, effect of T2* relaxation, fat spectral complexity, noise bias, and eddy currents) is currently being accepted as the best available imaging biomarker of liver steatosis [45]. MECSE-MR sequences were considered to be more precise than liver biopsy for therapy monitoring in patients with NASH [13, 46]. Other potential applications for these sequences are the living liver donors' evaluation and the pre- and postoperative assessment of patients undergoing bariatric surgery.

Nevertheless, most of the validation studies were conducted in patients with NAFLD and, consequently, could be biased to higher liver fat content. Distinct PDFF thresholds to diagnose hepatic steatosis and to discriminate different histologic grades have been reported in different studies. Therefore, proposed PDFF thresholds still require validation in large cohorts of patients with different clinical scenarios of diffuse liver diseases, before they become standardized and widely accepted. Moreover, the best strategy regarding the number of echoes, the curve fitting, or the multipeak fat modeling has yet to be defined.

20.3 Imaging Biomarkers of Liver Iron: MR Sequences, Measurements, and Biases

In contrast with fat quantification, MR imaging does not quantify iron directly but, instead, depicts the paramagnetic effect of iron on the neighborhood protons. Iron accelerates the T2 relaxation and mainly the T2* signal decay, resulting in a recognizable signal loss on T2- and T2*-weighted images, which is proportional to the iron content.

The MR imaging methods developed for liver iron quantification can be divided into *signal intensity ratio (SIR) methods* and *relaxometry methods*. All of these methods need calibration as they perform measurements (signal intensity ratios or relaxation times or rates) from MR images, which are compared to chemically determined LIC values from liver biopsies (the gold standard), to generate empirical calibration curves [8].

20.3.1 Signal Intensity Ratio (SIR) Methods

In SIR methods, the signal intensity (SI) of the liver is compared to the SI of a reference tissue that does not accumulate iron, usually the skeletal muscle. The most widely used *SIR method* [47] is performed with a body coil, acquiring

five IP GRE sequences, with constant TR but with different flip angles and increasing TE. On each sequence, three ROIs are placed within the right hepatic lobe and two ROIs within the paraspinal muscles to quantify liver and muscle SI. In a free online worksheet provided by the University of Rennes (http://www.radio.univ-rennes1.fr/Sources/FR/HemoCalc15.html), the LIC (μmol Fe/g) is estimated from the mean SI of each ROI. This computer-based algorithm was validated in 149 patients and provides accurate LIC measurements, over a range from 3 to 375 μmol Fe/g dry weight. However, in a subsequent multicenter study with 171 patients, which has compared LIC estimated by MR and LIC chemically quantified, the diagnostic accuracy was 61.4 %, with a tendency to overestimate overload [48]. The differences between the MR-estimated LIC and the biopsy LIC may be important in clinical practice. Nevertheless, this model is very useful to rule out disease, due to its low tendency to underestimate LIC: hemochromatosis is excluded at a cutoff point of less than 60 μmol Fe/g, with a negative predictive value of 100 % [48].

Alústiza et al. [49] proposed a mathematic model for estimating LIC using only two sequences (T2-* and PD-weighted sequences) and, therefore, reducing the acquisition time required for liver iron MR quantification as compared to the original Gandon's method. A free online worksheet for LIC quantification has been provided by the *Sociedad Española de Diagnóstico por Imagen del Abdómen* (SEDIA) (http://www.sedia.es/sedia_investiga/proyectos2007/calculo_hierro/calculoFE.php). In patients suspected of having hemochromatosis, an MR-estimated LIC >79 μmol/g has a positive predictive value of 100 %, whereas an estimated LIC <20 μmol/g has a negative predicted value of 100 %.

In spite of being widely available, these SIR methods have some important limitations. Firstly, they saturate with very high iron overload, with an upper limit value of LIC in the range of 350 μmol Fe/g. Many patients with transfusional hemochromatosis have LIC values higher than this range. Secondly, SI ratios may be confounded by coexisting hepatic steatosis and/or muscle fatty infiltration. Also, these protocols are not compatible with phased-array coils, and, quite relevant, they are not calibrated for 3-T machines. Finally, SIR methods are influenced by the sequence TR, TE, flip angle, and body habitus [8].

20.3.2 Relaxometry Methods

Relaxometry techniques measure relaxation time constants after acquiring series of images with increasing TEs, usually more than six echoes, and can be performed with surface coils. The liver SI is modeled as a function of TE, and signal decay constants are then calculated [8]. The T2 or T2* values (measured in ms) are calculated, depending on whether a spin-echo-based or a gradient-echo-based sequence is performed, respectively. The rates of signal decay R2 (=1/T2) or R2*(=1/T2*) may be used instead and are usually presented in s^{-1}. The underlying proof of concept is that liver T2 and T2* are related to liver iron concentrations: the greater the liver iron concentration, the higher the relaxation rates (R2 or R2*) and the lower the relaxation times (T2 or T2*) [50].

20.3.2.1 R2 Relaxometry

The most known R2 relaxometry method [51] was validated in over 100 patients, with LIC values ranging from 0.3 to 42.7 mg Fe/g dry weight (5–747 μmol Fe/g). Liver R2 had a curvilinear relationship with LIC, with a correlation coefficient of 0.98. This technique is currently available as a commercial service ("FerriScan®"), which has been approved by the Food and Drug Administration (FDA) in the USA. After calibration, MR images are acquired with five T2-weighted sequences, during free breathing, and are then forwarded for centralized image data analysis and R2 measurements. This method has shown a low inter-exam variability and good inter-machine reproducibility [52, 53]. Wood et al. [54] have reproduced the same calibration curve obtained by St. Pierre et al., using different imaging parameters, and a different fitting model (a monoexponential decay model rather than a biexponential decay model). In a recent

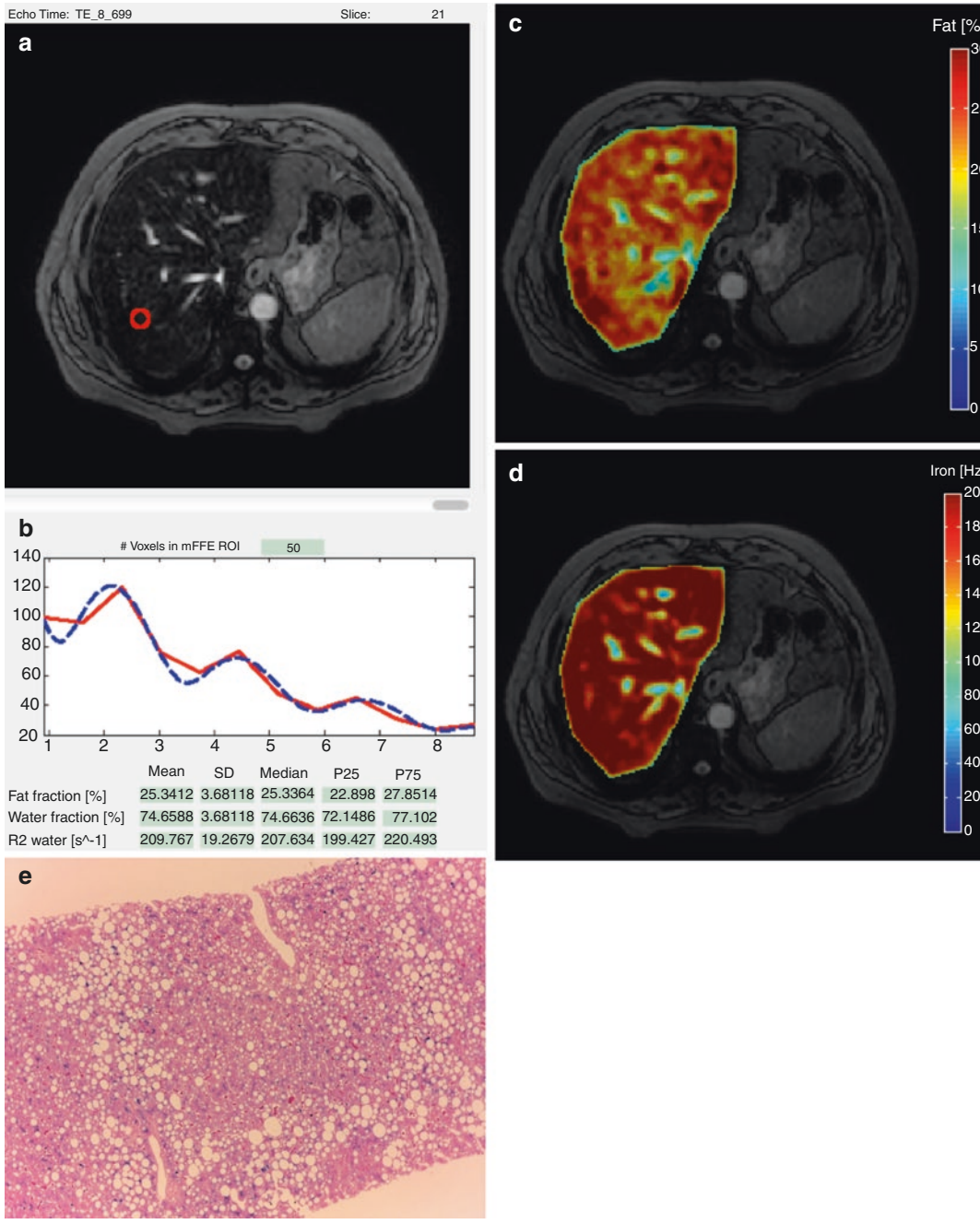


Fig. 20.4 PDFF and R2* quantification with a MECSE-MR sequence. (a) magnitude images used for fat and iron quantification, and the respective signal decay curve calculated with QLiver® software. A ROI area in the VII liver segment determined a PDFF=25% and R2*=208 s^{-1}. (b) Parametric maps of PDFF (c) and iron related-R2* (d) measurements are particularly useful to demonstrate the distribution of fat and iron deposits throughout the liver parenchyma. (e) Liver biopsy (after Perls' Prussian staining, 100x magnification) demonstrating severe steatosis (grade 3/3) and iron overload (grade 3/4). The MECSE-MR measurements were well related with liver biopsy

multicenter validation study, the calibration curve appeared independent of the patient age, stage of liver fibrosis, grade of necroinflammation, and use of chelation therapy. However, the limits of agreement between R2 LIC and biopsy LIC were very broad (between 74 and −71 %) [53].

Using R2 relaxometry, new MR methods have been proposed for separately quantifying the two principal forms of tissue storage iron [55–57]: the dispersed, soluble ferritin iron, which is rapidly mobilized, and the aggregated, insoluble hemosiderin iron, which acts as a long-term reserve. If the MR signal is obtained with varying echo spacing, the signal analysis can be decomposed into the two different forms of iron storage because ferritin iron has a monoexponential decay and the hemosiderin iron has a non-monoexponential. This method could improve the monitoring of iron-reducing therapies, since ferritin is in equilibrium with the cytosolic iron pool that changes rapidly with iron chelation.

20.3.2.2 R2* Relaxometry

R2* relaxometry methods usually are performed with SPGR multi-echo sequences, with increasing TEs. The liver R2* is calculated from the fitting of the rate of exponential signal decay, on either a voxel-by-voxel basis or averaging the measured signal within a ROI [58]. Ideally, the first echo should be as short as possible (1 ms or less), and the echo spacing should be short enough (approximately 1 ms or less), to guarantee that the signal decay is captured. This is particularly relevant in severe iron-overloaded livers: if the first TE is too long, most of the MR signal will have irreversibly disappeared by the first image acquisition [52]. A high number of echoes may result in improved R2* estimation (reduced standard deviation) in cases of low R2* values (low iron content), but not for severe iron-overloaded livers (R2* higher than 1000 s^{-1}), where most of the signal has fully decayed by a TE of 3–4 ms [59]. Finally, the last echo time of 10–15 ms is usually sufficient [60].

The most used R2* relaxometry method [54] was calibrated in 21 patients (23 liver biopsies), in a 1.5-T equipment, using a single breath-hold multi-echo GRE sequence with 17 TE, stepped at 0.25-ms intervals from 0.8 to 4.8 ms. Data were fitted to a monoexponential decay curve with constant offset on a pixel-by-pixel basis, obtaining liver R2* parametric maps. A linear relationship between R2-* and biopsy-based LIC (mg Fe/g) was derived, with a correlation coefficient of 0.97. Other authors have derived different

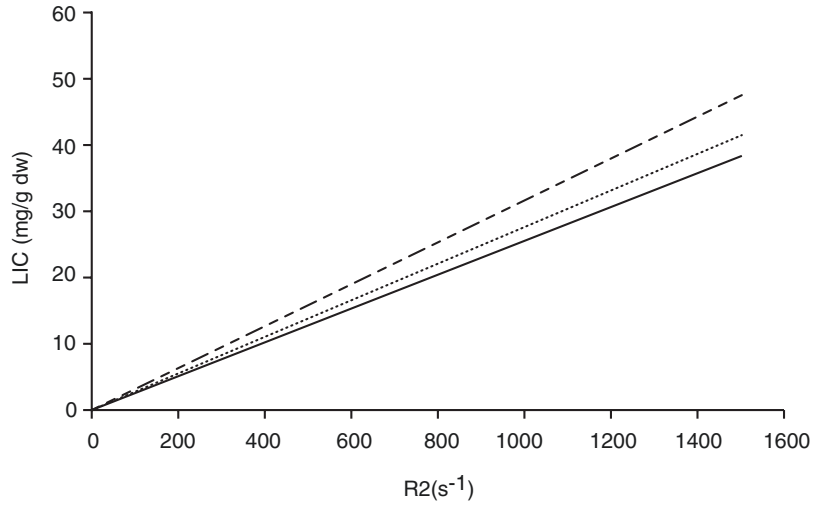


Fig. 20.5 Slope comparison of different R2* (s^{-1};) versus LIC (mg/g dry weight) calibration models, adapted from Garbowski et al [62] . Calibrations model by Garbowski et al (*dashed line*): LIC = 0.032(R2*)-0.14; by J. Wood et al (*solid line*): LIC = 0.0254(R2*) + 0.202; by J. Hankins et al (*dot line*): LIC = 0.028(R2*)-0.45

calibration curves to transform R2* values into LIC values (mg Fe/g or μmol Fe/g), most probably due to differences in methodologies (Fig. 20.5) [61, 62]. R2* relaxometry methods have shown a good reproducibility across different scanners and centers, as also a good interobserver reproducibility [58, 63]. For image analysis, a CE-marked and FDA-approved software is commercially available (Thalassaemia Tools – https://www.cmrtools.com).

The R2* exponential model fitting can be done using either the complex signal or the signal magnitude, the latter being prone to a "noise-floor" effect that can introduce systematic errors in the R2* measurements, which are particularly relevant in cases with severe iron overload (R2* of 1000 s^{-1} or higher) [58]. To overcome noise-floor effects, several approaches have been proposed, with advantage for using complex fitting models because it simultaneously obviates noise-floor effects and corrects for the confounding effect of fat in iron quantification.

Relaxation rates are dependent on magnetic field strength, and, thus, calibration curves obtained at one field cannot be directly transposed to another field strength. Some studies performed indirect calibration of 3 T T2*/R2* values against 1.5 T values and have found a linear relationship between R2* measurements at 1.5 and 3 T with a factor of 2 [64, 65].

Both R2 and R2* methods have theoretical advantages and disadvantages. R2 techniques are less sensitive to external magnetic inhomogeneities, while R2* measurements are less sensitive to variations in size of iron particles [54]. R2 techniques require a longer acquisition time (5–20 min) and are more prone to motion artifacts, while R2* methods can be performed within a single breath hold. Quite relevant, R2 and R2* methods have shown a strong agreement between the MR measurements and the biopsy-quantified LIC [51, 54], however with very broad limits of agreement [54]. This may be partially attributed to the spatial variability of iron deposits within the liver. In addition, relaxation rates depend not only on iron concentration but also on other features such as iron particle size, shape, and distribution and the coexistence of steatosis, inflammation, or fibrosis [58]. Specifically, if relaxometry techniques model the T2* decay as a monoexponential decay, they will not account for the signal oscillations as water and fat signals become in and out of phase, and, therefore, they will be confounded by the coexistence of liver fat.

20.3.2.3 MECSE-MR Sequences: Simultaneous Quantification of Iron and Fat

MECSE-MR sequences, the same used for PDFF quantification, can be used for R2* estimation and liver iron quantification [8, 14, 18, 58, 59] while simultaneously correcting and assessing for liver fat content (Fig. 20.3, Fig. 20.4, Fig. 20.6). The use of complex fitting and fat correction improves the robustness, noise performance, and accuracy of R2* measurements as an imaging biomarker of liver iron [59].

20.3.3 Proof of Effectiveness

Both R2 or R2* values can be used for LIC estimates in clinical practice, either for diagnosis of iron overload or treatment monitoring [50, 53, 66, 67], as long as these methods are performed with validated acquisition and analysis protocols. The choice of relaxometry technique may depend on software availability and local expertise at any given institution. R2* methods might be preferable because they have faster acquisition times [58]. Assessment of T2* relaxation rates at 3 T seemed to be feasible [64, 68, 69], reproducible, and reliable to quantify iron burden. However, the accuracy was lower for detecting heavy-moderate liver iron burden [65]. The increase in R2* (shortening of T2*) and the significantly higher T2* susceptibility artifacts at higher magnetic field strength make accurate quantification even more challenging due to the rapid signal decay, which may limit the dynamic range for accurate iron quantification at 3 T [65].

Although both R2 and R2* methods are considered accurate for iron quantification, differences exist between measurements obtained with different techniques and post-processing models [51, 54, 61, 62] (Fig. 20.5). Therefore, R2 and R2* measurements obtained with different methods should not be used interchangeably. It is important to highlight that these methods do not detect iron directly, but rely on empirical approaches and calibration with LIC obtained from liver biopsies. Comprehensibly, they are influenced by technical factors and fitting models and are subject to individual variability. Furthermore, the gold standard itself (LIC determined from liver biopsy) is subject to sampling bias and has a coefficient of variation up to 40 % in cirrhotic livers [67]. Moreover, differences in R2*-LIC calibrations may also result from different post-biopsy sample processing (e.g., LIC measurements obtained from paraffin-embedded liver samples are lower than those from fresh liver samples)

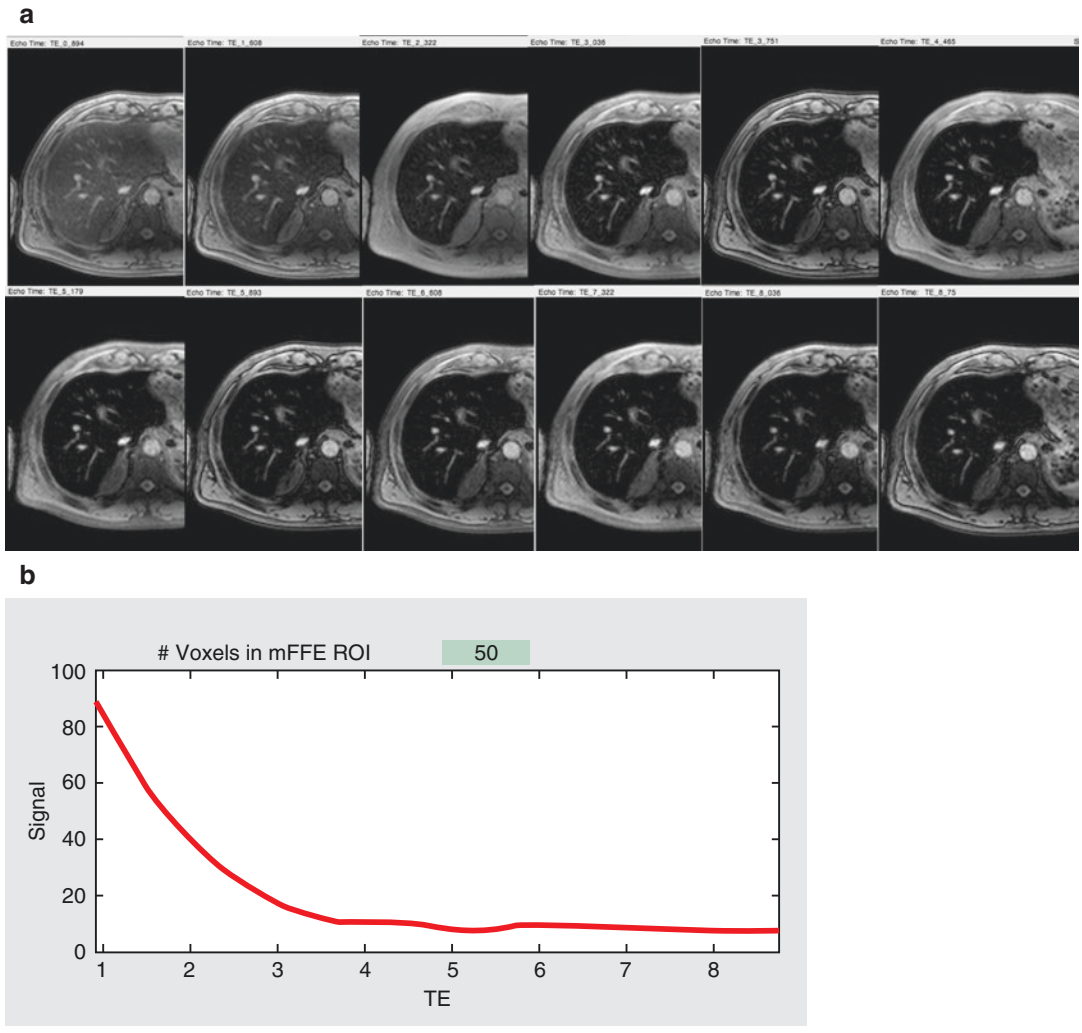


Fig. 20.6 Iron (R2*) quantification by a MECSE-MR sequence. (a) Magnitude images from 12 different TEs demonstrate severe liver signal dropout with increasing TEs, which is related to iron overload. (b) A plot of mean signal intensity within a circular ROI is modeled as a function of TE, as a bi-exponential decay curve. Estimated R2* is 710s-1. Patient had had a liver biopsy scored as grade 4 for iron overload and a LIC of 115 ½mol/g dry liver

[50, 62]. Therefore, to facilitate comparisons across different studies and methodologies, the R2 or R2* values should be converted into LIC values using the appropriate calibration curve [68]. Quite relevant, the same relaxometry technique should be used when following patients over time [67].

In a recent study [66], MR relaxometry was superior to liver biopsy for serial LIC observations, and it was proposed that R2*-derived LIC measurements could replace liver biopsy as a surrogate for chelator effectiveness in clinical trials.

Nevertheless, R2* relaxometry still lacks technical standardization. A consensus is needed regarding optimized imaging acquisition and post-processing strategies in order to standardize protocols. Also, there is a need to establish universally accepted MR-positivity thresholds and to investigate the effects of MR surveillance on patient outcomes [70].

Conclusion

Noninvasive imaging biomarkers for quantification of hepatic fat and iron deposits are now available using dedicated MR sequences and post-processing algorithms. When these biomarkers become widely standardized, these new quantitative imaging techniques might be performed as "virtual" biopsies. In the setting of diffuse liver diseases, these imaging biomarkers' measurements should be part of a structured radiological report, in order to be used for diagnosis and longitudinal monitoring of hepatic fat and/or iron overload.

References

1. Schwenzer NF, Springer F, Schraml C, Stefan N, Machann J, Schick F. Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. J Hepatol. 2009;51:433–45.
2. Wells SA. Quantification of hepatic fat and iron with magnetic resonance imaging. Magn Reson Imaging Clin N Am. 2021;22:397–416.
3. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology. 2012;55:2005–23.
4. Pietrangelo A. Iron in NASH, chronic liver diseases and HCC: how much iron is too much? J Hepatol. 2009;50:249–51.
5. Dongiovanni P, Fracanzani AL, Fargion S, Valenti L. Iron in fatty liver and in the metabolic syndrome: a promising therapeutic target. J Hepatol. 2011;55:920–32.
6. Nelson JE, Wilson L, Brunt EM, et al. Relationship between the pattern of hepatic iron deposition and histological severity in nonalcoholic fatty liver disease. Hepatology. 2010;53:448–57.
7. Powell EE, Ali A, Clouston AD, et al. Steatosis is a cofactor in liver injury in hemochromatosis. Gastroenterology. 2005;129:1937–43.
8. Sirlin CB, Reeder SB. Magnetic resonance imaging quantification of liver iron. Magn Reson Imaging Clin N Am. 2010;18:359–81.
9. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41:1313–21.
10. Deugnier Y, Turlin B. Pathology of hepatic iron overload. World J Gastroenterol. 2007;13:4755–60.
11. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology. 2005;128:1898–906.
12. Ligabue G, Besutti G, Scaglioni R, Stentarelli C, Guaraldi G. MR quantitative biomarkers of nonalcoholic fatty liver disease: technical evolutions and future trends. Quant Imaging Med Surg. 2013;3:192–5.
13. Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. J Magn Reson Imaging. 2011;34:spcone–spcone
14. Martí-Bonmatí L, Alberich-Bayarri A, Sánchez-González J. Overload hepatitides: quanti-qualitative analysis. Abdom Imaging. 2011;37:180–7.
15. Westphalen ACA, Qayyum A, Yeh BM, et al. Liver fat: effect of hepatic iron deposition on evaluation with opposed-phase MR imaging. Radiology. 2007;242:450–5.
16. O'Regan DP, Callaghan MF, Wylezinska-Arridge M, et al. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T – feasibility. Radiology. 2008;247:550–7.
17. Yokoo T, Shiehmorteza M, Hamilton G, et al. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology. 2011;258:749–59.
18. Kühn J-P, Hernando D, del Rio AM, et al. Effect of multipeak spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. Radiology. 2012;265:133–42.

19. McPherson S, Jonsson JR, Cowin GJ, et al. Magnetic resonance imaging and spectroscopy accurately estimate the severity of steatosis provided the stage of fibrosis is considered. J Hepatol. 2009;51:389–97.
20. Chebrolu VV, Hines CDG, Yu H, et al. Independent estimation of T* 2 for water and fat for improved accuracy of fat quantification. Magn Reson Med. 2010;63:849–57.
21. Horng DE, Hernando D, Hines CDG, Reeder SB. Comparison of R2* correction methods for accurate fat quantification in fatty liver. J Magn Reson Imaging. 2012;37:414–22.
22. Liu C-Y, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T1 and noise. Magn Reson Med. 2007;58:354–64.
23. Yokoo T, Browning JD. Fat and iron quantification in the liver: past, present, and future. Top Magn Reson Imaging. 2014;23:73–94.
24. Yu H, Shimakawa A, Hines CDG, et al. Combination of complex-based and magnitude-based multiecho water-fat separation for accurate quantification of fat-fraction. Magn Reson Med. 2011;66:199–206.
25. Hernando D, Hines CDG, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magn Reson Med. 2011;67:638–44.
26. Idilman IS, Aniktar H, Idilman R, et al. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology. 2013;267:767–75.
27. Tang A, Tan J, Sun M, et al. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology. 2013;267:422–31.
28. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging–estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. Radiology. 2015;274:416–25.
29. Meisamy S, Hines CDG, Hamilton G, et al. Quantification of hepatic steatosis with T1-independent, T2*-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology. 2011;258:767–75.
30. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn Reson Med. 2008;60:1122–34.
31. Hamilton G, Yokoo T, Bydder M, et al. In vivo characterization of the liver fat 1H MR spectrum. NMR Biomed. 2010;24:784–90.
32. Wang X, Hernando D, Reeder SB. Sensitivity of chemical shift-encoded fat quantification to calibration of fat MR spectrum. Magn Reson Med. 2015;75(2):845–51.
33. Yokoo T, Bydder M, Hamilton G, et al. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T1. Radiology. 2009;251:67–76.
34. Kühn J-P, Hernando D, Mensel B, et al. Quantitative chemical shift-encoded MRI is an accurate method to quantify hepatic steatosis. J Magn Reson Imaging. 2014;39:1494–501.
35. Permutt Z, Le T-A, Peterson MR, et al. Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease – MRI accurately quantifies hepatic steatosis in NAFLD. Aliment Pharmacol Ther. 2012;36:22–9.
36. Parente DB, Rodrigues RS, Paiva FF, et al. Is MR spectroscopy really the best MR-based method for the evaluation of fatty liver in diabetic patients in clinical practice? Bonino F, editor. PLoS One. 2014;9:e112574.
37. Sofue K, Mileto A, Dale BM, Zhong X, Bashir MR. Interexamination repeatability and spatial heterogeneity of liver iron and fat quantification using MRI-based multistep adaptive fitting algorithm. J Magn Reson Imaging. 2015;42(5):1281–90.
38. Johnson BL, Schroeder ME, Wolfson T, et al. Effect of flip angle on the accuracy and repeatability of hepatic proton density fat fraction estimation by complex data-based, T1-independent, T2*-corrected, spectrum-modeled MRI. J Magn Reson Imaging. 2013;39:440–7.
39. Negrete LM, Middleton MS, Clark L, et al. Interexamination precision of magnitude-based MRI for estimation of segmental hepatic proton density fat fraction in obese subjects. J Magn Reson Imaging. 2013;39:1265–71.
40. Tyagi A, Yeganeh O, Levin Y, et al. Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. Abdom Imaging. 2015;40(8):3070–7.
41. Hines CDG, Frydrychowicz A, Hamilton G, et al. T1 independent, T2* corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. J Magn Reson Imaging. 2011;33:873–81.
42. Kang GH, Cruite I, Shiehmorteza M, et al. Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. J Magn Reson Imaging. 2011;34:928–34.
43. Kühn J-P, Evert M, Friedrich N, et al. Noninvasive quantification of hepatic fat content using three-echo dixon magnetic resonance imaging with correction for T2* relaxation effects. Invest Radiol. 2011;46:783–9.
44. Kang B-K, Yu ES, Lee SS, et al. Hepatic fat quantification: a prospective comparison of magnetic resonance spectroscopy and analysis methods for chemical-shift gradient echo magnetic resonance imaging with histologic assessment as the reference standard. Invest Radiol. 2012;47:368–75.

45. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized mr-based biomarker of tissue fat concentration. J Magn Reson Imaging. 2012;36:1011–4.
46. Noureddin M, Lam J, Peterson MR, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology. 2013;58(6):1930–40.
47. Gandon Y, Olivié D, Guyader D, et al. Non-invasive assessment of hepatic iron stores by MRI. Lancet. 2004;363:357–62.
48. Castiella A, Alústiza JM, Emparanza JI, Zapata EM, Costero B, Díez MI. Liver iron concentration quantification by MRI: are recommended protocols accurate enough for clinical practice? Eur Radiol. 2010;21:137–41.
49. Alústiza JMJ, Artetxe JJ, Castiella AA, et al. MR quantification of hepatic iron concentration. Radiology. 2004;230:479–84.
50. PhD JCWM. Use of magnetic resonance imaging to monitor iron overload. Hematol Oncol Clin North Am. 2014;28:747–64.
51. St Pierre TG. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. Blood. 2005;105:855–61.
52. Wood JC, Ghugre N. Magnetic resonance imaging assessment of excess iron in thalassemia, sickle cell disease and other iron overload diseases*. Hemoglobin. 2008;32:85–96.
53. St Pierre TG, El-Beshlawy A, Elalfy M, et al. Multicenter validation of spin-density projection-assisted R2-MRI for the noninvasive measurement of liver iron concentration. Magn Reson Med. 2013;71:2215–23.
54. Wood JC. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. Blood. 2005;106:1460–5.
55. Tang H, Jensen JH, Sammet CL, et al. MR characterization of hepatic storage iron in transfusional iron overload. J Magn Reson Imaging. 2014;39:307–16.
56. Jensen JH, Tang H, Tosti CL, et al. Separate MRI quantification of dispersed (ferritin-like) and aggregated (hemosiderin-like) storage iron. Magn Reson Med. 2010;63:1201–9.
57. Wu EX, Kim D, Tosti CL, et al. Magnetic resonance assessment of iron overload by separate measurement of tissue ferritin and hemosiderin iron. Ann N Y Acad Sci. 2010;1202:115–22.
58. Hernando D, Levin YS, Sirlin CB, Reeder SB. Quantification of liver iron with MRI: state of the art and remaining challenges. J Magn Reson Imaging. 2014;40(5):1003–21.
59. Hernando D, Kramer JH, Reeder SB. Multipeak fat-corrected complex R2* relaxometry: theory, optimization, and clinical validation. Magn Reson Med. 2013;70:1319–31.
60. Hernando D, Wells SA, Vigen KK, Reeder SB. Magnetic resonance imaging. Magn Reson Imaging. 2015;33:43–50.
61. Hankins JS, McCarville MB, Loeffler RB, et al. R2* magnetic resonance imaging of the liver in patients with iron overload. Blood. 2009;113:4853–5.
62. Garbowski MW, Carpenter J-P, Smith G, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. J Cardiovasc Magn Reson. 2014;16:1–11.
63. Kirk P, He T, Anderson LJ, et al. International reproducibility of single breathhold T2* MR for cardiac and liver iron assessment among five thalassemia centers. J Magn Reson Imaging. 2010;32:315–9.
64. Storey P, Thompson AA, Carqueville CL, Wood JC, de Freitas RA, Rigsby CK. R2* imaging of transfusional iron burden at 3T and comparison with 1.5T. J Magn Reson Imaging. 2007;25:540–7.
65. Meloni A, Positano V, Keilberg P, et al. Feasibility, reproducibility, and reliability for the T*2 iron evaluation at 3 T in comparison with 1.5 T. Magn Reson Med. 2012;68:543–51.
66. Wood JC, Zhang P, Rienhoff H, Abi-Saab W, Neufeld EJ. Magnetic resonance imaging. Magn Reson Imaging. 2015;33:761–7.
67. Wood JC, Zhang P, Rienhoff H, Abi-Saab W, Neufeld E. R2 and R2* are equally effective in evaluating chronic response to iron chelation. Am J Hematol. 2014;89:505–8.
68. Meloni A, Rienhoff HY, Jones A, Pepe A. The use of appropriate calibration curves corrects for systematic differences in liver R2* values measured using different software packages. Br J Haematol. 2013;161(6):888–91.
69. Anwar M, Wood J, Manwani D, Taragin B, Oyeku SO, Peng Q. Hepatic iron quantification on 3 Tesla (3 T) magnetic resonance (MR): technical challenges and solutions. Radiol Res Pract. 2013;2013:1–7.
70. Sarigianni M, Liakos A, Vlachaki E, et al. Accuracy of Magnetic Resonance Imaging in Diagnosis of Liver Iron Overload: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2015;13(1):55–63.

Imaging Biomarkers in Clinical Trials

21

Yan Liu and John C. Waterton

21.1 What Is a Clinical Trial?

A clinical trial is a study of human subjects to test a medical hypothesis in order to understand the causes, development, and effects of diseases and improve preventive, diagnostic, and therapeutic interventions. Drugs are a very important class of preventive and therapeutic intervention. Moreover, the principles established in trials of investigational drug therapies are readily applied in the evaluation of other types of therapy, such as radiotherapy, surgery, ablation therapy, environmental modification, or talking therapies.

With very limited exceptions, a new drug must be approved by a regulatory authority before it can legally be manufactured, distributed, and prescribed. The United States (US) Food and Drug Administration (FDA) [1] is the longest established and perhaps the most influential regulator: important regulators in other jurisdictions include European Medicines Agency (EMA, Europe) [2], China Food and Drug Administration (CFDA, China) [3], and Ministry of Health, Labour and Welfare (MHLW, Japan) [4]. Regulatory procedures are to some extent standardized through International Conference on Harmonization (ICH) [5]. Regulators determine whether the likely benefit (*efficacy*) of an investigational new drug outweighs the risk of harm (*safety*), and they ensure that the drug can be supplied to consistent *quality*. Some jurisdictions impose a fourth hurdle, *cost-effectiveness*.

Definitive evidence for efficacy and safety is provided by "pivotal" or Phase III trials (typically 3–5 years), which typically compare the best currently available treatment with the investigational drug at intended dose in the intended patient population. Depending on the indication, they may involve as few as 100 patients up to many tens of thousands, lasting a few weeks for acute conditions up to several years. In pivotal trials, regulators are mainly interested in how a patient feels and functions or how long they survive, but imaging biomarkers can also have an important supporting role as will be discussed below.

Phase II trials (typically 2 years) in patients provide information needed in order to design Phase III. They typically seek evidence of safety issues and may explore different doses, formulations, and schedules, together with preliminary evidence of efficacy (so-called proof of principle or proof of concept) and pharmacodynamics (PD) ("what the drug does to the body"). Phase II trials may involve a few tens or hundreds of

Y. Liu (✉)
European Organisation for Research and Treatment of Cancer, Headquarters, Brussels, Belgium
e-mail: yan.liu@eortc.be

J.C. Waterton
Biomedical Imaging Institute, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, UK
e-mail: John.waterton@manchester.ac.uk

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6_21

patients, often in a small number of academic centers. Imaging biomarkers are often critically important in Phase II trials as will be discussed below.

Phase I trials (typically up to 1.5 years) are mainly designed to determine pharmacokinetics (PK) ("what the body does to the drug"). Usually they are performed in healthy volunteers, although in some serious diseases such as cancer, phase I trials may be performed in patients. Phase I trials may additionally seek preliminary evidence that the drug indeed modulates its target (so-called proof of mechanism) and PD. Imaging biomarkers are often used to assess PK and PD as will be discussed below.

A drug will only be taken into man following compelling evidence from animal and in vitro studies which is made available to trialists in the "Investigators' Brochure" (i.e., a comprehensive document summarizing investigational product information) in accordance with Article 21 of the Declaration of Helsinki [6]. It would be ethically questionable to expose human subjects for the first time to an investigational new drug in the absence of compelling preclinical evidence of safety and efficacy. Where it is proposed to expose patients to a costly and burdensome imaging test, the investigator may require evidence from preclinical studies that the imaging biomarker will likely provide useful data.

The term Phase 0 is sometimes used for trials where no drug is used, or the drug is tested in humans, at subtherapeutic doses with PK analysis using positron emission tomography (PET) or accelerator mass spectrometry (AMS).

Phase IV trials are studies performed after marketing authorization, to further the understanding of the drug in "real world" settings. Large post-marketing surveillance studies are often conducted on new drugs to identify rare adverse events. The entire drug development takes typically 15 years from invention to FDA approval.

Drug developers are facing enormous challenges [7], including the long development process and a productivity crisis of their own. Between 2002 and 2011, the pharmaceutical and biotech sector spent nearly 1.1 trillion dollars on research and development (R&D). The US FDA approved 308 new molecular entities and biologics in the 10 years to 2011. It suggests that the average cost per approved molecule ranged from 2.3 to 4.9 billion dollars [8]. Only about two of every ten marketed drugs generate sufficient revenues to cover their associated R&D cost [9]. Compared to other diseases, oncology has the highest attrition rate for late-stage clinical trials, and overall success rate from first-in-man to approval is about 5 % [10]. Novel technologies and optimized trial designs are needed to reduce those gaps and disease and the production attrition rate, and this is where imaging biomarkers can play a critical role.

21.2 Trial EndPoints

From the perspective of the regulatory authority, the measurements of interest in trials can be divided into "clinical endpoints" and "surrogate endpoints." Clinical trials use clinical endpoints (outcomes) to establish whether the therapy is safe and effective. A clinical endpoint is defined as a characteristic or variable that measures how a patient feels, functions, or survives. The classic endpoint of mortality determines whether the new therapy decreases the rate of death (i.e., overall survival) in comparison with that of a control group. Another endpoint is morbidity, in which investigators examine whether patients undergoing the therapy suffer less from the illness (e.g., function better, or enjoy a higher quality of life) than those who do not receive the therapy. Except in diseases with very poor prognosis, trials with these clinical outcomes often have a long duration and require a large number of subjects and are therefore extremely costly.

A surrogate endpoint is a special category of biomarker (and therefore derived, e.g., from a laboratory test, radiographic image, or physical sign) that in the opinion of the regulatory authority can substitute for a clinical endpoint in measuring treatment effects and in specific circumstances [11]. Regulatory approval may be granted when a drug has an effect on a surrogate endpoint that is likely to predict clinical benefits.

From January 1, 2010, to December 31, 2014, FDA approved 197 novel drugs and biologics, and 84 relied upon surrogate endpoints [12]. Many of these drugs have orphan designation (i.e., rare disease), so an accelerated approval helps a faster time to market, and patients with those life-threatening disease benefit from earlier access to new treatments.

FDA has often been urged to accept novel surrogate endpoints in order to reduce the time and cost of trials and bring new drugs more quickly to patients. However, if a poorly validated surrogate endpoint is used, there is a risk that the regulator causes an ineffective or harmful drug to enter the healthcare system, as occurred disastrously when class 1 antiarrhythmics were approved on the basis of an invalid electrophysiologic biomarker endpoint in the 1970s. Such mistakes are extremely difficult to rectify [13]. A true surrogate endpoint requires an exhaustive validation, i.e., strong evidence of positively predicting clinical outcome with negligible risk of error. A qualified biomarker accepted by the FDA as a surrogate endpoint must match several important criteria: (I) the endpoint must have an accepted, standardized definition; (II) data from multiple clinical studies must demonstrate a strong correlation of the surrogate endpoint with clinical outcome; (III) well-powered prospective studies must have been performed to validate the surrogate endpoint (i.e., truly predictive of clinical benefit with meaningful improvement in patient outcome); and (IV) prospective studies to determine if the surrogate endpoint can be generalized to other patient populations, other target organs, or drugs with other mechanisms [14]. The strength of evidence will vary, depending on whether the surrogate is intended for use in accelerated approval or definite regulatory approval.

Nearly half (84/197) of the new drugs approved by FDA in the last 5 years were approved with surrogate endpoints, and of these, 28 used imaging biomarker endpoints [12]. Even in oncology [15], where overall survival remains the gold standard, imaging biomarkers based on tumor morphology have been increasingly used as surrogate endpoints, specifically progression-free survival (PFS), i.e., tumor does not grow or upstage and objective response rate (ORR), i.e., the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period. Both PFS and ORR are imaging biomarkers. Those surrogate endpoints are more objective and faster to measure than clinical outcomes, allow smaller group sizes and may provide insight into tumor biology *in vivo*.

Regulatory authorities are also interested in so-called safety biomarkers as surrogate endpoints of harm (or, rather, lack of harm); imaging biomarkers such as bone mineral density (BMD) and left ventricular ejection fraction (LVEF) play a major role. Many anticancer therapies carry significant risk of cardiotoxicity. This is true not only for well-established treatments such as cytotoxic chemotherapy (e.g., doxorubicin) or radiotherapy but also for many recently introduced tyrosine kinase inhibitors, such as lapatinib and sunitinib. Indeed the FDA approvals *inter alia* for lapatinib, sunitinib, and doxorubicin each require the use of the LVEF imaging biomarker. Current guidelines [16] define cancer therapeutics-related cardiac dysfunction as a decrease in the LVEF of >10 percentage points to a value <53 % (normal reference value for two-dimensional echocardiography).

21.3 What Is the Value of Imaging in Therapeutic Drug Trials?

Imaging biomarkers are involved throughout the drug development process (Table 21.1) and serve many purposes other than providing surrogate endpoints. Imaging biomarkers can determinate patient eligibility, can stratify the patient population, can identify and validate therapeutic targets, and can provide evidence of drug efficacy. In addition, after the drug enters use, imaging biomarkers play important roles in monitoring for drug safety and for relapse.

The roles of imaging therefore change through the lifecycle of drug development and use and adapt to the objectives of the study, disease types, clinical situation, and feasibility. Some examples of using imaging biomarkers in the different stages of clinical trials are demonstrated in Table 21.2.

Table 21.1 The role of imaging in clinical trials

The role in a clinical trial	Imaging biomarker type	How widely used	Comments
Determinate eligibility	Morphologic imaging (e.g., X-ray, mammography, CT, MRI) Whole-body FDG-PET	Almost universal	Established biomarkers
Optimize patient population (patient stratification)	Molecular imaging (e.g., hypoxia imaging, metabolic imaging)	Emerging	Biomarkers have impact on patient management
Verify that the therapeutic target is present and accessible	Molecular imaging (e.g., drug-labeled radiotracer)	Commonly for certain targets	Biomarker used for proof of concept
Verify that the drug has the desired effect on the pathway and local physiology (e.g., define dose and schedule)	Molecular and functional imaging	Less often	Biomarker used for proof of concept
Evaluate efficacy	Conventional morphologic imaging	Almost universal	Established biomarkers
	Molecular and functional imaging	Commonly for certain targets	Biomarker used for proof of concept
Evaluate safety	Morphologic and functional imaging	Very often	Established biomarkers
Drug approval (safety and efficacy)	Conventional morphologic imaging	Especially for accelerated approvals	Surrogate endpoint

21.3.1 Determinate Eligibility

Eligibility criteria in clinical trials are designed to ensure the safety of research participants and tailored based on scientific objective of a trial. The eligibility criteria differ from trial to trial, but a clear diagnosis of disease characteristics and disease staging are often essential before patient enrollment. In oncology, the TNM staging system is widely used in trials in all solid tumors and remains a vital component of eligibility criteria. The TNM may be classified as an imaging biomarker, as it usually relies on imaging to define the size and extent of the primary tumor (T), any lymphatic involvement (N), and the presence of metastases (M). Whole-body PET imaging improves TNM assessment from morphologic imaging, as it reveals metastases not evident on conventional CT or MR imaging. Molecular imaging techniques are increasingly used to augment conventional imaging to confirm patient eligibility. For example, imaging with [^{18}F]-FDG has been used in the EORTC LungTech trial, to ensure that enrolled patients are with inoperable early stage non-small cell lung cancer (NSCLC) [17].

21.3.2 Optimize Patient Population (Patient Stratification)

Clinical trials can be enriched by selecting subpopulations that may be more responsive to treatments, so as to improve the chance of trial success. Molecular and functional imaging provides additional information on tumor microenvironment, which could help to "preselect" sub-patient populations. For example, identification of tumor hypoxia could facilitate the use of hypoxia-stimulated pro-drugs, which selectively kill hypoxic cells. Tirapazamine, a cytotoxic agent with high selective toxicity toward hypoxic cells, is a good example. The relatively limited benefit obtained in a trial of NSCLC reported by the CATAPULT I study group was likely due to poor patient stratification with inclusion of patients with better-oxygenated tumors [18]. Rischin and his colleagues compared the cisplatin/5-FU versus cisplatin/tirapazamine regimen in patients with head and neck squamous-cell cancer (HNSCC), where [^{18}F]-fluoromisonidazole (FMISO)-PET

Table 21.2 Imaging biomarker at different stages of clinical trials

Phase	Objective	Scope	The role of imaging biomarkers	Examples of imaging biomarker used
Phase 0	First in human to test PK and PD	Single subtherapeutic doses of the study drug or treatment are given to a small number of subjects (10–15) to gather preliminary data on the agent's PK and PD	Visualize whole-body drug bio-distribution and potentially provide insight into tumor characteristic *in vivo*	Radio-labeled tracers such as ^{89}Zr-trastuzumab visualize their targets in the case of HER2 in breast cancer (NCT02065609) or ^{89}Zr-bevacizumab in antiangiogenic treatment (NCT01894451)
Phase 1	Gathering evidence for the safety and appropriate dose of the drug in healthy volunteers or in patients	Testing within a small group of people (20–80) to evaluate safety, determine safe dosage ranges, and begin to identify side effects	Evaluate safety and optimize drug dose escalation	DCE-MRI used for antiangiogenic drug dose selection [26] (efficacy) LVEF used as a cardiac functional biomarker in tyrosine kinase inhibitor trials [49] (safety)
Phase 2	Gathering evidence for the safety and efficacy of the drug in patients	Testing with a larger group of people (100–200) to see if it is effective and to further evaluate its safety	Assess objective response and continue evaluate safety	RECIST [50] or modified RECIST [51, 52] to evaluate tumor objective response (efficacy) MRI is commonly employed for safety monitoring in Alzheimer's disease, especial in trials with investigational immunotherapy where vasogenic edema is a recognized risk [53]
Phase 3	Final confirmation of safety and efficacy to satisfy regulators	Testing with large groups of people (500–3000) to confirm its effectiveness, monitor side effects, and compare it to commonly used treatments	Confirm objective response or progression	Various criteria [54] based on imaging biomarkers obtained from tumor phenotype changes (e.g., size, density) or new lesion presence Cartilage thickness to evaluate response in knee osteoarthritis trials [55]
Phase 4	Sentry studies after marketing authorization	Post-marketing studies delineate additional information, including the treatment's risks, benefits, and optimal use	Routine imaging to evaluate response, progression, or safety	Imaging modalities used for routine disease evaluation

hypoxia imaging was used to stratify the tumors into hypoxic and non-hypoxic ones. The study showed that tirapazamine improved local tumor control in hypoxic but not in non-hypoxic tumors [19]. Similar examples are imaging with ^{99m}Tc-etarfolatide used to identify folate receptor-positive patients with advanced ovarian cancer, who are most likely to benefit from treatment with vintafolide, a folate receptor-targeted therapy [20], and [^{18}F]-FDG used in patients with gefitinib-treated NSCLC, which showed a low baseline SUV of [^{18}F]-FDG associated with a higher response rate (53 % versus 18 %) and a prolonged PFS (median, 33.1 weeks versus 8.6 weeks) [21]. The prognostic value of those imaging biomarkers needs to be further validated with multicentric prospective studies, to

confirm that baseline imaging without any treatment effect can predict the degree of risk for disease occurrence, or progression and ultimately to implement this patient stratification strategy in therapeutic drug trails.

21.3.3 Verify That the Therapeutic Target Is Present and Accessible

A very direct approach in using imaging to verify therapeutic targets is to image the distribution of the drug itself. Typically, the drug molecule would be synthesized with a positron-emitting isotope (e.g., ^{11}C or ^{18}F) replacing the natural isotope, allowing detection by PET. This approach is common in neuroscience although rare in other diseases. Alternatively, for detection with single-photon emission computed tomography (SPECT), a gamma-emitting moiety may be attached to the drug molecule (albeit with the risk that the drug's properties are thereby altered). An example is radiolabeled derivative of trastuzumab, an antihuman epidermal growth factor receptor 2 (HER2) monoclonal antibody used to treat breast cancer patients with HER2-expressing tumors [22, 23]. Radiolabeled trastuzumab helps to visualize the affinity of the drug in vivo and provides useful information about the PK properties of the drug, such as injected dose versus accumulated drug concentration in the organs and its regional bio-distribution. The noninvasive whole-body imaging overcomes problems associated with biopsies, including sampling errors and discordance of expression between primary tumors and metastases. More importantly, the drug uptake by the target tissue can be quantified at sequential imaging scans, and might provide insight into drug's action at the target tissue and its association with tumor response. Indeed, for the radiotherapeutic drug [^{131}I]-tositumomab, FDA requires that PK be verified by imaging before a therapeutic dose be given [24].

21.3.4 Verify That the Drug Has the Desired Effect on the Pathway and Local Physiology (e.g., Define Dose and Schedule)

Dose escalation and schedule, the main purpose of phase I trials, is usually undertaken to define the maximum tolerated dose (MTD), with the assumption that the most pronounced changes are likely to be detected at the highest dose. But, target saturation may already be reached at lower dose levels, and imaging changes are likely to be apparent and help to define an optimal biological dose. The effects of antiangiogenic therapies on dynamic contrast enhanced (DCE)-MRI have been documented in 39 phase I and II trials with a significant reduction in the transfer constant (K^{trans}) and/or initial area under the gadolinium curve (IAUGC) with multiple agents [25]. In a study of brivanib, a dual tyrosine kinase inhibitor of both vascular endothelial growth factor receptor and fibroblast growth factor receptor, DCE-MRI has been used in several dose schedules and then selected the optimal schedule for a phase II trial [26]. A similar approach has been used with vatalanib [27] and cediranib [28] in advanced cancer, and with sorafenib in renal cancer [29].

21.3.5 Evaluate Efficacy (Response Assessment)

Using imaging biomarkers as surrogate endpoint to assess new drug efficacy has been addressed in the above content. Clinical outcome such as mortality often takes years of follow-up to establish, and the determination of morbidity could be subjective. The use of imaging biomarkers to assess response or progression can reduce sample size, trial duration, and cost and accelerate the introduction of new drugs. A notable example is the approval of etanercept, a tumor necrosis factor inhibitor for the treatment of rheumatoid arthritis. A phase III trial [30] was designed to evaluate the efficacy and safety of

etanercept and methotrexate (the standard of care) in patients with early rheumatoid arthritis, based on two sets of criteria: (1) the American College of Rheumatology scores, which used patient outcome report such as pain and function, in addition to serum C-reactive protein level and (2) imaging evidence of progression, such as joint-space narrowing and erosion. The clinical scoring only showed etanercept a more rapid treatment effect within the first 6 months but was approximately the same thereafter between two groups. Imaging-based erosion score showed significant differences of both immediate and long-term effect. The FDA granted etanercept a conditional marketing authorization. Subsequent studies demonstrated that etanercept achieved sustained improvement compared to methotrexate on both clinical and imaging scores [31]. A conditional approval is a continuum approach that drug developers should continue conducting studies to collect data on the effectiveness of drugs in use after initial approval and eventually provide evidence to keep the drug on the market and obtain a full approval within a predefined timeline.

21.3.6 Predict Response and Resistance (Early Assessment)

Many tyrosine kinase inhibitors and monoclonal antibodies targeting signal transduction inhibitor of tumor growth without tumor regression. Imatinib mesylate is such an example, a tyrosine kinase inhibitor for the treatment of chronic myelogenous leukemia and gastrointestinal stromal tumors (GIST). As opposed to the classical cytotoxic treatment, response rates are low, despite a high percentage of patients having prolonged stable disease and sometimes improvements in survival compared to standard therapies. Acute changes in tumor size are seldom significant in GIST patients treated by imatinib. However, when using [^{18}F]-FDG-PET as a biomarker of tumor metabolism, response could be detected as early as 8 days following the start of treatment and was also associated with a longer PFS [32]. Other examples of early imaging assessment by MRI biomarkers have been reported [33, 34].

The use of functional and molecular imaging to predict the likelihood of response to a particular treatment at early stage is very attractive for drug development. This strategy has been used in early phase trials for proof of concept, although they still await validation to prove their relationships with clinical benefit. In these proof-of-concept trials, drug safety and efficacy will require the confirmation in subsequent trials.

21.3.7 Monitor Safety

Noninvasive imaging biomarkers have a huge potential in monitoring drug safety during clinical trials. In addition to LVEF and BMD, which have achieved surrogate endpoint status, emerging imaging biomarkers can detect drug-induced changes of vital organs and provide region-specific information about tissue abnormality, while serum and urine biomarkers can still be normal due to the functional reserve of the affected organs. The reserve of kidney function can compensate for up to 75% of the loss, so serum and urine biomarkers are insensitive for early renal damage. A wide range of drugs can cause renal papillary necrosis, and the diagnosis tends to be made when irreversible destructive changes have occurred. Contrast-enhanced multiphasic CT can identify early signal changes of renal papillary necrosis and medullary necrosis and at the same time monitor lesion progression and regression. The liver is also a metabolic organ, often suffers from drug toxicity. In some cases, drug-induced hepatic steatosis can lead to a rapid evaluation of severe hepatic failure and ultimately death [35]. Both morphologic and functional liver imaging can display manifestations of hepatotoxicity, easier and earlier than histology techniques.

21.4 Pitfalls of Imaging Biomarker Implementation in Clinical Trials

Although many molecular and functional imaging biomarkers have been explored in clinical trials, few have yet proved adequate to support decision-making in drug development, in regulatory approval, or in clinical practice. Table 21.3 summarizes some common pitfalls of imaging biomarker implementation in trials, and each pitfall is accompanied by proposed solutions.

21.4.1 Poor Study Methodology

Drug development requires methodologically robust and practical clinical trials. A successful imaging-driven trial relies on sound study design, appropriate criteria, optimal timing of observation, and sufficient statistical power. Inappropriate utilization of imaging biomarkers will discredit them in drug development and clinical therapy. To address this issue with imaging evaluations intended to demonstrate new drug efficacy, it is essential to communicate with each regulatory

Table 21.3 Pitfalls of imaging biomarker implementation in clinical trials and proposed solutions

Pitfalls	Possible consequence	Proposed solutions
Poor study methodology		
Insufficient or low-quality preclinical and early phase clinical data	Drug development is wrongly stopped, wrongly continued, or wrong dose/schedule selected	Follow guidelines before study initiation Critical development and review of guidelines by experts from different disciplines (clinicians and imagers)
Inappropriate design selected (e.g., using an investigational imaging biomarker to assess an investigational therapeutic drug)	Unreliable data	Using well-established imaging biomarkers for patient eligibility or response assessment in new therapeutic drug trials Using well-known standard of care treatment as backbone studies to evaluate imaging biomarkers
Inappropriate statistical consideration: low or inaccurate sample size calculation	Underpowered (false negative) or overpowered (unnecessary patient exposure, unnecessary cost)	Sample size calculation based on the objective of the trial and statistical plan should be written before data analysis
Incorrect imaging acquisition or interpretation methods (e.g., miss the right detecting time window, inappropriate criteria used, data misinterpretation)	Data not accepted by regulatory authority	Follow guidelines before study initiation Critical development and review of guidelines by experts from different disciplines (clinicians and imagers)
Lack of standardization		
Missing standardized imaging acquisition protocol	Trial failure due to variation	Imaging acquisition protocol should be developed based on consensus guidelines and previous data It should be delivered to participating centers with adequate training before any patient enrollment
Incompliance with imaging acquisition guidelines	Trial failure due to unreliable data	Good training, quality control, and properly documented and enforced operation procedures
Poor selection of appropriate imaging analysis software	Trial failure due to bias and errors	Software should be tested and validated against test objects

Table 21.3 (continued)

Pitfalls	Possible consequence	Proposed solutions
Variability cross different vendors (low reproducibility)	Impossible to pool data from multicenter trials	Accreditation using scanner calibration by test objects across centers
Variability within the same vendor and same patient (low repeatability)	Trial failure due to unreliable data	Consider those confounders in protocol design The imaging equipment should be accredited throughout the whole trial period
Failure in quality control		
Failure to deliver within required turnaround time	Delay	Ensure adequate manpower and back up team to deliver in time
Inadequate budget for quality control	Trial failure due to unreliable data	Fully forecast the trial cost at the start and regular review the spending
Inappropriately handling imaging data (e.g., alter or delete imaging data by mistake, transfer imaging data with patient private information)	Loss of data or patient privacy	Secured imaging platform for data transmission and analysis, with computer-generated record times and types of action

authority about the trial design and methods before its initiation. FDA have issued specific guidance on "standards for clinical trial imaging endpoint" [36] and "developing medical imaging drug and biological products" [37–39], to encourage imaging biomarker-driven trials to be effective.

21.4.2 Lack of Standardization

Changes in the imaging biomarker due to treatment can only be detected if they are greater than the intrinsic and extrinsic variability of the biomarker in the absence of treatment [40]. Many factors increase the variability of biomarker measurements, as a result of the complexity of data acquisition methods, a number of imaging post-processing procedures with advanced multi-vendor software and the differences of inter- and intra-reader performance. All of these are magnified when imaging biomarkers progress from single-center to multicenter trials.

The standardization of imaging acquisition and analysis is essential to allow comparison and pooling of data among patients in multicenter studies, and more generally in meta-analyses. In this regard, efforts to reduce the difference and increase the repeatability and reproducibility have led to an accreditation program on [^{18}F]-FDG-PET conducted by EANM Research Ltd (EARL) [41]. The National Cancer Institute (NCI) has also developed the Centers for Quantitative Imaging Excellence (CQIE) program [42], in which accredited PET and MR scanners in the NCI-designated Cancer Centers are used to provide reliable quantitative imaging for clinical trials. Consensus papers on how to conduct DCE-MRI and analyze the biomarker have been developed by European and US groups [43, 44]. In Europe, the Innovative Medicines Initiative (IMI) project QuIC-ConCePT [45] is addressing similar issues of harmonization of diffusion-weighted MRI biomarkers, with the aim to provide early readout, robust, and reliable imaging biomarkers for new drug selection in early phase I trials.

21.4.3 Failure in Quality Control

Last but not least, adequate operational support should be in place through the entire conduct of the trial. Resource for performing imaging quality assurance and quality control procedures in a timely manner must be built into the budget. Security is required to protect patient privacy and maintain the integrity of coded data. An imaging

platform with electronic security must employ audit trials for data transmission, storage, analysis, and reporting.

21.5 Perspectives

21.5.1 More Imaging-Driven Trials in Early Drug Development

Pharmaceutical companies spend only 7% on average of their budget on target/mechanism selection and validation [8]. Given the high-failure rate in late-stage drug development, there is an opportunity to improve the selection of targets and the selection of patients most likely to benefit, using inter alia molecular and functional imaging biomarkers. Preclinical imaging studies in rodent tumor models should be designed to assist the design and interpretation of imaging studies in humans. Imaging biomarker-driven phase I/II studies based on a strong biological rationale will help to understand drug mechanism and provide reliable data.

21.5.2 Imaging-Driven Precision Medicine for Early Patient Management

The concept of precision medicine or personalized healthcare is increasingly influencing drug development, aiming to tailor medical treatment to the specific molecular drivers of each patient's disease. In oncology, imaging biomarkers can assess mutation status, while imaging biomarkers of tumor microenvironment which change early (within several days) can identify responders. A reliable early identification through noninvasive imaging techniques will be extremely valuable in guiding patient management and treatment, so as to avoid unnecessary toxicity related to therapy in nonresponsive patients. Nevertheless, very few imaging biomarkers can be used to change treatment strategy at early stage, and imaging-guided patient stratification only occurs in clinical trials. Although many promising findings have been reported, most are retrospective and single-center studies. Large prospective multicenter clinical trials are needed to assess the degree of correlation by comparing a predefined threshold of imaging biomarker change to clinical outcome, so that those qualified early readout biomarkers could be eventually used in clinical practice.

21.5.3 Multi-stakeholder Collaboration and Data Sharing Platform

To improve success, clinical trials require strong collaboration between drug developers and imagers in industry and academia, as well as insight from regulators and payers, utilizing the different strengths of each stakeholder. For example, as a result of the RECIST [46] warehouse, which contains datasets (>10,000 patients) from industry and academic trials on targeted agents, new evidence-based RECIST criteria are being developed, aiming to provide more accurate and objective assessment tools for drug efficacy assessment. The Cancer Imaging Archive [47] provides a large achieve of medical images, clinical, and genomic data of cancer accessible to public. Similar platform is also available with neuroimaging in Alzheimer's disease [48]. Warehouses with good quality data can not only help power analysis, increase data accuracy, and transparency but also, in some cases, allow future imaging biomarker discovery in an efficient and timely manner. However it must be recognized that most new imaging biomarkers require novel tracers or novel acquisition, and in such cases, archived data cannot substitute for new imaging biomarker validation trials.

References

1. Food and Drug Administration. FDA official web. http://www.fda.gov/.
2. European Medicines Agency. EMA offical web. http://www.ema.europa.eu/ema/.
3. China Food and Drug Administration. CFDA official web. http://eng.cfda.gov.cn/WS03/CL0755/.
4. Ministry of Health Labour Welfare. MHLW official web. http://www.mhlw.go.jp/english/.

5. International Conference on Harmonization. ICH official web. http://www.ich.org/home.html.
6. WMA. WMA declaration of Helsinki – ethical principles for medical research involving human subjects. 2013. http://www.wma.net/en/30publications/10policies/b3/.
7. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11:191–200. doi:10.1038/nrd3681.
8. PWC. From vision to decision pharma 2020. 2013. https://www.pwc.com/jp/ja/japan-knowledge/archive/assets/pdf/pharma-2020-vision-to-decision.pdf.
9. https://report.nih.gov/investigators_and_trainees/acd_bwf/pdf/Phrma_Industry_Profile_2011.pdf
10. Bria E, Di Maio M, Carlini P, et al. Targeting targeted agents: open issues for clinical trial design. J Exp Clin Cancer Res. 2009;28:66. doi:10.1186/1756-9966-28-66.
11. Atkinson AJ, Colburn WA, DeGruttola VG, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89–95. doi:10.1067/mcp.2001.113989.
12. FDA. Novel drugs approved using surrogate endpoints. 2015. http://www.fda.gov/downloads/NewsEvents/Testimony/UCM445375.pdf.
13. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. N Engl J Med. 1991;324(12):781–8. doi:10.1056/NEJM199103213241201.
14. Wilson WH, Schenkein DP, Jernigan CL, Woodcock J, Schilsky RL. Reevaluating the accelerated approval process for oncology drugs. Clin Cancer Res. 2013;19:2804–9. doi:10.1158/1078-0432.CCR-13-0315.
15. FDA. Guidance for industry clinical trial endpoints for the approval of cancer drugs and biologics. 2007. http://www.fda.gov/downloads/Drugs/…/Guidances/ucm071590.pdf.
16. Lancellotti P, Nkomo VT, Badano LP, et al. Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy. J Am Soc Echocardiogr. 2013;26:1013–32. doi:10.1016/j.echo.2013.07.005.
17. Adebahr S, Collette S, Shash E, et al. LungTech, an EORTC phase II trial of stereotactic body radiotherapy for centrally located lung tumours – a clinical perspective. Br J Radiol. 2015;88:20150036. doi:10.1259/bjr.20150036.
18. Von Pawel J, Von Roemeling R, Gatzemeier U, et al. Tirapazamine plus cisplatin versus cisplatin in advanced non-small-cell lung cancer: a report of the international CATAPULT I study group. J Clin Oncol. 2000;18:1351–9. http://www.ncbi.nlm.nih.gov/pubmed/10715308
19. Rischin D, Hicks RJ, Fisher R, et al. Prognostic significance of [18F]-misonidazole positron emission tomography-detected tumor hypoxia in patients with advanced head and neck cancer randomly assigned to chemoradiation with or without tirapazamine: a substudy of Trans-Tasman Radiation Oncology Group Study. J Clin Oncol. 2006;24(13):2098–104. doi:10.1200/JCO.2005.05.2878.
20. Morris RT, Joyrich RN, Naumann RW, et al. Phase II study of treatment of advanced ovarian cancer with folate-receptor-targeted therapeutic (vintafolide) and companion SPECT-based imaging agent (99mTc-etarfolatide). Ann Oncol. 2014;25(4):852–8. doi:10.1093/annonc/mdu024.
21. Na II, Byun BH, Kang HJ, et al. 18F-fluoro-2-deoxyglucose uptake predicts clinical outcome in patients with gefitinib-treated non-small cell lung cancer. Clin Cancer Res. 2008;14:2036–41. doi:10.1158/1078-0432.CCR-07-4074.
22. Dijkers EC, Oude Munnink TH, Kosterink JG, et al. Biodistribution of 89Zr-trastuzumab and PET imaging of HER2-positive lesions in patients with metastatic breast cancer. Clin Pharmacol Ther. 2010;87:586–92. doi:10.1038/clpt.2010.12.
23. Chang AJ, DeSilva R, Jain S, Lears K, Rogers B, Lapi S. 89Zr-radiolabeled trastuzumab imaging in orthotopic and metastatic breast tumors. Pharmaceuticals. 2012;5:79–93. doi:10.3390/ph5010079.
24. FDA. Prescribing information (tositumomab and Iodine I 131 Tositumomab). 2003. http://www.accessdata.fda.gov/drugsatfda_docs/label/2012/125011s102lbl.pdf.
25. Zweifel M, Padhani AR. Perfusion MRI in the early clinical development of antivascular drugs: decorations or decision making tools? Eur J Nucl Med Mol Imaging. 2010;37 Suppl 1:S164–82. doi:10.1007/s00259-010-1451-z.
26. Jonker DJ, Rosen LS, Sawyer MB, et al. A phase I study to determine the safety, pharmacokinetics and pharmacodynamics of a dual VEGFR and FGFR inhibitor, brivanib, in patients with advanced or metastatic solid tumors. Ann Oncol. 2011;22:1413–9. doi:10.1093/annonc/mdq599.
27. Thomas AL, Morgan B, Horsfield MA, et al. Phase I study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of PTK787/ZK 222584 administered twice daily in patients with advanced cancer. J Clin Oncol. 2005;23(18):4162–71. doi:10.1200/JCO.2005.09.034.
28. Drevs J, Siegert P, Medinger M, et al. Phase I clinical study of AZD2171, an oral vascular endothelial growth factor signaling inhibitor, in patients with advanced solid tumors. J Clin Oncol. 2007;25(21):3045–54. doi:10.1200/JCO.2006.07.2066.
29. Hahn OM, Yang C, Medved M, et al. Dynamic contrast-enhanced magnetic resonance imaging pharmacodynamic biomarker study of sorafenib in metastatic renal carcinoma. J Clin Oncol. 2008;26:4572–8. doi:10.1200/JCO.2007.15.5655.
30. Bathon JM, Martin RW, Fleischmann RM, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. N

Engl J Med. 2000;343:1586–93. doi:10.1056/nejm200011303432201.
31. Genovese MC, Bathon JM, Martin RW, et al. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. Arthritis Rheum. 2002;46:1443–50. doi:10.1002/art.10308.
32. Stroobants S, Goeminne J, Seegers M, et al. 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec®). Eur J Cancer. 2003;39:2012–20. doi:10.1016/S0959-8049(03)00073-X.
33. Sinkus R, Van Beers BE, Vilgrain V, DeSouza N, Waterton JC. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. Eur J Cancer. 2012;48:425–31. doi:10.1016/j.ejca.2011.11.034.
34. O'Connor JPB, Jackson A, Parker GJM, Jayson GC. DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. Br J Cancer. 2007;96:189–95. doi:10.1038/sj.bjc.6603515.
35. Brunt EM. Nonalcoholic steatohepatitis. Semin Liver Dis. 2004;24:3–20. doi:10.1055/s-2004-823098.
36. FDA. Guidance for industry standards for clinical trial imaging endpoints guidance for industry standards for clinical trial imaging endpoints. 2015. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm268555.pdf.
37. FDA. Guidance for industry developing medical imaging drug and biological products part 2: clinical indications. 2004. http://www.gmp-compliance.org/guidemgr/files/MEDIMAGEIND.PDF.
38. FDA. Guidance for industry developing medical imaging drug and biological products part 1: conducting safety assessments. 2004. http://www.fda.gov/downloads/Drugs/…/Guidances/ucm071600.pdf.
39. FDA. Developing medical imaging drug and biological products part 3: design, analysis, and interpretation of clinical studies. 2004. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071604.pdf.
40. Waterton JC, Pylkkanen L. Qualification of imaging biomarkers for oncology drug development. Eur J Cancer. 2012;48:409–15. doi:10.1016/j.ejca.2011.11.037.
41. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging. 2010;37(1):181–200. doi:10.1007/s00259-009-1297-4.
42. NCI. NCI-CQIE qualification materials. http://www.acrin.org/CORELABS/NCICQIEQUALIFICATIONPROGRAM/SITEQUALIFICATIONMATERIALS.aspx.
43. QIBA. https://www.rsna.org/QIBA_Protocols_and_Profiles.aspx.
44. Leach MO, Morgan B, Tofts PS, et al. Imaging vascular function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging. Eur Radiol. 2012;22(7):1451–64. doi:10.1007/s00330-012-2446-x.
45. QuIC-ConCePT. http://www.quic-concept.eu/.
46. RECIST. RECIST web. http://www.eortc.org/recist/.
47. Achieve cancer imaging. Cancer imaging archive. http://www.cancerimagingarchive.net.
48. Alzheimer disease neuroimaging initiative. ADNI web. http://adni.loni.usc.edu/data-samples/access-data/.
49. Chu TF, Rupnick MA, Kerkela R, et al. Cardiotoxicity associated with tyrosine kinase inhibitor sunitinib. Lancet. 2007;370:2011–9. doi:10.1016/S0140-6736(07)61865-0.
50. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228–47. doi:10.1016/j.ejca.2008.10.026.
51. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis. 2010;30:52–60. doi:10.1055/s-0030-1247132.
52. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. Ann Oncol. 2004;15:257–60. doi:10.1093/annonc/mdh059.
53. Salloway S, Sperling R, Gilman S, et al. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. Neurology. 2009;73:2061–70. doi:10.1212/WNL.0b013e3181c67808.
54. Liu Y, Litière S, de Vries EG, et al. The role of response evaluation criteria in solid tumour in anticancer treatment evaluation: results of a survey in the oncology community. Eur J Cancer. 2014;50(2):260–6. doi:10.1016/j.ejca.2013.10.011.
55. Hunter DJ, Zhang W, Conaghan PG, et al. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. Osteoarthritis Cartilage. 2011;19:557–88. doi:10.1016/j.joca.2010.10.029.

Index

L. Martí-Bonmatí, A. Alberich-Bayarri (eds.), *Imaging Biomarkers*,
DOI 10.1007/978-3-319-43504-6