

Issam El Naqa

Abstract

Radiotherapy treatment outcomes are determined by complex interactions among treatment, anatomical, and patient-related variables. A key component of radiation oncology research is to predict at the time of treatment planning, or during the course of fractionated radiation treatment, the probability of tumor eradication and normal tissue risks for the type of treatment being considered for that particular patient. Traditionally, these outcomes are modeled using information about the dose distribution and the fractionation. However, it is recognized that radiation response is multifactorial including clinical prognostic factors and, more recently, inherited genetic variations have been suggested as playing an important role in radiation response. Therefore, recent approaches have utilized increasingly data-driven models incorporating advanced bioinformatics and machine learning tools in which dose-volume metrics are mixed with other patient- or disease-based prognostic factors in order to improve outcomes prediction. Accurate prediction of treatment outcomes would provide clinicians with better tools for informed decision-making about expected benefits versus anticipated risks. In this chapter, we provide an overview of the current status of data-driven outcome modeling techniques for patients who receive radiation treatment with special focus on its big data notion and the emerging role of machine learning approaches to improve outcome modeling and response prediction.

I. El Naqa

Department of Oncology, McGill University, Montreal, QC, Canada

Department of Radiation Oncology, University of Michigan, Ann Arbor, USA

e-mail: issam.elnaqa@mcgill.ca; ielnaqa@med.umich.edu

© Springer International Publishing Switzerland 2015

I. El Naqa et al. (eds.), *Machine Learning in Radiation Oncology: Theory and Applications*, DOI 10.1007/978-3-319-18305-3_16

263

16.1 Introduction

Recent years have witnessed tremendous technological advances in radiotherapy treatment planning, image guidance, and treatment delivery [1, 2]. Moreover, clinical trials examining treatment intensification in patients with locally advanced cancer have shown incremental improvements in local control and overall survival [3]. However, radiation-induced toxicities remain major dose-limiting factors [4, 5]. Therefore, there is a need for studies directed toward predicting treatment benefit versus risk of failure. Clinically, such predictors would allow for more individualization of radiation treatment plans. In other words, physicians may prescribe a more or less intense radiation regimen for an individual based on model predictions of local control benefit and toxicity risk. Such an individualized regimen would aim toward an optimized radiation treatment response while keeping in mind that a more aggressive treatment with a promised improved tumor control will not translate into improved survival unless severe toxicities are accounted for and limited during treatment planning. Therefore, improved models for predicting both local control and normal tissue toxicity should be considered in the optimal treatment planning design process.

Radiotherapy outcomes are usually characterized by two metrics: the tumor control probability (TCP) and the normal tissue complication probability (NTCP) of surrounding normal tissues [2, 6]. TCP/NTCP models could be used during the consultation period as a guide for ranking treatment options [7, 8]. Alternatively, once a decision has been reached, these models could be included in an objective function, and the optimization problem driving the actual patient's treatment plan can be formulated in terms relevant to maximizing tumor eradication benefit and minimizing complication risk [9–11]. Traditional models of TCP/NTCP models and their variations use information only about the dose distribution and fractionation. However, it is well known that radiotherapy outcomes may also be affected by multiple clinical and biological prognostic factors such as stage, volume, tumor hypoxia, etc. [12, 13] as depicted in Fig. 16.1. Therefore, recent years have witnessed the emergence of data-driven models utilizing informatics techniques, in

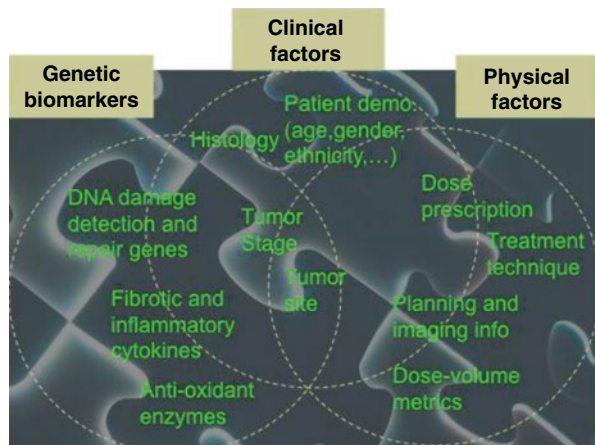


Fig. 16.1 Radiotherapy treatment involves complex interaction of physical, biological, and clinical factors. The successful informatics approach should be able to resolve this interaction “puzzle” in the observed treatment outcome (e.g., local control or toxicity) for each individual patient [21]

which dose-volume metrics are combined with other patient- or disease-based prognostic factors [4, 5, 14–20].

In this chapter, we provide an overview of the current status of data-driven outcome modeling techniques for predicting tumor response and normal tissue toxicities for patients who receive radiation treatment with special focus on the emerging role of machine learning approaches to improve outcome modeling and response prediction. Then, we present examples of radiotherapy data and its big data notion. Finally, we discuss the potentials and challenging obstacles to applying bioinformatics and machine learning strategies to radiotherapy outcome modeling.

16.2 Data-Driven Outcome Modeling

Radiotherapy outcome models could be divided according to the underlying principle into (1) analytical models, which employ biophysical understanding of irradiation effects such as the linear quadratic (LQ) model, and (2) data-driven models, which are phenomenological models and depend on parameters available from the collected clinical and dosimetric data [20]. In the context of data-driven and multi-variable modeling of outcomes, the observed treatment outcome (e.g., TCP or NTCP) is considered as the result of mathematical mapping of several dosimetric, clinical, or biological input variables [19]. Mathematically this is expressed as: $f(\mathbf{x}; \mathbf{w}^*): X \rightarrow Y$ where R^N (an input variable vector of N dimensions) is composed of the input metrics (dose-volume metrics, patient disease specific prognostic factors, or biological markers). The expression $y_i \in Y$ is the corresponding observed treatment outcome scalar. The variable \mathbf{w}^* includes the optimal parameters of model $f(\cdot)$ obtained by optimizing a certain objective functional. Learning is defined in this context as estimating dependencies from data [22]. The two common types of learning could be applied: supervised and unsupervised. Supervised learning is used when the endpoints of the treatments such as tumor control or toxicity grade are known; these endpoints are provided by experienced oncologists following RTOG or NCI criteria, and it is the most commonly used learning method in outcome modeling. Nevertheless, unsupervised methods such as principal component analysis (PCA) are also used to reduce dimensionality and to aid visualization of multivariate data and selection of learning method parameters [23]. The selection of the functional form of the model $f(\cdot)$ is closely related to the prior knowledge of the problem. In analytical models, the shape of the functional form is selected based on the clinical or biological process at hand; however, in data-driven models, the objective is usually to find a functional form that fits the data [24].

16.3 Radiotherapy as a Big Data Resource

A typical radiotherapy treatment scenario can generate a large pool of “big data” that comprise but are not limited to patient demographics, volumetric dosimetric data about radiation exposure to the tumor and surrounding tissues, and 3D and 4D

anatomical and functional disease longitudinal imaging features (radiomics), in addition to genomics and proteomics data derived from peripheral blood and tissue specimens. Accordingly, big data in radiotherapy could be divided based on its nature into four categories: clinical, dosimetric, imaging, and biological. These four categories of radiotherapy big data are described in the following.

16.3.1 Clinical Data

Clinical data in radiotherapy typically refers to cancer diagnostic information (site, histology, stage, grade, etc.) and patient-related characteristics (age, gender, comorbidities, etc.). In some instances, other treatment modalities information (surgery, chemotherapy, hormonal treatment, etc.) would be also classified under this category. The mining of such data could be challenging if the data is unstructured; however, there are good opportunities for natural language processing (NLP) techniques to assist in the organization of data [25].

16.3.2 Dosimetric Data

This type of data is related to the treatment planning process in radiotherapy, which involves radiation dose simulation using computed tomography imaging, specifically dose-volume metrics derived from dose-volume histograms (DVHs) graphs. Dose-volume metrics have been extensively studied in the radiation oncology literature for outcome modeling [14–17, 26, 27]. These metrics are extracted from the DVH such as volume receiving certain dose (V_x); minimum dose to $x\%$ volume (D_x); mean, maximum, and minimum dose; etc. More details are in our review chapter [20]. Moreover, we have developed a dedicated software tool called “Dose response explorer” (DREES) for deriving these metrics and modeling of radiotherapy response [28].

16.3.3 Radiomics (Imaging Features)

kV x-ray computed tomography (kV-CT) has been historically considered the standard modality for treatment planning in radiotherapy because of its ability to provide electron density information for target definition, structures, and heterogeneous dose calculations [2, 29]. However, additional information from other imaging modalities could be used to improve treatment monitoring and prognosis in different cancer types. For example, physiological information (tumor metabolism, proliferation, necrosis, hypoxic regions, etc.) can be collected directly from nuclear imaging modalities such as SPECT and PET or indirectly from MRI [30, 31]. The complementary nature of these different imaging modalities has led to efforts toward combining information to achieve better treatment outcomes. For instance,

PET/CT has been utilized for staging, planning, and assessment of response to radiation therapy [32, 33]. Similarly, MRI has been applied in tumor delineation and assessing toxicities in head and neck cancers [34, 35]. Moreover, quantitative information from hybrid-imaging modalities could be related to biological and clinical endpoints, a new emerging field referred to as “radiomics” [36, 37]. In our previous work, we demonstrated the potential of this new field to monitor and predict response to radiotherapy in head and neck [38], cervix [38, 39], and lung [40] cancers, in turn allowing for adapting and individualizing treatment.

16.3.4 Biological Markers

A biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention” [41]. Biomarkers can be categorized based on the biochemical source of the marker into exogenous or endogenous. Exogenous biomarkers are based on introducing a foreign substance into the patient’s body such as those used in molecular imaging as discussed above. Conversely, endogenous biomarkers can further be classified as (1) “expression biomarkers,” measuring changes in gene expression or protein levels, or (2) “genetic biomarkers,” based on variations, for tumors or normal tissues, in the underlying DNA genetic code. Measurements are typically based on tissue or fluid specimens, which are analyzed using molecular biology laboratory techniques [42]. Expression biomarkers are the result of gene expression changes in tissues or bodily fluids due to the disease or normal tissues’ response to treatment [43]. These biomarkers can be further divided into single parameter (e.g., prostate-specific antigen (PSA) levels in blood serum) versus bio-arrays. These can be based on disease pathophysiology or pharmacogenetic studies or they can be extracted from several methods, such as high-throughput gene expression (aka transcriptomics) [44–46], resulting protein expressions (aka proteomics) [47, 48], or metabolites (aka metabolomics) [49, 50]. On the other hand, the inherent genetic variability of the human genome is an emerging resource for studying disposition to cancer and the variability of patient responses to therapeutic agents. These variations in the DNA sequences of humans, in particular single-nucleotide polymorphisms (SNPs), have strong potential to elucidate complex disease onset and response in cancer [51]. Methods based on the candidate gene approach and high throughput (genome-wide associations (GWAS) studies) are currently heavily investigated to analyze the functional effect of SNPs in predicting response to radiotherapy [52–54]. There are several ongoing SNP genotyping initiatives in radiation oncology, including the pan-European GENEPI project [55], the British RAPPER project [56], the Japanese RadGenomics project [57], and the US Gene-PARE project [58]. An international consortium has been also established to coordinate and lead efforts in this area [59]. Examples include the identification of SNPs related to radiation toxicity in prostate cancer treatment [60–62].

16.4 Systems Radiobiology

To integrate heterogeneous big data in radiotherapy, engineering-inspired system approaches would have great potential to achieve this goal. Systems biology has emerged as a new field to apply systematic study of complex interactions to biological systems [63], but its application to radiation oncology, despite this potential, has been unfortunately limited to date [64, 65]. Recently, Eschrich et al. presented systems biology approach for identifying biomarkers related to radiosensitivity in different cancer cell lines using linear regression to correlate gene expression with survival fraction measurements [66]. However, such a linear regression model may lack the ability to account for higher-order interactions among the different genes and neglect the expected hierarchal relationships in signaling transduction of highly complex radiation response. It has been noted in the literature that modeling of molecular interactions could be represented using graphs of network connections as in power line grids. In this case, radiobiological data can be represented as a graph (network) where the nodes represent genes or proteins and the edges may represent similarities or interactions between these nodes. We have utilized such approach based on Bayesian networks for modeling dosimetric radiation pneumonitis relationships [67] and more recently in predicting local control from biological and dosimetric data [68].

In the more general realm of bioinformatics, this systems approach could be represented as a part of a feedback treatment planning system as shown in Fig. 16.2,

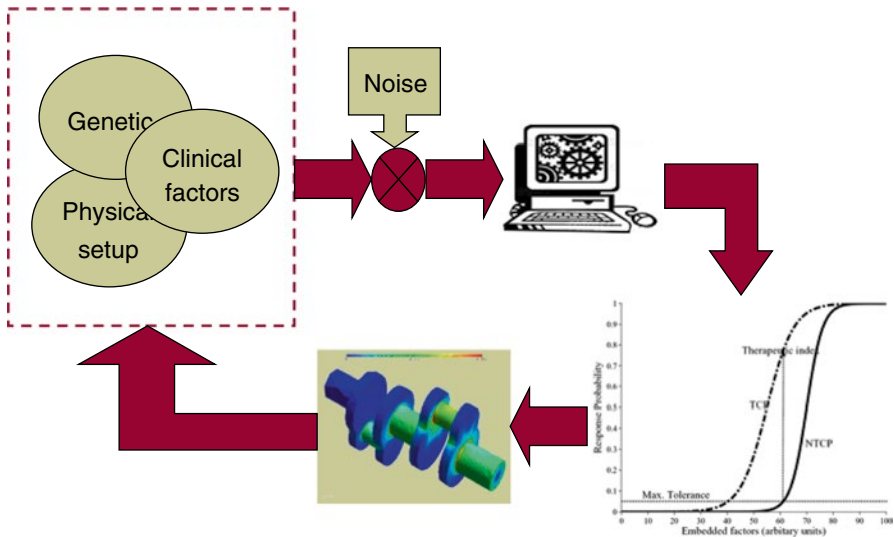


Fig. 16.2 The bioinformatics understanding of heterogeneous variable interactions as a feedback into the treatment planning system to improve patient's outcomes. A heterogeneous list of variables with their noisy characteristics are acquired from retrospective or prospective studies and fed into a learning algorithm to derive estimates of TCP/NTCP, which is typically corrected based on feedback of newly tested patients or scientific and clinical findings

in which bioinformatics understanding of heterogeneous variables interactions could be used as an adaptive learning process to improve outcome modeling and personalization of radiotherapy regimens.

16.5 Software Tools for Outcome Modeling

Many of the TCP/NTCP outcome modeling methods require dedicated software tools for implementation. Examples of such software tools in the literature are BIOPLAN and DREES. BIOPLAN (BIOlogical evaluation of treatment PLANs) uses several analytical models for evaluation of radiotherapy treatment plans [69], while DREES is an open-source software package developed by our group for dose-response modeling using analytical and data-driven methods [28] presented in Fig. 16.3. It should be mentioned that several commercial treatment planning systems have currently incorporated different TCP/NTCP models, mainly analytical ones that could be used for ranking and biological optimization purposes. A discussion of these models and their quality assurance guidelines is provided in TG-166 [11].

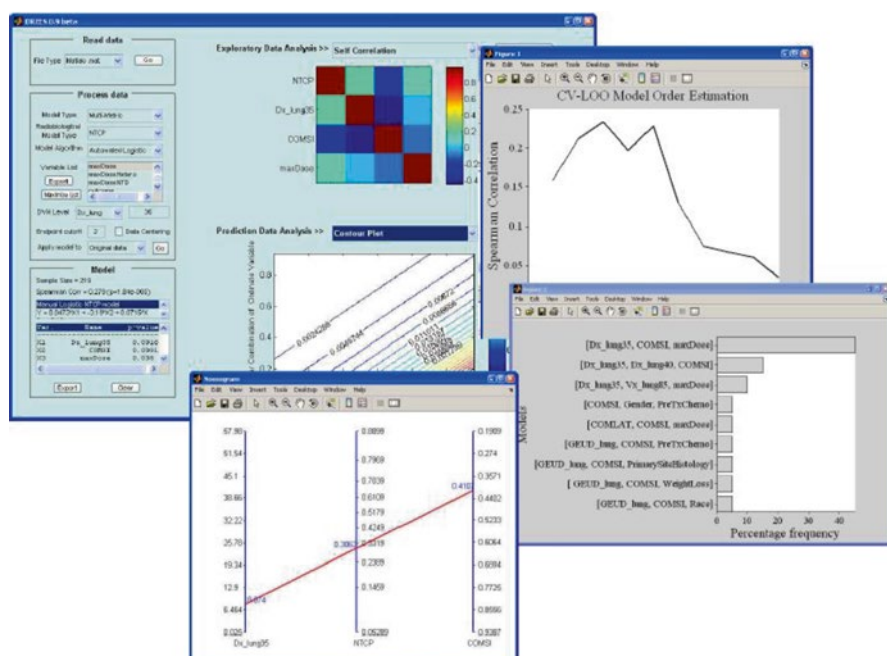


Fig. 16.3 DREES allows for TCP/NTCP analytical and multivariate modeling of outcomes data. The example is for lung injury. The components shown here are Main GUI, model order and parameter selection by resampling methods, and a nomogram of outcome as function of mean dose and location

16.6 Discussion

16.6.1 Data Sharing

Successful outcome modeling requires large datasets to meet statistical requirements, and sharing data is necessary to achieve this purpose. However, data sharing remains an issue for nontechnical issues [70]. Therefore, the Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) consortium has suggested that cooperative groups adopt a policy of anonymizing clinical trial data and making these data publicly accessible after a reasonable delay. This delay would enable publication of all the investigator-driven, planned studies while encouraging the establishment of key databanks of linked treatment planning, imaging, and outcomes data [71]. An alternative approach is to apply rapid learning as suggested by the Maastricht clinic group at Maastricht, in which innovative information technologies are developed that support semantic interoperability and enable distributed learning and data sharing without the need for the data to leave the hospital or the institution [72]. An example of multi-institutional data sharing is developed by the groups of Maastricht clinic and the Policlinico Universitario Agostino Gemelli in Rome, Italy (Gemelli) [73].

16.6.2 Lack of Web Resources for Radiobiology

As of today, there are no dedicated web resources for bioinformatics studies in radiation oncology. Nevertheless, radiotherapy biological marker studies can still benefit from existing bioinformatics resources for pharmacogenomic studies that contain databases and tools for genomic, proteomic, and functional analysis as reviewed by Yan [74]. For example, the National Center for Biotechnology Information (NCBI) site hosts databases such as GenBank, dbSNP, Online Mendelian Inheritance in Man (OMIM), and genetic search tools such as BLAST. In addition, the Protein Data Bank (PDB) and the program CPHmodels are useful for protein structure three-dimensional modeling. The Human Genome Variation Database (HGVD) contains information on physical and functional relationships between sequence variations and neighboring genes. Pattern analysis using PROSITE and Pfam databases can help correlate sequence structures to functional motifs such as phosphorylation [74]. Biological pathway construction and analysis is an emerging field in computational biology that aims to bridge the gap between biomarker findings in clinical studies with underlying biological processes. Several public databases and tools are being established for annotating and storing known pathways such as KEGG and Reactome projects or commercial ones such as the IPA or MetaCore [75]. Statistical tools are used to properly map data from gene/protein differential experiments into the different pathways such as mixed effect models [76] or enrichment analysis [77].

16.6.3 Protecting the Confidentiality and Privacy of Clinical Phenotype Data

QUANTEC offered a solution to radiotherapy digital data (treatment planning, imaging, and outcomes data) accessibility by asking cooperative groups to adopt a policy of anonymizing clinical trial data and making the data publicly accessible after a reasonable delay [71]. With regard to blood or tissue samples, no recommendation was made, however, by extending the same work and making any gene or protein expression assay measurements available under the same umbrella, while raw specimen data could be accessed from the biospecimen resource. For example, in the RTOG biospecimen standard operating procedure (SOP), it is highlighted that biospecimens received by the RTOG Biospecimen Resource are de-identified of all patient health identifiers and are enrolled in an approved RTOG study. Each patient being enrolled by an institution has to qualify and consent to be part of the study before being assigned a case and study ID by the RTOG Statistical Center. No information containing specific patient health identifiers is maintained by the Resource Freezerworks database, which is primarily an inventory and tracking system. In addition, information related to medical identifiers and any code lists could be removed completely from the dataset after a certain period say 10 years or so. Moreover, it has been argued that current measures by the Health Insurance Portability and Accountability Act (HIPPA) of 18 data elements are not sufficient and techniques based on research in privacy-preserving data mining, disclosure risk assessment data de-identification, obfuscation, and protection may need to be adopted to achieve better protection of confidentiality [78].

16.7 Future Research Directions

The ability to maintain high-fidelity large-scale data for radiotherapy studies remains a major challenge despite the high volume of clinical generated data on almost daily basis. As discussed above there have been several ongoing institutional and multi-institutional initiatives such as the RTOG, radiogenomics consortium, and EuroCAT to develop such infrastructure; however, there is plenty of work to be done to overcome issues related to, data sharing hurdles, patient confidentiality issues lack of signaling pathways databases of radiation response, development of cost-effective multicenter communication systems that allows transmission, storage, and query of large datasets such images, dosimetry, and biomarkers information. The use of NLP techniques is a promising approach in organizing unstructured clinical data. Dosimetry and imaging data can benefit from existing infrastructure for Picture Archiving and Communication Systems (PACS) or other medical image databases. Methods based on the new emerging field of systems radiobiology will continue to grow on a rapid pace, but they could also benefit immensely from the development of specialized radiation response signaling pathway databases analogous to the currently existing pharmacogenomics databases. Data sharing among

different institutions is a major hurdle, which could be solved through cooperative groups or distributed databases by developing in a cost-effective manner the necessary bioinformatics and communication infrastructure using open-access resources through partnership with industry.

Conclusion

Recent evolution in radiotherapy imaging and biotechnology has generated enormous amount of big data that spans clinical, dosimetric, imaging, and biological markers. This data provided new opportunities for reshaping our understanding of radiotherapy response and outcome modeling. However, the complexity of this data and the variability of tumor and normal tissue responses would render the utilization of advanced bioinformatics and machine learning methods as indispensable tools for better delineation of radiation complex interaction mechanisms and basically a cornerstone to “making data dreams come true” [79]. However, it also posed new challenges for data aggregation, sharing, confidentiality, and analysis. Moreover, radiotherapy data constitutes a unique interface between physics and biology that can benefit from the general advances in biomedical informatics research such as systems biology and available web resources while still requiring the development of its own technologies to address specific issues related to this interface. Successful application and development of advanced data communication and bioinformatics tools for radiation oncology big data so to speak is essential to better predicting radiotherapy response to accompany other aforementioned technologies and usher significant progress toward the goal of personalized treatment planning and improving the quality of life for radiotherapy cancer patients.

References

1. Bortfeld T, Schmidt-Ullrich R, De Neve W, Wazer D, editors. Image-guided IMRT. Berlin: Springer; 2006.
2. Webb S. The physics of three-dimensional radiation therapy: conformal radiotherapy, radio-surgery, and treatment planning. Bristol, UK. Philadelphia: Institute of Physics Pub; 2001.
3. Halperin EC, Perez CA, Brady LW. Perez and Brady's principles and practice of radiation oncology. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
4. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys.* 2010;76:S3–9.
5. Jackson A, Marks LB, Bentzen SM, Eisbruch A, Yorke ED, Ten Haken RK, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome. *Int J Radiat Oncol Biol Phys.* 2010;76:S155–60.
6. Steel GG. Basic clinical radiobiology. 3rd ed. London/New York: Arnold/Oxford University Press; 2002.
7. Armstrong K, Weber B, Ubel PA, Peters N, Holmes J, Schwartz JS. Individualized survival curves improve satisfaction with cancer risk management decisions in women with BRCA1/2 mutations. *J Clin Oncol.* 2005;23:9319–28.

8. Weinstein MC, Toy EL, Sandberg EA, Neumann PJ, Evans JS, Kuntz KM, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health*. 2001;4:348–61.
9. Moiseenko V, Kron T, Van Dyk J. Biologically-based treatment plan optimization: a systematic comparison of NTCP models for tomotherapy treatment plans. In: Proceedings of the 14th international conference on the use of computers in radiation therapy, Seoul. 2004.
10. Brahme A. Optimized radiation therapy based on radiobiological objectives. *Semin Radiat Oncol*. 1999;9:35–47.
11. Allen Li X, Alber M, Deasy JO, Jackson A, Ken Jee KW, Marks LB, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM. *Med Phys*. 2012;39:1386–409.
12. Choi N, Baumann M, Flentje M, Kellokumpu-Lehtinen P, Senan S, Zamboglou N, et al. Predictive factors in radiotherapy for non-small cell lung cancer: present status. *Lung Cancer*. 2001;31:43–56.
13. Fu XL, Zhu XZ, Shi DR, Xiu LZ, Wang LJ, Zhao S, et al. Study of prognostic predictors for non-small cell lung cancer. *Lung Cancer*. 1999;23:143–52.
14. Blanco AI, Chao KS, El Naqa I, Franklin GE, Zakarian K, Vicic M, et al. Dose-volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy. *Int J Radiat Oncol Biol Phys*. 2005;62:1055–69.
15. Bradley J, Deasy JO, Bentzen S, El-Naqa I. Dosimetric correlates for acute esophagitis in patients treated with radiotherapy for lung carcinoma. *Int J Radiat Oncol Biol Phys*. 2004;58:1106–13.
16. Marks LB. Dosimetric predictors of radiation-induced lung injury. *Int J Radiat Oncol Biol Phys*. 2002;54:313–6.
17. Hope AJ, Lindsay PE, El Naqa I, Bradley JD, Vicic M, Deasy JO. Clinical, dosimetric, and location-related factors to predict local control in non-small cell lung cancer. In: *ASTRO 47th annual meeting*. Denver. 2005. p. S231.
18. Tucker SL, Cheung R, Dong L, Liu HH, Thames HD, Huang EH, et al. Dose-volume response analyses of late rectal bleeding after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys*. 2004;59:353–65.
19. El Naqa I, Bradley JD, Lindsay PE, Blanco AI, Vicic M, Hope AJ, et al. Multi-variable modeling of radiotherapy outcomes including dose-volume and clinical factors. *Int J Radiat Oncol Biol Phys*. 2006;64:1275–86.
20. Deasy JO, El Naqa I. Image-based modeling of normal tissue complication probability for radiation therapy. *Cancer Treat Res*. 2008;139:215–56.
21. Spencer S, Bonnin DA, Deasy J, Bradley JD, El Naqa I. *Bioinformatics methods for learning radiation-induced lung inflammation from heterogeneous retrospective and prospective data*. New York: Hindawi Publishing Corporation; 2009.
22. Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning data mining, inference, and prediction: with 200 full-color illustrations*. New York: Springer; 2001.
23. Härdle W, Simar L. *Applied multivariate statistical analysis*. Berlin/New York: Springer; 2003.
24. El Naqa I, Bradley J, Lindsay PE, Hope A, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. *Phys Med Biol*. 2009;54:S9–30.
25. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc*. 2013;21(2):221–30.
26. Hope AJ, Lindsay PE, El Naqa I, Alaly JR, Vicic M, Bradley JD, et al. Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters. *Int J Radiat Oncol Biol Phys*. 2006;65:112–24.
27. Levegrun S, Jackson A, Zelefsky MJ, Skwarchuk MW, Venkatraman ES, Schlegel W, et al. Fitting tumor control probability models to biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer: pitfalls in deducing radiobiologic parameters for tumors from clinical data. *Int J Radiat Oncol Biol Phys*. 2001;51:1064–80.

28. El Naqa I, Suneja G, Lindsay PE, Hope AJ, Alaly JR, Vicic M, et al. Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. *Phys Med Biol*. 2006;51:5719–35.
29. Khan FM. *Treatment planning in radiation oncology*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2007.
30. Condeelis J, Weissleder R. In vivo imaging in cancer. *Cold Spring Harb Perspect Biol*. 2010;2:a003848.
31. Willmann JK, van Bruggen N, Dinkelborg LM, Gambhir SS. Molecular imaging in drug development. *Nat Rev Drug Discov*. 2008;7:591–607.
32. Bussink J, Kaanders JHAM, van der Graaf WTA, Oyen WJG. PET-CT for radiotherapy treatment planning and response monitoring in solid tumors. *Nat Rev Clin Oncol*. 2011;8:233–42.
33. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging*. 2010;37:2165–87.
34. Newbold K, Partridge M, Cook G, Sohaib SA, Charles-Edwards E, Rhys-Evans P, et al. Advanced imaging applied to radiotherapy planning in head and neck cancer: a clinical review. *Br J Radiol*. 2006;79:554–61.
35. Piet D, De Frederik K, Vincent V, Sigrid S, Robert H, Sandra N. Diffusion-weighted magnetic resonance imaging to evaluate major salivary gland function before and after radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;71(5):1365–71.
36. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
37. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30:1234–48.
38. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recogn*. 2009;42:1162–71.
39. Kidd EA, El Naqa I, Siegel BA, Dehdashti F, Grigsby PW. FDG-PET-based prognostic nomograms for locally advanced cervical cancer. *Gynecol Oncol*. 2012;127:136–40.
40. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol*. 2012;102:239–45.
41. Group BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69:89–95.
42. El Naqa I, Craft J, Oh J, Deasy J. Biomarkers for early radiation response for adaptive radiation therapy. In: Li XA, editor. *Adaptive radiation therapy*. Boca Raton: Taylor & Francis; 2011. p. 53–68.
43. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx*. 2004;1:182–8.
44. Nuyten DS, van de Vijver MJ. Using microarray analysis as a prognostic and predictive tool in oncology: focus on breast cancer and normal tissue toxicity. *Semin Radiat Oncol*. 2008;18:105–14.
45. Ogawa K, Murayama S, Mori M. Predicting the tumor response to radiotherapy using microarray analysis (Review). *Oncol Rep*. 2007;18:1243–8.
46. Svensson JP, Stalpers LJ, Esveldt-van Lange RE, Franken NA, Haveman J, Klein B, et al. Analysis of gene expression using gene sets discriminates cancer patients with and without late radiation toxicity. *PLoS Med*. 2006;3, e422.
47. Wouters BG. Proteomics: methodologies and applications in oncology. *Semin Radiat Oncol*. 2008;18:115–25.
48. Alaiya A, Al-Mohanna M, Linder S. Clinical cancer proteomics: promises and pitfalls. *J Proteome Res*. 2005;4:1213–22.
49. Tyburski JB, Patterson AD, Krausz KW, Slavik J, Fornace Jr AJ, Gonzalez FJ, et al. Radiation metabolomics. 1. Identification of minimally invasive urine biomarkers for gamma-radiation exposure in mice. *Radiat Res*. 2008;170:1–14.
50. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res*. 2009;15:431–40.

51. Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. *Br J Cancer*. 2004;90:747–51.
52. West CML, Elliott RM, Burnet NG. The genomics revolution and radiotherapy. *Clin Oncol*. 2007;19:470–80.
53. Andreassen CN, Alsner J. Genetic variants and normal tissue toxicity after radiotherapy: a systematic review. *Radiother Oncol*. 2009;92:299–309.
54. Alsner J, Andreassen CN, Overgaard J. Genetic markers for prediction of normal tissue toxicity after radiotherapy. *Semin Radiat Oncol*. 2008;18:126–35.
55. Baumann M, Hölscher T, Begg AC. Towards genetic prediction of radiation responses: ESTRO's GENEPI project. *Radiother Oncol*. 2003;69:121–5.
56. Burnet NG, Elliott RM, Dunning A, West CML. Radiosensitivity, radiogenomics and RAPPER. *Clin Oncol*. 2006;18:525–8.
57. Iwakawa M, Noda S, Yamada S, Yamamoto N, Miyazawa Y, Yamazaki H, et al. Analysis of non-genetic risk factors for adverse skin reactions to radiotherapy among 284 breast cancer patients. *Breast Cancer*. 2006;13:300–7.
58. Ho AY, Atencio DP, Peters S, Stock RG, Formenti SC, Cesaretti JA, et al. Genetic predictors of adverse radiotherapy effects: the Gene-PARE project. *Int J Radiat Oncol Biol Phys*. 2006;65:646–55.
59. West C, Rosenstein BS. Establishment of a Radiogenomics Consortium. *Int J Radiat Oncol Biol Phys*. 2010;76:1295–6.
60. Kerns SL, Ostrer H, Stock R, Li W, Moore J, Pearlman A, et al. Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys*. 2010;78:1292–300.
61. Kerns SL, Stock R, Stone N, Buckstein M, Shao Y, Campbell C, et al. A 2-stage genome-wide association study to identify single nucleotide polymorphisms associated with development of erectile dysfunction following radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys*. 2013;85(1):e21–8.
62. Rosenstein BS, West CM, Bentzen SM, Alsner J, Andreassen CN, Azria D, et al. Radiogenomics: radiobiology enters the era of big data and team science. *Int J Radiat Oncol Biol Phys*. 2014;89:709–13.
63. Alon U. An introduction to systems biology: design principles of biological circuits. Boca Raton: Chapman & Hall/CRC; 2007.
64. Feinendegen L, Hahnfeldt P, Schadt EE, Stumpf M, Voit EO. Systems biology and its potential role in radiobiology. *Radiat Environ Biophys*. 2008;47:5–23.
65. El Naqa I. Machine learning methods for predicting tumor response in lung cancer. *Wiley Interdiscip Rev: Data Min Knowl Discov*. 2012;2:173–81.
66. Eschrich S, Zhang H, Zhao H, Boulware D, Lee J-H, Bloom G, et al. Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform. *Int J Radiat Oncol Biol Phys*. 2009;75:497–505.
67. Oh JH, El Naqa I. Bayesian network learning for detecting reliable interactions of dose-volume related parameters in radiation pneumonitis. In: International conference on machine learning and applications (ICMLA), Miami. 2009.
68. Oh JH, Craft J, Al Lozi R, Vaidya M, Meng Y, Deasy JO, et al. A Bayesian network approach for modeling local failure in lung cancer. *Phys Med Biol*. 2011;56:1635–51.
69. Sanchez-Nieto B, Nahum AE. Bioplan: software for the biological evaluation of radiotherapy treatment plans. *Med Dosim*. 2000;25:71–6.
70. Sullivan R, Peppercorn J, Sikora K, Zalberg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncol*. 2011;12:933–80.
71. Deasy JO, Bentzen Sr M, Jackson A, Ten Haken RK, Yorke ED, Constine LS, et al. Improving normal tissue complication probability models: the need to adopt a 'Data-Pooling' Culture. *Int J Radiat Oncol Biol Phys*. 2010;76:S151–4.
72. Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, et al. 'Rapid learning health care in oncology' – an approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncology*. 2013;109:159–64.

73. Roelofs E, Dekker A, Meldolesi E, van Stiphout RGPM, Valentini V, Lambin P. International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining. *Radiother Oncol.* 2014;110(2):370–4.
74. Yan Q. Biomedical informatics methods in pharmacogenomics. *Methods Mol Med.* 2005;108:459–86.
75. Viswanathan GA, Seto J, Patil S, Nudelman G, Sealfon SC. Getting started in biological pathway construction and analysis. *PLoS Comput Biol.* 2008;4, e16.
76. Wang L, Zhang B, Wolfinger RD, Chen X. An integrated approach for the analysis of biological pathways using mixed models. *PLoS Genet.* 2008;4, e1000115.
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545–50.
78. Krishna R, Kelleher K, Stahlberg E. Patient confidentiality in the research use of clinical medical databases. *Am J Public Health.* 2007;97:654–8.
79. Nature Editorial. Making data dreams come true. *Nature.* 2004;428:239.