Literature Evaluation and Critique

4

Andrew J. Scally

4.1 Introduction

In the global modern healthcare environment, there is an expectation that you, as a healthcare professional, should base your practice upon the best available research evidence. National and international professional organisations for diagnostic imaging and radiotherapy practitioners emphasise the importance of an understanding, and the implementation, of an evidence-based approach to service development [1, 2].

Given that new 'evidence' continues to emerge at a rapid rate, all health professionals must be able to evaluate findings that are relevant to their practice and judge whether to incorporate change when this is necessary. This ability to critically appraise claims from research that are published in the literature, and independently evaluate the strength of such claims, is vital to diagnostic imaging and radiotherapy.

Although the idea of evidence-based medicine, or more generally evidence-based healthcare practice, has been traced back to the nineteenth century, the quality of health research has improved steadily. This does not mean that nowadays all research is conducted in a way that ensures the robustness of the conclusions. Established best practice is not always followed by researchers, and even where it is there is still the potential for hidden biases to be present in research that cannot easily be identified, eliminated, or controlled.

Standards for best practice in healthcare research have been published within recent years. These publications are an extremely valuable resource for students and qualified practitioners to help them make informed judgements on the quality and relevance of published research. Several organisations have developed critical

A. J. Scally (✉)
Diagnostic Radiography, School of Clinical Therapies, University College Cork, Cork, Ireland
e-mail: andrew.scally@ucc.ie

A. Ramlaul (ed.), *Medical Imaging and Radiotherapy Research: Skills and Strategies*, https://doi.org/10.1007/978-3-030-37944-5_4

appraisal tools which enable you critically to appraise research papers in a systematic way that involves the consistent application of the same relevant key questions for a given research design (see Appendix for web addresses).

4.2 Hierarchies of Evidence

A natural hierarchy of research evidence quality has emerged that is informed by the ease with which potential biases can be avoided or controlled. This is covered in useful texts by Sackett and colleagues and by Greenhalgh [3–5]. Although there are variations on the precise structure of this hierarchy, in particular to take account of qualitative research [6], it is broadly outlined in Table 4.1.

The highest level of evidence is widely considered to be a systematic review of well-designed randomised controlled clinical trials (RCTs), all of which aim to answer the same research question. However, the quality of a systematic review is necessarily constrained by the quality of the individual trials of which it is composed. Not all systematic reviews are reviews of clinical trials. Also, it is worth bearing in mind that published randomised controlled trials may be relatively uncommon in diagnostic imaging and radiotherapy. It is perfectly reasonable to perform a systematic review of observational studies, when there is little or no evidence from RCTs in a particular area of interest. The second level in the standard hierarchy is a large, well-designed RCT. It is therefore considered by some to represent the strongest kind of evidence. The third level of evidence is an observational cohort study; the fourth and fifth are observational case–control studies and cross-sectional studies (surveys) and the sixth is an observational ecological study (where individual-level exposure data are lacking, but aggregate, population-level data are available). Ecological studies are not common in medical imaging or radiotherapy, but are conducted occasionally. The lowest level of evidence

Table 4.1 The traditional hierarchy of evidence

Rank	Study design	Comment
1	Systematic review	Ideally of well-designed homogeneous RCTs. Status in the hierarchy may be relegated if RCTs are heterogeneous or if the review is of observational studies. A systematic review may or may not include a meta-analysis
2	Randomised controlled trial	Judgement required of the size and quality of the study and whether the results are definitive
3	Cohort study	A large, well-designed study may be more persuasive than a weak RCT, but a cohort study is more prone to bias
4	Case–control study	Causal inference more difficult to establish and more prone to bias than in a cohort study. Efficient design for rare conditions
5	Cross-sectional study (survey)	Causal inferences cannot be made. Provides information at a single instance of time
6	Ecological study	An observational study that uses aggregate level data, in the absence of an assessment of individual exposures
7	Case reports	Lack generalisability due to very limited sample size and their selective nature

is considered to be individual case reports, owing to their usual lack of generalisability.

Although this hierarchy reflects the reliability of the different study designs in terms of researchers' ability to eliminate or control biases within them, it should not be assumed that an observational study is always inferior to an RCT. Randomised controlled trials have their own potential problems and may not always be the most appropriate design for diagnostic imaging studies. They are not even ethical or appropriate in many situations, for example, when studying the health effects due to exposure to toxic agents, such as ionising radiation or chemical pollutants. Furthermore, the need for and appropriateness of such a hierarchy has been challenged. Wherever the design methods rank in the hierarchy, a well-designed study produces results that are more plausible than those from a poorly designed study. It has also been suggested that rigid adherence to this hierarchy has seriously misrepresented, or under-reported, the evidence supporting the more widespread use of new imaging methods in oncology [7].

In medical imaging, studies are commonly designed to measure the diagnostic accuracy of alternative imaging techniques and their combinations, or of the diagnostic performance of individuals or groups of observers/interpreters or even combinations of technology and observers. Diagnostic accuracy is commonly determined by the sensitivity of a test (the ability to detect disease when it is present), and the specificity of a test (the ability to exclude disease when it is absent). Although standards of good practice have been developed specifically for the design and reporting of such studies [8], a diagnostic accuracy study can have the characteristics of either an RCT or an observational study and so can be evaluated broadly within the evaluation framework relevant to an RCT or observational study.

4.3 Examples of Different Research Designs in Medical Imaging and Radiotherapy

Many different types of studies are published in the medical and health science literature. However, not all of them are primary or secondary evaluations of patient outcomes or, in the case of medical imaging studies, diagnostic performance. The research literature is broad and may cover many aspects of professional practice, for example, clinical audits, development of guidelines, developments in education and training, surveys of professional practice, surveys of user views and experiences, and experimental studies relating to assessment of health technologies. Although many of the principles addressed in this chapter can be applied to the appraisal of such articles, the main focus of this chapter is on research involving patient outcomes and diagnostic performance. It is from such studies that suggested changes can be made to clinical practice and improvements made in patient care.

Some examples from the medical imaging and radiotherapy/oncology literature that have used the primary research designs identified above are outlined in Table 4.2 [9–49]. No attempt here is made to appraise these studies but they could serve as helpful examples to which you could apply an appropriate critical appraisal tool from the options presented in the Appendix.

Table 4.2 Examples of the different study designs used in medical imaging and radiotherapy research

Authors	Title	Purpose
Systematic reviews		
Younger et al. [9]	Describing ionising radiation risk in the clinical setting: a systematic review	A systematic review seeking to identify and explore the techniques advocated for disclosing the risk to patients of ionising radiation from clinical medical imaging examinations
Sierinka et al. [10]	Systematic review of flexion/extension radiography of the cervical spine in trauma patients	To investigate whether flexion/extension (F/E) radiography adds diagnostic value to CT or MRI in the detection of cervical spine ligamentous injury and/or clinically significant cervical spine instability of blunt trauma patients
Gupta et al. [11]	Systematic review and meta-analyses of intensity-modulated radiation therapy versus conventional two-dimensional and/or or three-dimensional radiotherapy in curative-intent management of head and neck squamous cell carcinoma	To compare IMRT with conventional two-dimensional (2D) and/or three-dimensional (3D) radiotherapy (RT) in curative-intent management of HNSCC regarding disease-related endpoints
Harris et al. [12]	Systematic review of endoscopic ultrasound in gastro-oesophageal cancer	To review the literature about the use of endoscopic ultrasound for the preoperative staging of gastro-oesophageal cancer, especially staging performance and impact
Bryant et al. [13]	Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review	To conduct a systematic review of the clinical effectiveness and cost-effectiveness of cardioprotection against the toxic effects of anthracyclines given to children with cancer
Brealey et al. [14]	Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis	To quantify how accurately radiographers report plain radiographs in clinical practice compared with a reference standard
Randomised controlled trials		
Gupta et al. [15]	Neoadjuvant chemotherapy followed by radical surgery versus concomitant chemotherapy and radiotherapy in patients with stage IB2, IIA, or IIB squamous cervical cancer: a randomized controlled trial	To compare the efficacy and toxicity of neoadjuvant chemotherapy followed by radical surgery versus standard cisplatin-based chemoradiation in patients with locally advanced squamous cervical cancer

Table 4.2 (continued)

Authors	Title	Purpose
Brealey et al. [16]	Influence of magnetic resonance of the knee on GPs' decisions: a randomised trial	To assess the effect of early access to MRI, compared with referral to an orthopaedic specialist, on GPs' diagnoses and treatment plans for patients with knee problems
Bartholomew et al. [17]	A randomised controlled trial comparing lateral skull computerised radiographs with or without a grid	To investigate the effect on perceived image quality of the use or non-use of a secondary radiation grid for lateral skull radiography
Harrison et al. [18]	Randomized controlled trial to assess the effectiveness of a videotape about radiotherapy	To investigate whether the provision of a videotape, in addition to the standard information booklet, reduced pre-treatment worry about radiotherapy in cancer patients
Sala et al. [19]	A randomized controlled trial of routine early abdominal computed tomography in patients presenting with non-specific acute abdominal pain	To compare the effect of initial early computed tomography (CT) versus standard practice (SP) on the length of hospital stay, diagnostic accuracy, and mortality of adult patients presenting with acute abdominal pain
Ravasco et al. [20]	Dietary counseling improves patient outcomes: a prospective, randomized controlled trial in colorectal cancer patients undergoing radiotherapy	To investigate the impact of dietary counselling or nutritional supplements on several outcome measures (nutritional intake, nutritional status, and quality of life) in colorectal cancer patients
Cohort studies		
Slaar et al. [21]	Plain radiography in children with spoke wheel injury: a retrospective cohort study	To evaluate the type of radiographs that are obtained in children with BSI, to assess in which anatomical regions fractures occur, and to evaluate on which radiographs a fracture can be detected in children with bicycle spoke injury (BSI)
Damen et al. [22]	Additional value of different radiographic views on the identification of early radiographic hip and knee osteoarthritis and its progression: a cohort study	To investigate the prevalence and progression of early radiographic osteoarthritis (OA) of the hip and knee on different radiographic views, to determine whether different radiographic views have additional value in detecting early hip and knee radiographic OA cases or progression
Trakada et al. [23]	Pulmonary radiographic findings and mortality in hospitalized patients with lower respiratory tract infections	To identify whether specific radiographic findings in patients with lower respiratory tract infections predict mortality

(continued)

Table 4.2 (continued)

Authors	Title	Purpose
Aktas et al. [24]	Concomitant radiotherapy and hyperthermia for primary carcinoma of the vagina: a cohort study	To evaluate the supplementary value of adding hyperthermia to radiotherapy in patients with primary vaginal cancer
Virtanen et al. [25]	Angiosarcoma after radiotherapy: a cohort study of 332,163 Finnish cancer patients	To evaluate the risk of angiosarcoma after radiotherapy among cancer patients in Finland
Jaremko et al. [26]	Do radiographic indices of distal radius fracture reduction predict outcomes in older adults receiving conservative treatment?	To investigate whether radiographic deformities suggesting inadequate reduction would be associated with adverse clinical outcomes
Case–control studies		
Zhang et al. [27]	Diagnostic radiography exposure increases the risk for thyroid microcarcinoma: a population-based case–control study	A population-based case–control study to investigate whether there is an association between ionising radiation-based medical imaging procedures and incidence of thyroid cancer
Darby et al. [28]	Risk of ischemic heart disease in women after radiotherapy for breast cancer	A population-based case–control study of major coronary events (myocardial infarction, coronary revascularisation, or death from ischemic heart disease) to investigate if there is an increased risk due to receiving radiotherapy for breast cancer
Sernik et al. [29]	Ultrasound features of carpal tunnel syndrome: a prospective case-control study	To examine the most adequate cut-off point for median nerve cross-sectional area and additional ultrasound features supporting the diagnosis of carpal tunnel syndrome (CTS)
Cheng et al. [30]	Yoga and lumbar disc degeneration disease: MR imaging based case control study	To identify whether lumbar disc degenerative disease was reduced in practicing yoga instructors compared to a control group
Spruit et al. [31]	Regional radiotherapy versus an axillary lymph node dissection after lumpectomy: a safe alternative for an axillary lymph node dissection in a clinically uninvolved axilla in breast cancer. A case control study with 10 years follow up	To compare disease-free survival and overall survival in patients with clinically uninvolved axilla undergoing radiotherapy or axillary lymph node dissection following lumpectomy for breast cancer
Finlay et al. [32]	Advanced presentation of lung cancer in Asian immigrants: a case-control study	To determine if Asian immigrants to the USA present with more advanced lung cancer compared to non-Asians

Table 4.2 (continued)

Authors	Title	Purpose
Cross-sectional studies (surveys)		
Nightingale et al. [33]	A national survey of current practices of preparation and management of radical prostate radiotherapy patients during treatment	To gain insight into the variation of radiotherapy practices in the UK, focusing on pre-treatment preparations, on-treatment review, and management of radical prostate cancer patients undergoing radiotherapy
Snaith et al. [34]	A UK survey exploring the assistant practitioner role across diagnostic imaging: current practice, relationships and challenges to progression	An electronic survey of individual assistant practitioners (APs) within the NHS in the UK to explore utilisation, role scope, and aspirations
Goense et al. [35]	Patient perspectives on repeated MRI and PET/CT examinations during neoadjuvant treatment of esophageal cancer	To evaluate the experienced burden associated with repeated MRI and positron emission tomography with integrated CT (PET/CT) examinations during neoadjuvant treatment for oesophageal cancer from the perspective of the patient
Lutz et al. [36]	Survey on use of palliative radiotherapy in hospice care	Hospice professionals were surveyed to assess the need for palliative radiotherapy in the hospice setting
Davies et al. [37]	Radiation protection practices and related continuing professional education in dental radiography: a survey of practitioners in the North-east of England	To survey the opinion of practitioners on the availability of related postgraduate courses in the region
Jones and Manning [38]	A survey to assess audit mechanisms practised by skeletal reporting radiographers	To survey the role of plain film reporting radiographers and the methods they employ to evaluate the quality of their performance
Power et al. [39]	Videofluoroscopic assessment of dysphagia: a questionnaire survey of protocols, roles and responsibilities of radiology and speech and language therapy personnel	To survey videofluoroscopic practice and identify the roles and responsibilities of radiology and speech and language therapy personnel
Studies of diagnostic test accuracy		
Yi et al. [40]	Detection of noncalcified breast cancer in patients with extremely dense breasts using digital breast tomosynthesis compared with full-field digital mammography	To evaluate the tumour visibility and diagnostic performance of digital breast tomosynthesis (DBT) plus full-field digital mammography (FFDM), compared to FFDM alone, in patients with noncalcified T1 breast cancer

(continued)

Table 4.2 (continued)

Authors	Title	Purpose
Wooten et al. [41]	Bedside ultrasound versus chest radiography for detection of pulmonary edema: a prospective cohort study	This study compared the sensitivity and specificity of bedside ultrasound and chest radiography in diagnosing pulmonary edema
Grisaru et al. [42]	The diagnostic accuracy of ^{18}F-Fluorodeoxyglucose PET/CT in patients with gynaecological malignancies	To compare the diagnostic accuracy of PET/CT with standard imaging (CT/MRI/US) in patients with suspected recurrence of gynaecological malignancy
Burling et al. [43]	Virtual colonoscopy: effect of computer-assisted detection (CAD) on radiographer performance	To determine whether CAD as a "second reader" improves polyp detection by trained radiographers reporting on virtual colonoscopy examinations
MERCURY Study Group [44]	Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study	To assess the accuracy of preoperative staging of rectal cancer with magnetic resonance imaging to predict surgical circumferential resection margins.
Dai et al. [45]	Does three-dimensional power Doppler ultrasound improve the diagnostic accuracy for the prediction of adnexal malignancy?	To investigate the diagnostic accuracy of 3-D power Doppler ultrasound in the differentiation between benign and malignant adnexal masses
Qualitative studies		
Nightingale et al. [46]	A qualitative analysis of staff-client interactions within a breast cancer assessment clinic	An exploration of the culture of staff–client interactions within a breast cancer assessment clinic, using an ethnographic approach: the impact upon client experience
Nagle et al. [47]	Exploring general practitioners' experience of informing women about prenatal screening tests for foetal abnormalities: a qualitative focus group study	To explore GPs' experience of informing women of prenatal genetic screening tests for foetal abnormality
Poulos and Llewellyn [48]	Mammography discomfort: a holistic perspective derived from women's experiences	To use qualitative research methods to consider discomfort from a holistic perspective of the mammography experience derived from the women themselves
Colyer [49]	The role of the radiotherapy treatment review radiographer	A qualitative study to gain an understanding of the role of the radiotherapy treatment review radiographer

4.4 Basic Concepts in Critical Appraisal

There are some common concepts of critical appraisal of research literature that are relevant to most study designs. Some key pointers to evaluating a piece of published research are indicated below.

- Are there clear aims and objectives?
- Is there a defined research question?
- Do the authors have a good grasp of previous research in this field?
- Is the study relevant to clinical practice and carried out in 'real-world' circumstances?
- Is the method clear and well reported?
- Is the sample group sufficient and representative?
- Have appropriate inclusion and exclusion criteria been defined and used?
- Is the analysis of findings (quantitative and/or qualitative) appropriate and are the results appropriately interpreted?
- Are all study participants accounted for in the analysis?
- Are there any possible sources of bias, identified or unidentified by the authors, and have these been controlled or adjusted for in the analysis?
- Are unexpected events or negative findings discussed?
- Are weaknesses in the study acknowledged by the authors?
- Are the authors balanced in their views and conclusions?
- Are there useful recommendations?

All research should be designed to produce valid results. Validity is concerned with the extent to which inferences can be drawn from a study, in particular generalisations extending beyond the study sample, having taken into account study methods and the representativeness of the study sample. Two types of study validity can be distinguished.

- Internal validity relates to the ability of a research method to show a real relationship between cause and effect, such as whether observed differences in patient outcome can be attributed to the effect of the intervention under investigation.
- External validity is concerned with how generalisable the findings from a study are to a wider population, based on the sample of patients included in the study.

Bias, confounding, and chance can all reduce internal validity and may provide alternative explanations for an observed difference between study groups. Bias is often related to faults in the study design and can arise, for example, from an unrepresentative or skewed selection of patients for a study (selection bias), or a partial or unbalanced collection of data (information bias). To prevent bias, wherever feasible and necessary, good study designs will blind (or mask) the patients, clinicians, and even the researchers so that they are kept ignorant of anything that could lead them to a change in behaviour that might affect study findings.

Confounding occurs when an apparent effect of the intervention on patient outcome is in fact due to the action of a variable other than the intervention. When confounding is known or suspected, it can be controlled for in the design (e.g. randomisation, matching) or in the analysis (e.g. multivariable analysis). The effects of unknown confounders can be reduced by randomisation, but can never be eliminated entirely. A confounder is defined as an additional variable that is related to the dependent variable (e.g. disease or other outcome), but is not a consequence of this outcome. It is also related to the independent variable under study (e.g. intervention or exposure), but is not a consequence of this variable either.

The effect of any intervention can also be explained by chance. Even a randomised trial, which protects against systematic differences between groups, does not prevent differences between samples arising by chance although this does diminish as a sample size increases. The probability of an observed difference occurring by chance when no real difference exists is demonstrated by a *p*-value. A *p*-value of, for example, $P = 0.01$, informs us that assuming there is no real difference between treatments, the probability of uneven randomisation explaining the difference is around 1 in 100. Therefore you would not expect the play of chance to explain your study findings. External validity is likely to be threatened when only a small sample of patients is obtained from a single geographical location or there is self-selection of patients into a study (e.g. volunteers). This is therefore addressed by conducting research at multiple sites, increasing the sample size, and, when possible, selecting a random sample of patients into a study so that every eligible patient has an equal chance of being selected and thus the sample should be representative of the target population. It is often very difficult, if not impossible, to obtain a truly random sample. Many studies therefore use a convenience sample of, for example, consecutive eligible patients attending a department/clinic. In such circumstances particular attention should be given to the representativeness of a sample chosen for a study.

4.5 The Typical Structure of a Research Paper

Most research articles are similarly structured, though the precise structure may vary according to the editorial policy of a journal and the design of a study. The general structure of a published research article is as follows.

- *Title*—Making clear the purpose and design of the study.
- *Authors*—Including names, qualifications, and affiliations.
- *Abstract*—Summarising the background and purpose, structure, results, and conclusions of the study.
- *Introduction*—Presenting the background to the study and its rationale, including reference to previous relevant research.
- *Methods*—Including a thorough description of the study design, an outline of the practicalities of how it was done, an explanation of how potential biases were addressed, and a description of the data analysis methods used.

- *Results*—Presentation of the results, with emphasis on the primary outcome measure identified for the study.
- *Discussion*—Interpretation of findings, recognition of any limitations of the study, the discussion of the findings in the context of what was previously known, and suggested implications for practice.

4.6 Preliminary Steps in a Critical Appraisal

When setting out to identify relevant research in an area of practice, the first task should be a systematic search of the literature using the methods discussed in Chap. 3.

From the outset, it is important to understand that there is no such thing as a perfect research study. Even the best conducted studies have potential flaws that are impossible to avoid. For example, almost all research involving patients or staff needs informed consent from the participants. If those who refuse to give that consent are over-represented in particular subgroups, such as gender, age or ethnicity, then the representativeness of the sample could be open to challenge. Also, all research is subject to logistical and economic constraints and so compromises have to be made when considering what is feasible. It is far easier to criticise the work of others than to design a study that is beyond criticism. It is thus important, when critically appraising a paper, to consider unavoidable constraints within which researchers are working and to assess whether they have implemented all measures reasonably available to them to optimise the robustness of the study. A critical evaluation of a study is not just about finding fault. We should also praise when this seems appropriate.

Another issue to bear in mind is that there is a difference between the assessment of the method and findings of a research article and the assessment of the written presentation of that article, although both are important. Students often focus too much on the presentation of a research study when evaluating it, leading to a critique which is descriptive and uncritical. Reports of studies of high inherent quality may be poorly presented by the authors, meaning that some information may be lacking and a fair assessment of study quality is hard to undertake. Conversely, a weak study could be well presented, with strong structure and great detail, and yet could contain flaws so significant that no meaningful inferences can be drawn from it.

Once you have identified a research article that may be of relevance to you and that you may wish to critically evaluate, there are a few preliminary steps and questions that should be considered before progressing further.

1. A reading of the abstract may clearly identify whether or not the paper is relevant to your purpose. If it is still unclear after reading the abstract, a quick reading of the article may be necessary before you are able to make a decision. Is the nature of, and emphasis within the study relevant to the purpose of your literature search and evaluation? Do not spend too much time on articles that are peripheral or irrelevant to your purpose.

2. Does the title accurately reflect the content of the study or is it uninformative or misleading?
   a. The title may give the impression that the study comprises fresh (primary) data, but it may in fact be a review of previously published work. In this case the article may be of help to you in appraising some of the other pieces of published research to which it makes reference, but this is no substitute for your own independent assessment of the original studies.
   b. Is the study measuring the outcome(s) it says it is measuring, or are surrogate outcome measures being used? (A surrogate outcome measure is one that is presumed—with or without good evidence—to be associated with the primary outcome of interest, but is usually easier to measure.)
3. Does the list of authors suggest that they have the relevant expertise in all important aspects of the research? You should never assume that eminence in a particular field guarantees the quality of the research, nor that an unknown author, or an author from a different discipline should not be trusted or believed. All research should be appraised on its merits, but extra vigilance in the appraisal of the robustness of the research may be suggested where certain relevant expertise may appear to be lacking, for example the absence of a medical statistician from the list of authors of a paper that utilises seemingly complex data analysis methods.
4. Is the study design what it says it is? Not all studies reported as RCTs are randomised or adequately controlled; some studies reported as cohort studies could more accurately be described as cross-sectional studies. The answer to this question is not always clear-cut and the paper may require more thorough evaluation before it can be definitively answered.
5. Has the paper been commented upon already? Peer-reviewed journals normally include a letters section, in the printed edition and/or online, within which members of the health/scientific community pass informed comment on research previously published in the journal. In the online content pages of peer-reviewed journals, letters commenting on the research are often identified adjacent to the original article. It is always worthwhile to read the published views of other commentators on a research article, though of course these comments themselves should be subject to critical appraisal.
6. In the introduction to the paper, have the authors adequately identified and summarised the available evidence in the relevant subject area and justified the need for their own study? The Declaration of Helsinki, which governs the ethics of biomedical research, requires that research involving people should be underpinned by a thorough knowledge of the scientific literature in order that research volunteers are not subject to unnecessary harm or inconvenience.

4.7 Critical Evaluation Strategies According to Design Method

We next consider the specific requirements for a critical evaluation of studies comprising the designs illustrated in Table 4.2. Some key resources have been identified to assist students, and qualified practitioners alike, in performing the

evaluation, including reference to key publications explaining the rationale for giving attention to specific aspects of the study design and an evaluation tool/checklist that provides a pro forma for a systematic evaluation. A few of the key issues for each study design are briefly outlined, but a more thorough explanation of the importance of each issue is provided in the essential resources indicated.

4.7.1 Critical Evaluation of Systematic Reviews

- *Useful resources*
- The PRISMA Statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [50]; an appropriate checklist from Appendix.
- *Specific issues to consider*
  The purpose of a systematic review is to help healthcare providers and other decision-makers to make clinical decisions about best practice. Rather than reflecting the views of the authors or being based on a possibly biased selection of published literature, a systematic review involves locating all the available evidence in relation to a specific research question, appraising the quality of the evidence identified, synthesising the available evidence, and if relevant, statistically aggregating the evidence of all relevant studies. Systematic reviews, and the statistical meta-analytic methods they use, were originally developed to synthesise the results from several homogeneous randomised controlled trials. In today's healthcare environment, they have a much broader application and can incorporate more heterogeneous RCTs and observational and qualitative studies, respectively. Systematic reviews should adhere to strict scientific design in order to make them more comprehensive and to minimise the chance of bias (systematic errors) and random errors (mistakes occurring by chance), thus providing more reliable results from which to draw conclusions and make decisions. The following should therefore be considered when critically appraising the quality of a systematic review.
- *Research question*
  What question did the systematic review address? The main research question should be clearly stated and preferably describe the relationship between population, intervention (or test or exposure), comparison intervention, and outcome (PICO). Knowing the population is important to decide whether the review applies to your specific patient group. The intervention is a planned course of action and the exposure something that happens. These again need to be described in detail, as should the comparison intervention, to ensure clarity and to help you determine what contributed to the outcome. The most important outcomes, beneficial or harmful, should also be clearly defined. The title, abstract, or final paragraph of the introduction should clearly state the research question. See Chap. 2 for more guidance on finding and formulating a research question.
- *Searching*
  Is it unlikely that important, relevant studies were missed? The information sources searched should be clearly described (e.g. databases, registers, personal files, expert informants, hand-searching) and any restrictions (e.g. years considered, publication

status, language of publication). A comprehensive search for all relevant studies should include the major bibliographic databases (e.g. Medline, EMBASE, Cochrane), a search of reference lists from relevant studies, contact with experts to inquire about, in particular, unpublished studies, and the search should, ideally, not include English language only. The search strategy should be clear, explicit, and reproducible and be described in the methods section of the paper.

- *Study selection*
  Were the criteria used to select articles for inclusion appropriate? The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design) should be clearly defined before the search is undertaken to ensure the consistent and appropriate selection of eligible studies into the review. The methods section should describe in detail these criteria.
- *Validity assessment*
  Were the included studies sufficiently valid for the type of questions asked? There should be predetermined criteria used to assess the quality (e.g. randomisation, blinding, completeness of follow-up) of each included study depending on the type of clinical question being asked. The process of assessing validity should also be described, for example, masking the reviewers to who were the authors of the study and whether two reviewers independently applied the quality criteria. The methods section should describe the quality criteria used and the process of applying the criteria. The results section should provide information on the quality of studies and, if applicable, extent of agreement between reviewers when appraising studies.
- *Study characteristics*
  Were the study characteristics similar? The type of study design, participants' characteristics, details of intervention, and outcomes should be described. Heterogeneity, or inconsistency of results across different studies, could be explained by differences in study characteristics. The possibility of heterogeneity should be explored visually through the examination of forest plots of the results of studies or, more formally, with statistical tests such as chi-square (see Chap. 15 for common statistical tests).
- *Data synthesis*
  Were the methods used to combine the findings of the relevant studies reported (if applicable)? The principal measures of effect (e.g. relative risk), method of combining results (statistical testing and confidence intervals), and a priori sensitivity and subgroup analyses should all be reported in the methods section and the findings in the results.

4.7.2 Critical Evaluation of Randomised Controlled Trials

- *Useful resources*
  The CONSORT 2010 Statement (Consolidated Standards of Reporting Trials) and associated resources [51]; an appropriate checklist from Appendix.
- *Specific issues to consider*

A randomised controlled trial is usually regarded as the strongest type of primary research study in health care, although it may not be feasible in all situations and for some types of research question it is not appropriate. For example, determination of the prevalence of a particular disease in a population requires a well-designed survey. In an RCT, subjects (usually patients) are allocated in a controlled but random way to two or more groups, receiving different interventions. An RCT is generally the best available design to test (a) whether a medical intervention works at all (e.g., a drug, surgical technique, exercise regimen, radiotherapy treatment, or diagnostic screening test), by comparing outcomes of the intervention group with the placebo or control group; (b) whether a new intervention is superior to existing treatment, by comparing outcomes with a group receiving standard care; or (c) whether a new, cheaper, or less invasive intervention is equivalent in its effect to the current expensive or invasive procedure.

- *Participants*
  What were the eligibility criteria for participants? What were the settings (primary care, secondary care, community) and geographical locations from which recruitment was made? What were the specific inclusion and exclusion criteria and were they appropriate? Were participants randomly selected (how subjects are randomly allocated is discussed later) or was a convenience sample used (e.g. all consecutive patients over a 1-month period)? Are the sample characteristics representative of the population of patients in whom you are interested, for example in terms of age, gender, ethnicity, socioeconomic characteristics, disease type, and severity? If not, caution may be advised in generalising the results to your patient population.
- *Interventions*
  Is the precise nature of the experimental and control interventions clear? How and when were the interventions administered? Did the control group receive a placebo or standard care? Was this a comparison of a new intervention compared to standard care or a new intervention in addition to standard care?
- *Outcome measures*
  Ideally there should be only one primary outcome measure, though occasionally more than one may be justified. Several secondary outcome measures may also be identified, but should be interpreted with more caution. Were the outcome measures adequately defined and accurately measured? Were they measured just once or repeated measures made over time? If the latter, then there will be important statistical issues to consider. Did the primary outcome measure evaluate the real concept of interest or were surrogate outcome measures used?
- *Sample size*
  Was an appropriate prospective sample size estimation undertaken? If so, was previous research used to estimate a likely effect size (true difference in outcomes between the human groups included in the trial) or was a judgement made regarding the minimum effect size that would represent a clinically important effect? If no sample size estimation was undertaken, then there is a serious risk of the study being overpowered (an unnecessarily large sample) or underpowered (too small a sample to make any valid findings).

- *Randomisation*
  Was the randomisation method adequately described and was it open to abuse? Were random number tables used, or a computer-generated random number sequence? Was simple randomisation used or a restricted method, for example random number blocks, stratification, or minimisation (a method used to minimise differences in baseline characteristics between groups)? Was the group allocation of all participants adequately concealed?
- *Blinding*
  Ideally, the group allocation of all participants should remain unknown to participants and to those responsible for the administration of the intervention and data collection and for their general medical care, until after the data are analysed. Sometimes this is very difficult and may at times not be logistically feasible. Lack of blinding, or its inadequacy, can in some circumstances seriously compromise the validity of a study (due to complex psychological issues affecting both patients and those responsible for their care), but in other circumstances it may be of limited importance (e.g., lack of blinding of a patient is unlikely to seriously compromise a study evaluating the diagnostic accuracy of alternative tests, since the outcome measure relates to observer interpretation of imaging signs rather than to the degree of improvement in the health status of the patient). Were all reasonable steps taken to ensure adequate blinding? What more could have been done?
- *Statistical methods*
  Were appropriate statistical methods chosen to analyse all outcome measures? For simple analyses the answer to this question should be within the scope of all readers. Although RCTs can be complex to undertake, the statistical methods chosen for their analysis (at least that of the primary outcome measure) are usually relatively simple because the groups should be fairly well balanced on all factors that may affect outcome, apart from the intervention group to which they have been assigned. More complex methods may be used for some secondary outcome measures. The primary analysis of a clinical trial should be based on 'intention to treat'. In other words, patients should be analysed within the group to which they were randomly allocated rather than according to the treatment they may actually have received.
- *Results*
  Was the flow of participants through each stage of the trial made clear? Were all important baseline characteristics of participants summarised and were they very similar between trial groups? Were all participants accounted for, with the number of dropouts evaluated and reasons given for all missing data? Were the results of statistical analyses adequately reported (effect size, confidence intervals (where possible), and statistical significance)? Were secondary and further exploratory analyses identified as such? Was an appropriate account taken of multiple analyses in determining the threshold for statistical significance?
- *Interpretation*
  Are the researchers' claims justified by their results, in the context of what is already understood from previous research? Were the limitations of the study (in

terms of inclusion criteria, uncontrolled potential biases, sample size, and precision) adequately recognised by the authors? Is the evidence presented sufficiently strong to confirm, or warrant reconsideration of, current practice?

4.7.3 Critical Evaluation of Observational Studies

- *Essential resources*
- The STROBE Statement (Strengthening the Reporting of Observational Studies in Epidemiology) [52]; an appropriate checklist from Appendix.
- Specific issues to consider are presented below.

4.7.3.1 Cohort Studies

- A cohort is a group of people with shared or common characteristics for the purpose of health research and is often followed longitudinally over time. As in the case of an RCT, groups within the cohort (sample) are compared with one another. The main difference between an RCT and a cohort study is that in the latter, subjects are not allocated at random to interventions or exposures. This lack of random allocation makes it harder to eliminate or control biases due to systematic baseline differences between cohort subgroups to be compared. Otherwise, the characteristics of cohort studies are similar to RCTs. A cohort study is usually the best available study design in situations where an RCT is either unethical or impractical. Cohort studies are not the most efficient design for studies investigating rare occurrences or diseases with long latency periods.
- *Participants*
  Settings, locations, and periods of recruitment, follow-up, and data collection should all be stated. What were the eligibility criteria for inclusion and were they appropriate? If two or more sub-cohorts were compared, might there be any other systematic differences between them (e.g., different prior information, recruited at different times)?
- *Exposure*
  What was the nature of the "exposure", how was it measured, and how did it vary across the cohort? Was it measured reliably? In most cohort studies, the exposure consists of some agent which the subject physically receives, for example, a vaccine, drug, other medical intervention, or an environmental toxin such as a radiation exposure or inhalation of some toxic chemical agent. In many medical imaging studies, such as those by Trakada et al. [23] and Jaremko et al. [26] in Table 4.2, the role of the exposure is taken by imaging findings because we want to assess the degree to which the imaging appearances can predict patient outcome.
- *Outcome measures*
  Was a primary outcome measure adequately defined and was it appropriately and adequately measured? What were the additional outcome measures? If the study was longitudinal (repeated measurements over time), was a specific time point

identified as the primary time point or was the trend over time of primary interest?

- *Other variables*
  Unlike the case with RCTs, in a cohort study we cannot be assured of reasonable balance between groups in a cohort study, so baseline differences between groups may need to be accounted for in the analysis. There may be a number of potential confounders (other variables associated with both the exposure and the outcome measure) that need to be adjusted for in the analysis. All variables of importance in a study, their method of measurement/determination, and their role (measure of exposure, outcome measure, or confounder) should be identified.
- *Sample size*
  The same considerations are applicable as for RCTs, but the methods of estimation could potentially be more complex due to a necessarily more complex statistical analysis.
- *Control of biases*
  Did the authors identify all serious potential sources of bias in the study and make all reasonable efforts to control them?
- *Statistical methods*
  In some cohort studies, the analysis methods used can be quite straightforward, but often various types of regression model are required to accommodate repeated measures on individuals and/or adjustment for confounders. The authors should explain clearly the nature of the analyses proposed.
- *Results*
  All relevant details relating to the recruitment of participants should be reported, including the total number of people eligible for participation, the numbers declining consent, any missing data, and the numbers lost to follow-up. Actual numbers, rather than just percentages, should be reported. The analysis process should be adequately described, including unadjusted and adjusted estimates, and the confounders adjusted for. Effect size and measures of uncertainty should be presented as well as statistical significance.
- *Interpretation*
  As for RCTs, but potential limitations due to uncontrolled biases require even more careful consideration.

4.7.3.2 Case–Control Studies

Case–control studies involve comparing people with a disease or characteristic (the cases) with otherwise similar people who lack that disease or characteristic (the controls). These studies have proved very useful for investigating cause and effect, for example, linking smoking with lung cancer. They are most appropriately used in situations where a disease process being investigated is rare. They are however more prone to hidden biases than cohort studies. A cohort study in such cases would need to be inordinately large to ensure that sufficient cases of disease were included in order to effect comparisons between subgroups. In a case–control study, the cases of disease are identified first; appropriate controls are then selected for comparison, and the focus is on a comparison of an exposure of interest between the two groups.

Direct inference of causation cannot be made from case–control studies because our starting point is the identification of cases that already have the disease of interest.

- *Participants*
  Particular care is required in explaining how case ascertainment was determined because misclassification is a serious potential bias in studies of this type. Suitable controls are often also problematic to recruit. A control group should be similar in all its characteristics to a case group except with regard to their disease status and, potentially, their 'exposure'. Were there equal numbers of cases and controls or are two or more controls recruited for each case? Were controls matched or unmatched to cases? If matched, what were the matching criteria?
- *Exposure, outcome, other variables, sample size, and control of biases*
  As for cohort studies
- *Statistical methods*
  As for cohort studies. An additional issue for case–control studies arises when the cases and controls are matched. In this case, the matching has to be specifically accounted for in the methods of analysis. For example, McNemar's test should be used to analyse case–control pairs, rather than a simple chi-squared test that compares groups at the aggregate level. If a logistic regression model were to be used for a cohort study, then for a matched case–control study a conditional logistic regression model, which incorporates the matching variables, should be used.
- *Interpretation*
  As for cohort studies.

4.7.3.3 Cross-Sectional Studies

Studies of this type involve a 'snap-shot' investigation of some phenomenon of interest at a particular instant or over a short period of time. In epidemiology, they are often used to ascertain the prevalence of a particular disease at a moment in time in a well-defined geographical area or subject group. Surveys are usually examples of this design and are used widely in studies involving both patients and health professional groups.

- *Participants*
  Were the eligibility criteria for inclusion clearly stated? What were the settings and locations of recruitment? What were the methods of recruitment? Are the characteristics of the sample similar to those of your population of interest? What potential biases are present in the methods of sample selection?
- *Variables*
  In epidemiological studies, this study design is often used to determine the prevalence of a disease in a population of interest. More broadly, surveys can be used to obtain information on a wide and complex range of issues using simple or complex, single or multiple questionnaires. Were all quantitative variables adequately defined and were the measures valid? If a questionnaire was used, has it been previously validated and was it suitable for the purpose for which it was used?

- *Statistical methods*
  Analyses could comprise simple evaluations of prevalence of disease (or other concept of interest), where confidence intervals should also be provided if a random sample of the population of interest is used. Commonly, surveys are based on non-random samples, in which case any statistical inferences should be treated with caution. Many surveys are essentially descriptive in nature, with assessment of responses to a large number of questions. The validity of any statistical comparisons in such circumstances is even more open to question unless efforts were made to minimise the number of formal comparisons and account for multiple testing. It is not possible to ascertain causality from cross-sectional studies.
- *Interpretation*
  As for cohort studies.

4.7.4 Critical Evaluation of Studies of Diagnostic Test Accuracy

- *Useful resources*
  The STARD (Standards for Reporting of Diagnostic Accuracy Studies) 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration [53]. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies [54]. This is a generic tool used to appraise the quality of primary studies in systematic reviews of diagnostic accuracy. An appropriate appraisal tool from Appendix.
- *Specific issues to consider*
  Diagnostic accuracy studies are integral to the evaluation of new and existing imaging technologies and to the measurement of their ability to distinguish patients with and without the target disorder. Studies that assess the performance (or accuracy) of a medical imaging modality, such as magnetic resonance imaging of the knee, should apply the modality to a prospective and consecutive series of patients with and without the target disease, such as meniscal or ligamentous injury, and then the patients undergo a second gold standard or reference test, such as arthroscopy. The relationship between the results of the imaging modality (or index test) and disease status, as determined by the gold standard, is described using probabilistic measures such as sensitivity (correct abnormal diagnosis of patients with disease) and specificity (correct normal diagnosis of patients without disease). It is important that the results of the gold standard are close to the truth, or the performance of the imaging modality will be poorly estimated.
- *Patient selection*
  Was the setting for the evaluation described? Was the patient spectrum representative of patients who will receive the test in practice? Were selection criteria clearly described? Patient selection processes affect which patients enter a study and this can affect both its internal validity (in that a biased selection of patients could inflate the index test performance) and external validity (in that a

narrow selection of patients could limit the generalisability of the findings). The setting, such as a specialised centre, could be referred rare or problem cases which could affect the prevalence and severity of disease in a patient sample and thus study generalisability. Similarly, an appropriate spectrum of patients should be selected in terms of demographics and clinical features; a limited spectrum can considerably bias the sensitivity and specificity of a test. Predetermined selection criteria should be described to ensure the explicit and reproducible selection of patients into the study.

- *Observer selection*
  Was the effect of the characteristics of observers on test performance considered? Was observer variability determined? The characteristics of the observers involved in the interpretation of images are important in diagnostic accuracy studies of imaging modalities, as they can affect estimates of test performance and generalisability. For example, a study that includes a single, highly specialist observer is likely to have low external validity. In contrast, such an observer could help to produce the best estimates of test accuracy and so increase internal validity. Characteristics of observers that have been considered important in the appraisal of a diagnostic accuracy study include allocation of images to be read by observers; number, experience, and training of observers; profession of observers; and assessment of observer variability and examination of its effect on test accuracy. The variability of an observer, or the reproducibility with which an observer interprets an image, can be assessed as different observers interpreting the same sample of images (interobserver) or the same observers interpreting the same images on separate occasions (intra-observer). The greater the observer variability, the less reliable are the results of the imaging modality (see receiver operating characteristics (ROC) in Chap. 12).
- *Choice and application of the reference (gold) standard*
  Was the reference standard likely to correctly classify the target condition? Was the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? Did the whole sample or a random selection of the sample receive verification using a reference standard diagnosis? Did patients receive the same reference standard regardless of the index test result? Was the reference standard independent of the index test? The reference standard is the method used to determine the presence or absence of the target condition and is assumed to be 100% sensitive and specific. In reality, every test is fallible, but if the reliability of a reference standard is high then methods can be used to account for imperfection. The choice and application of the reference standard is therefore very important in determining estimates of an index test performance. A valid reference standard should be chosen that correctly classifies the target condition and is applied within a clinically acceptable timeframe after the index test to prevent a change in the target condition explaining a difference in the results between the index test and reference standard. The same reference standard should be applied regardless of the results of the index test and preferably to the whole or at least a random sample of patients. Not applying the same reference standard to deter-

mine the definitive diagnosis in the sample of patients could also explain differences in results between an index test and reference standard and thus estimates of test performance. Nor should an index test form part of the reference standard as this too will introduce bias.

- *Independence of interpretation*
  Were the index test results interpreted without knowledge of the results of the reference standard? Were the reference standard results interpreted without knowledge of the results of the index test? Assessments that involve clinical judgement, such as the interpretation of medical images, are susceptible to bias owing to prior expectation. Therefore, the interpretation of the results of a test under evaluation should be undertaken independently, blind to the results of the reference standard. Similarly, the results of a reference standard should be interpreted blind to the results of an index test. Not avoiding this bias may lead to inflated measures of diagnostic accuracy.
- *Measurement of results*
  Were uninterpretable/intermediate test results reported? Were withdrawals from a study explained? Indeterminate index test results might arise due to factors such as technical faults or inferior image quality. Patients might also withdraw from a study before the results of either or both of an index test and reference standard are known. This could be for many uncontrollable reasons such as death, changing residency, or unwilling to continue co-operation. A study should fully report these indeterminate test results and withdrawals. If they are essentially random and not related to the true disease status, they should not introduce bias but could affect generalisability [55].

4.7.5 Critical Evaluation of Qualitative Studies

- *Useful resources*
  Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups [56]; an appropriate appraisal tool from Appendix.
- *Specific issues to consider*
  Qualitative research aims to provide an in-depth understanding of social phenomena such as people's experiences and perspectives in the context of their personal circumstances or settings. To explore the phenomena from the perspective of those being studied, qualitative studies are characterised by the use of unstructured methods which are sensitive to the social context of the study; the capture of data which are detailed, rich, and complex; a mainly inductive rather than deductive analytic process; and answering 'what is,' 'how,' and 'why' questions. It employs a variety of methods including interviews, focus groups, observations, conversation, discourse and narrative synthesis, documentary, and video analysis.
- *Sampling*

Were the criteria for selecting the sample clearly described? Was the sampling strategy comprehensive to ensure the generalisability of the analyses? Were the characteristics of the sample adequately described? As with quantitative studies, it is important for exclusion and inclusion criteria to be clearly specified. This will help you to judge whether the appropriate characteristics of participants according to age, gender, ethnicity, and other relevant demographic features were identified. Unlike quantitative research that requires the selection of a consecutive or random sample of patients that are representative of a population, qualitative research requires the selection of specific groups of people that possess characteristics relevant to the phenomena being studied. Convenience sampling might be used for pragmatic reasons and involves choosing individuals that are easiest to reach, but this might introduce bias. Alternatively, there is purposive sampling when patients/participants are deliberately selected because they possess a certain characteristic and this helps to ensure a range of viewpoints are represented. The characteristics of a sample must be described to help you judge whether an appropriate selection of patients/participants has been included.

- *Data collection*
  Were the data collection methods appropriate for the research objectives and setting? Common methods of data collection include observations, interviews, focus groups, or document analysis. Observation is used to record social phenomena directly by the investigators themselves or indirectly through audiotape or videotape recording. Direct observation requires an investigator to spend time in the social context under investigation and collect data through their nonparticipation or participation in a setting. In nonparticipant observation a researcher does not get involved in the social interactions being observed. It is therefore important to consider whether an observer is likely to be ignored or could inadvertently affect the behaviour of those observed. In participant observation a researcher is part of the social setting, but again it must be considered whether their dual role as observer and participant influences social interactions. Collecting data using interviews might include semi-structured or unstructured individual interviews or may be conducted in focus group settings. Individual interviews are more useful for evoking personal experience, in particular, on sensitive topics; focus groups use group interaction to generate data, but their public forum might inhibit candid disclosure. You should consider the rationale for the choice of a particular method of data collection and its appropriateness for the topics being studied. Finally, analysis of documents such as charts, journals, and correspondence might provide qualitative data. This can be achieved by counting specific content elements (e.g., frequency of specific words being used) or interpreting text (e.g., seeking nuances of meaning). The former rarely provides adequate information for analysis. You should consider whether multiple methods of collecting data are included. This approach can improve the rigour of a study as it allows investigators to examine subjects' perspectives and behaviour from different angles and to capture information with one method that was not possible with another.

- *Validity*
  Are the results of the study valid? This is concerned with whether the data collected truly reflect the phenomena under scrutiny. One method to achieve this is to use triangulation, which refers to the collection of data from different sources using different research methods to identify patterns of convergence. Another approach to validating data is to feed the findings back to the subjects to see if they consider the findings a reasonable account of their experience. There should also be appropriate consideration of 'negative' or 'deviant' cases by a researcher who should give a fair account of these occasions and explore reasons for why the data may vary.
- *Data analysis*
  Were the data appropriately analysed? Qualitative research begins with a general exploratory question and preliminary concepts. Relevant data are collected, patterns observed, and a conceptual framework is developed. This process is iterative, with new data being incorporated that may corroborate or challenge an emerging framework. The process should continue until the framework stabilises. Further data would thus not substantially affect the process. At this point theoretical saturation or informational redundancy is said to have been achieved. Qualitative data, and their interpretation, should be cross-referenced across multiple sources, using triangulation, in order to ensure the robustness of the analysis. Data synthesis should also, ideally, be undertaken by more than one person, and consensus agreement reached, to reduce the risk of researcher bias due to preconceived ideas about the phenomena investigated.

4.8 Conclusions

It is an expectation of all health professionals that they maintain an awareness of relevant research developments in their area(s) of practice in order to inform continuous improvement in patient care. In medical imaging and radiotherapy, rapidly evolving technology continually leads to the refinement of existing diagnostic/therapeutic techniques, and the development of new diagnostic and therapeutic methods.

Evidence-based practice requires the use of current best evidence in making decisions about patient care. This can only be achieved through (a) an understanding of research concepts; (b) an awareness of the characteristics, application, and limitations of commonly used research designs; and (c) an ability to critically appraise and evaluate research evidence in order that appropriate decisions can be made regarding when and how practice should evolve or change.

Key steps in terms of adopting a systematic approach to critical evaluation of the literature are presented. In addition internationally accepted standards, detailing best practice in research design for all commonly used research approaches and methods, are highlighted. Links are also provided to a variety of critical appraisal templates that can be applied to individual research studies, thereby aiding a consistent and systematic approach. References to several professionally relevant

examples of published research studies are provided, for each type of research design, which students, educators, and practitioners can use to practise their critical appraisal skills.

Appendix: Resources for Critical Appraisal

The following resources have been developed to assist medical and health practitioners in the critical appraisal of research appropriate to their practice. The checklists have similarities and some differences, so it is worth exploring a few of them to find a checklist that you think is best suited to your purpose:

- BestBETs (Best Evidence Topics)—critical appraisal worksheets for a wide range of study types: https://bestbets.org/links/BET-CA-worksheets.php (accessed 19 May 2019)
- Boynton PM and Greenhalgh T. Hands-on guide to questionnaire research: Selecting, designing, and developing your questionnaire. BMJ (2004);328:1312–1315.—Checklists for questionnaire design and the critical evaluation of a questionnaire based studies. Table E Critical appraisal checklist for a questionnaire study available at: https://www.bmj.com/content/suppl/2004/05/27/328.7451.1312.DC1 (accessed 19 May 2019)
- Centre for Evidence-Based Medicine (CEBM)—Critical appraisal tools for systematic reviews, RCTs, diagnostic accuracy, prognostic and qualitative studies: https://www.cebm.net/2014/06/critical-appraisal/ (accessed 19 May 2019)
- Critical Appraisal Skills Programme (CASP)—Critical appraisal tools for systematic reviews, qualitative studies, RCTs, cohort, case-control and diagnostic accuracy studies, economic evaluation studies and clinical prediction rules: https://casp-uk.net/casp-tools-checklists/ (accessed 19 May 2019)
- The Scottish Intercollegiate Guidelines Network (SIGN)—Critical appraisal notes and checklists: https://www.sign.ac.uk/checklists-and-notes.html (accessed 19 May 2019)
- The Joanna Briggs Institute (JBI)—Critical appraisal tools for a broad range of study designs: https://www.joannabriggs.org/critical_appraisal_tools (accessed 19 May 2019)

References

1. EFRS statement on radiography education. 2019. The European Federation of Radiographer Societies (EFRS).
2. European Qualifications Framework (EQF). Level 6 benchmarking document: radiographers. 2nd ed. 2018. European Federation of Radiographer Societies (EFRS).
3. Greenhalgh T. How to read a paper: the basics of evidence-based medicine and healthcare. 6th ed. Wiley Blackwell: Hoboken; 2019.
4. Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence based medicine: what it is and what it isn't. Br Med J. 1996;312:71–2.

5. Sackett DL, Straus SE, Richardson WS, et al. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.
6. Daly J, Willis K, Small R, et al. A hierarchy of evidence for assessing qualitative health research. J Clin Epidemiol. 2007;60:43–9.
7. Hicks RJ. Health technology assessment and cancer imaging: who should be setting the agenda? Cancer Imaging. 2004;4(2):58–60.
8. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
9. Younger CWE, Wagner MJ, Douglas C, Warren-Forward H. Describing ionising radiation risk in the clinical setting: a systematic review. Radiography. 2019;25:83–90.
10. Sierinka JC, van Lieshout WAM, Beenen LFM, et al. Systematic review of flexion/extension radiography of the cervical spine in trauma patients. Eur J Radiol. 2013;82:974–81.
11. Gupta T, Kannan S, Ghosh-Laskar S, Agarwal JP. Systematic review and meta-analyses of intensity-modulated radiation therapy versus conventional two-dimensional and/or or three-dimensional radiotherapy in curative-intent management of head and neck squamous cell carcinoma. PLoS One. 2018;13(7):e0200137.
12. Harris KM, Kelly S, Berry E, et al. Systematic review of endoscopic ultrasound in gastro-oesophageal cancer. Health Technol Assess. 1998;2(18):1–134.
13. Bryant J, Picot J, Levitt G, et al. Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review. Health Technol Assess. 2007;11(27):1–84.
14. Brealey S, Scally A, Hahn S, et al. Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. Clin Radiol. 2005;60(2):232–41.
15. Gupta S, Maheshwari A, Parab P, et al. Neoadjuvant chemotherapy followed by radical surgery versus concomitant chemotherapy and radiotherapy in patients with stage IB2, IIA, or IIB squamous cervical cancer: a randomized controlled trial. J Clin Oncol. 2018;36(16):1548–55.
16. Brealey SD, DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team. Influence of magnetic resonance of the knee on GPs' decisions: a randomised trial. Br J Gen Pract. 2007;57(541):622–9.
17. Bartholomew AL, Denton ERE, Shaw M, et al. A randomised controlled trial comparing lateral skull computerised radiographs with or without a grid. Radiography. 2004;10:201–4.
18. Harrison R, Dey P, Slevin NJ, et al. Randomized controlled trial to assess the effectiveness of a videotape about radiotherapy. Br J Cancer. 2001;84(1):8–10.
19. Sala E, Watson CJE, Beardsmoore C, et al. A randomized controlled trial of routine early abdominal computed tomography in patients presenting with non-specific acute abdominal pain. Clin Radiol. 2007;62:961–9.
20. Ravasco P, Monteiro-Grillo I, Vidal PM, et al. Dietary counseling improves patient outcomes: a prospective, randomized controlled trial in colorectal cancer patients undergoing radiotherapy. J Clin Oncol. 2005;23:1431–8.
21. Slaar A, Karsten IHCM, Ludo FM, Beenena LFM, et al. Plain radiography in children with spoke wheel injury: a retrospective cohort study. Eur J Radiol. 2015;84:2296–300.
22. Damen J, Runhaar J, Kloppenburg M, et al. Additional value of different radiographic views on the identification of early radiographic hip and knee osteoarthritis and its progression: a cohort study. Arthritis Care Res. 2017;69(11):1644–50.
23. Trakada G, Pouli A, Goumas P. Pulmonary radiographic findings and mortality in hospitalized patients with lower respiratory tract infections. Radiography. 2006;12:20–5.
24. Aktas M, de Jong D, Nuyttens JJ, et al. Concomitant radiotherapy and hyperthermia for primary carcinoma of the vagina: a cohort study. Eur J Obstet Gynecol Reprod Biol. 2007;133:100–4.
25. Virtanen A, Pukkala E, Auvinen A. Angiosarcoma after radiotherapy: a cohort study of 332,163 Finnish cancer patients. Br J Cancer. 2007;97:115–7.
26. Jaremko JL, Lambert RGW, Rowe BH, et al. Do radiographic indices of distal radius fracture reduction predict outcomes in older adults receiving conservative treatment? Clin Radiol. 2007;62:65–72.
27. Zhang Y, Chen Y, Huang H, et al. Diagnostic radiography exposure increases the risk for thyroid microcarcinoma: a population-based case–control study. Eur J Cancer Prev. 2015;24(5):439–46.

28. Darby SC, Ewertz M, McGale P, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med. 2013;368:987–98.
29. Sernik RA, Abicalaf CA, Pimental BF, et al. Ultrasound features of carpal tunnel syndrome: a prospective case-control study. Skelet Radiol. 2008;37:49–53.
30. Cheng T-C, Jeng C-M, Kung C-H, et al. Yoga and lumbar disc degeneration disease: MR imaging based case control study. Chin J Radiol. 2008;33:73–8.
31. Spruit PH, Siesling S, Elferink MAG, et al. Regional radiotherapy versus an axillary lymph node dissection after lumpectomy: a safe alternative for an axillary lymph node dissection in a clinically uninvolved axilla in breast cancer. A case control study with 10 years follow up. Radiat Oncol. 2007;2:40.
32. Finlay GA, Joseph B, Rodrigues CR, et al. Advanced presentation of lung cancer in Asian immigrants: a case-control study. Chest. 2002;122(6):1938–43.
33. Nightingale H, Conroy R, Elliott T, Coyle C, Wylie JP, Choudhury A. A national survey of current practices of preparation and management of radical prostate radiotherapy patients during treatment. Radiography. 2017;23:87–93.
34. Snaith B, Harris MA, Palmer D. A UK survey exploring the assistant practitioner role across diagnostic imaging: current practice, relationships and challenges to progression. Br J Radiol. 2018;91:20180458.
35. Goense L, Borggreve AS, Heethuis SE, van Lier ALHMW, van Hillegersberg R, Mook S, et al. Patient perspectives on repeated MRI and PET/CT examinations during neoadjuvant treatment of esophageal cancer. Br J Radiol. 2018;91:20170710.
36. Lutz S, Spence C, Chow E, et al. Survey on use of palliative radiotherapy in hospice care. J Clin Oncol. 2004;22:3581–6.
37. Davies C, Grange S, Trevor MM. Radiation protection practices and related continuing professional education in dental radiography: a survey of practitioners in the north-east of England. Radiography. 2005;11:255–61.
38. Jones HC, Manning D. A survey to assess audit mechanisms practised by skeletal reporting radiographers. Radiography. 2008;14(3):201–5.
39. Power M, Laasch H, Kasthuri RS, et al. Videofluoroscopic assessment of dysphagia: a questionnaire survey of protocols, roles and responsibilities of radiology and speech and language therapy personnel. Radiography. 2006;12(1):26–30.
40. Yi A, Chang JM, Shin SU, Chu AJ, Cho N, Noh D-Y, et al. Detection of noncalcified breast cancer in patients with extremely dense breasts using digital breast tomosynthesis compared with full-field digital mammography. Br J Radiol. 2019;92:20180101.
41. Wooten WM, Shaffer LET, Hamilton LA. Bedside ultrasound versus chest radiography for detection of pulmonary edema: a prospective cohort study. J Ultrasound Med. 2019;38:967–73.
42. Grisaru D, Almog B, Levine C, et al. The diagnostic accuracy of 18F-Fluorodeoxyglucose PET/CT in patients with gynaecological malignancies. Gynecol Oncol. 2004;94:680–4.
43. Burling D, Moore A, Marshall M, et al. Virtual colonoscopy: effect of computer-assisted detection (CAD) on radiographer performance. Clin Radiol. 2008;63:549–56.
44. MERCURY Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. Br Med J. 2006;333:779.
45. Dai S, Hata K, Inubashiri E, et al. Does three-dimensional power Doppler ultrasound improve the diagnostic accuracy for the prediction of adnexal malignancy? J Obstet Gynaecol Res. 2008;34(3):364–70.
46. Nightingale JM, Murphy F, Eaton C, Borgen R. A qualitative analysis of staff-client interactions within a breast cancer assessment clinic. Radiography. 2017;23:38–47.
47. Nagle C, Lewis S, Meiser B, et al. Exploring general practitioners' experience of informing women about prenatal screening tests for foetal abnormalities: a qualitative focus group study. BMC Health Serv Res. 2008;28(8):114.
48. Poulos A, Llewellyn G. Mammography discomfort: a holistic perspective derived from women's experiences. Radiography. 2005;11(1):17–25.
49. Colyer H. The role of the radiotherapy treatment review radiographer. Radiography. 2000;6(4):253–60.

50. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). http://www.prisma-statement.org/. Accessed 17 May 2019.
51. CONSORT 2010 Statement (Consolidated Standards of Reporting Trials). http://www.consort-statement.org/. Accessed 17 May 2019.
52. The STROBE Statement (Strengthening the Reporting of Observational studies in Epidemiology). https://www.strobe-statement.org/index.php?id=strobe-home. Accessed 17 May 2019.
53. STARD (Standards for Reporting of Diagnostic Accuracy Studies) 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. Access via the EQUATOR network (Enhancing the Quality and Transparency of Health Research). http://www.equator-network.org/reporting-guidelines/stard/. Accessed 17 May 19.
54. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36.
55. Kelly S, Berry E, Roderick P, et al. The identification of bias in studies of the diagnostic performance of imaging modalities. Br J Radiol. 1997;70:1028–35.
56. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Access via the EQUATOR network (Enhancing the Quality and Transparency of Health Research). http://www.equator-network.org/reporting-guidelines/coreq/. Accessed 17 May 2019.

Further Readings

Azzam Al-Jundi A, Sakka S. Critical appraisal of clinical research. J Clin Diagn Res. 2017;11(5):JE01–5.

How to read a paper – links to articles by Trisha Greenhalgh on the BMJ website. https://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper. Accessed 19 May 2019.

Williams V, Boylan A-M, Nunan D. Critical appraisal of qualitative research: necessity, partialities and the issue of bias. https://doi.org/10.1136/bmjebm-2018-111132.