

Critical appraisal of quantitative and qualitative research literature

Tony Smith

University Department of Rural Health, Faculty of Health, University of Newcastle, Newcastle, New South Wales 2300, Australia.
Correspondence tony.smith@hnehealth.nsw.gov.au

Abstract Critical appraisal of research articles can be used to inform the design of new research studies. It can also be used by clinicians who wish to improve service quality by using the best available evidence to inform their practice. This paper describes a broad framework of critical appraisal of published research literature that covers both quantitative and qualitative methodologies.

The aim is the heart of a research study. It should be robust, concisely stated and specify a study factor, outcome factor(s) and reference population. Quantitative study designs, including sampling methods, can be ranked in order of the quality of the evidence they produce, with randomised control trials being ranked as level 1. The strength of evidence from qualitative research studies depends on the degree of rigour used in data collection and analysis, using techniques like theoretical sampling, triangulation and participant validation. Whatever the study design, it must be appropriate to address the aim of the study.

In critically appraising all research papers, there is a need to reflect on how well the conclusions flow logically from the results of the analysis and answer the original research question and how well the research applies to the population we are interested in.

Introduction

Reading critically and analysing the quality of research literature are skills that are used in designing valid and reliable research studies. Consequently, formal postgraduate research training includes an element of critical appraisal. Undergraduate Medical Radiation Science programmes also include a component of critical appraisal, linked to the generic graduate health professional attribute of “using research findings in clinical practice”¹ – evidence based practice. Critical appraisal of research literature is an essential skill for all members of the health care team, including diagnostic radiographers, radiation therapists and sonographers, in order to develop models of evidence based practice that focus on optimal outcomes. Maintenance and improvement of the quality and safety of health care demands the measurement of these outcomes. Hence, there is a further need for clinicians to have a grasp of research methodologies so that they can design and implement effective quality assurance programmes using methods that are both valid and reliable. This paper, therefore, aims to provide a framework for critical appraisal that is relevant to medical radiation professionals involved in research, education or clinical practice.

There is a need to understand the terms “validity” and “reliability” to appreciate the rationale behind critical appraisal. Validity can be broadly divided into “construct” (or “internal”) validity and “external” validity. Construct validity is the degree to which a study uses methods and measurement techniques that allow legitimate inferences to be made from the findings – Is the methodology sound? In fact, construct validity has a number of types (face validity, criterion validity, concurrent validity and so on). These are well explained by Trochim on his excellent web-site,² as are many other quantitative and qualitative research concepts. External validity relates to the question of whether, given the methodology used, it is reasonable to generalise the findings to other populations or settings. The term reliability, on

the other hand, refers to the “consistency” or “repeatability” of a study – Is the study reproducible? Well designed research, with good reliability, could be repeated at a different time or using a different population and give comparable results. Like validity, reliability has a number of types,² the best known being inter-rater or inter-observer reliability.

All research is not of equal quality in terms of validity and reliability. When reading a research paper we are justified in reading critically and questioning the findings, especially whether or not we should accept the conclusions drawn from the study and adopt the recommendations. However, being critical is a challenge for both clinicians and early career researchers, generally because they assume that the researchers must be more expert than themselves. There is a tendency to believe that any research study that has made it into print must be of the highest quality. This is not universally true. Editors and publishers choose papers for a variety of reasons. Furthermore, journals also vary in terms of the quality of the papers they publish and they are rated according to an impact factor.³ However, even journals with a high impact factor publish research of dubious quality, again, for various reasons. The first step in critically appraising a research article, therefore, is to reflect on the quality of the journal in which it is published. The second is to examine the track record of the authors in the particular field of study – Where are they from? What are their qualifications? Have they published in this field before? A search of Google Scholar or the Medline database can quickly answer these questions.

Some excellent resources are available to help develop critical appraisal skills. Concise, logically structured, analytical approaches are described by Darzins, *et al.*⁴ and by Greenhalgh.⁵ The latter wrote a series of papers on research methods in the *British Medical Journal* – a valuable resource! The critical appraisal framework described below synthesises the work of these and other authors.

Table 1: Various quantitative and qualitative study types, terminology and design considerations.

Quantitative		Qualitative	
Meta-analysis	Review of studies on a research question and hypothesis Stringent inclusion criteria (e.g. only RCTs – below) Uses statistics to combine samples and analyse results Increased sample size gives increased statistical power	Cross-sectional survey *	Subjects asked about behaviour, actions, experiences, etc. Self-administered (questionnaire) or structured interview “Snap shot” at one point in time Data from a large number of subjects but lacks depth
Systematic review	Review of literature focused on a research question Search strategy used that may include “grey” literature Structured critical appraisal techniques applied	Structured interview	Predetermined questions, as for questionnaires Predominantly closed-ended questions (limited responses) May also include some open-ended questions
Randomised Control Trial (RCT)	Prospective study design (experimental) Subjects randomly allocated to an intervention group (study factor) and control group (no intervention or placebo) Pre-determined time-frame and outcome factor(s)	Semi-structured interview	Questions around the topic and aim of the study Interview guide/schedule used but the wording is flexible Large amount of data (in-depth) from a small sample
Cohort study	Prospective and longitudinal study design Subjects with causative behaviour or activity (study factor) Control cohort does not engage in the same Subjects and controls compared for the outcome factor(s)	Unstructured interview	Broad topic of enquiry with minimal limitations Open-ended questions without categories Relies on deep interaction between interviewer and subject
Case-control study	Retrospective study design (non-experimental) Subjects have the condition or intervention (study factor) Controls have no intervention (may be matched to subjects) Cases and controls compared for outcome factor (s)	Focus groups	Groups of 6–10 subjects with some commonality Discuss an issue of common interest, with a moderator In-depth discussion and interaction between participants
Case study	Analysing outcomes of interesting or rare cases No statistical analysis Poor generalisability to populations	Observational study	Systematically watching interactions between individuals Recording physical features, behaviour, clothing, etc. May be at a particular location or in various settings
Longitudinal study	Observation or measurement over an extended period Data collected recurrently – e.g. 0, 3, 6, 12, 24 months Incorporates other study designs	Document analysis	Searching and reading related documents and records Extracting data around a particular research question Categorising data using comparative analysis techniques
Quasi-experimental study	Involves non-randomised study and control groups True experimental study is not possible (e.g. ethically) Includes pre- and post-intervention measurement	Narrative analysis	Stories give meaning and context to peoples' lives They give insight into behaviour, experiences, attitudes, etc. May use large units of data – biography or whole interview

* Surveys may also be used in quantitative studies, provided they yield quantifiable data.

Study aim

The “aim” is the heart of a research study. High quality studies have a robust and clearly stated aim that flows logically from the rationale for the study, and around which the study is designed. The aim under-pins the research question or questions. In quantitative research this will be stated in the form of a hypothesis that can be tested statistically. If the study is concerned with the impact of an intervention in the form of a treatment or diagnostic test, both the aim and the hypothesis will clearly identify the “study factor(s)” (the principal independent variable(s) being investigated and controlled by the researcher) and the “outcome factors” (the dependent variable(s) used to measure the results), including how they are measured. The reference population to which the study is relevant will also be defined.

In qualitative research, the aim and research question focus on “how?” and “why?” rather than “what?” and “how many?”, with the purpose of developing an integrated conceptual or theoretical understanding of an observed phenomenon.⁶ Ultimately, the theory or model will reflect the data from which it is derived. It is said to be grounded in the data and the most commonly used methodology is called “grounded theory”,⁷ as originally described by Barney Glaser and Anselm Strauss of the Chicago School of Sociology. This methodology has been used extensively in a variety of modified forms in health related research. Therefore, in general, the aim of

qualitative research is to use such methods to investigate, document and describe the knowledge, experiences, behaviour, opinions, values, attitudes and/or feelings of the individual study subjects in relation to a phenomenon.⁸

Study design

Table 1 lists a variety of study designs in both quantitative and qualitative research, listing the salient features of each. Quantitative study designs are ranked according to the “level of evidence” that is produced, as shown in Table 2. Randomised control trials produce the highest level of evidence. Qualitative research cannot be assigned a rank in the same way, according to the study design. The strength and quality of the evidence in qualitative research correlates closely with the degree of rigour applied in both data collection and analysis. The various techniques used to ensure rigour in qualitative research are described elsewhere⁸ and many of these, if not all, should be reflected in the description of the study design and methodology in a qualitative research article.

Whether quantitative or qualitative, the study design must be appropriate to address the aim of the study and answer the research question. Ask yourself – Is it appropriate? Are the study and outcomes factor(s) clearly defined? How are they measured? Do they target the critical variables? Are any important outcome factor(s) excluded? If so, why?

Table 2: Hierarchy of evidence.

Level of evidence	Type of quantitative study design
Level 1a	Systematic reviews or meta-analysis of randomised controlled trials (RCTs)
Level 1b	At least one RCT
Level 2a	At least one quasi-experimental clinical trial (i.e. without randomisation)
Level 2b	At least one other type of quasi-experimental study (e.g. cohort study)
Level 3	Non-experimental or descriptive comparative, correlational or case-control studies
Level 4	Expert committee reports and/or clinical experience of respected authorities
NICE	National Institute of Clinical Evidence Guidelines or health technology assessment
HSC	Health service circular(s)

Table 3: A range of random and non-random sampling methods.

Random sampling		Non-random sampling	
Simple	Sample chosen randomly from a population Equal possibility of being selected	Convenience	Subjects chosen by availability/presence (e.g. patients on a particular day)
Systematic	Population is ordered or ranked Sample at regular intervals (e.g. every 10th) until the sample size is reached Does not give an equal chance of selection	Purposive or Theoretical	Selection of subjects with specific traits (e.g. experienced and inexperienced) Preferred method for qualitative research Poor generalisability in quantitative (bias)
Stratified	Population grouped by a characteristic (e.g., male/female, inpatient/outpatient) Sample randomly and equally from groups Avoids unequal representation or bias	Snowballing	Subjects asked to nominate others who fit the inclusion criteria
Quota	Stratified with specific numbers per group Groups may be unequally represented	Volunteer	Canvassing or advertising for subjects Inviting people to fill-out a questionnaire
Cluster	Population divided into sub-populations or clusters (e.g. electorate, health service) Randomly select clusters as needed Include all individuals in selected clusters		

Sampling and sample size

Table 3 lists a variety sampling methods. A sample (size n) is drawn from a population of much larger size (N). Members of the sample share some commonality (e.g. a disease or condition) with each other and with the reference population and should thus be a reasonable representation of that population. Ask yourself – Is this the case? In quantitative research random sampling produces the strongest level of evidence. However, this may require a large amount of money, time and effort. Furthermore, recruitment, particularly of a control cohort, can be difficult and there are justifiably strong ethical constraints relating to experimental study designs. In fact, in reading any good research article, it should be possible to find a statement that the study has been approved by a human research ethics committee. If not, the validity must be questioned.

The necessary sample size in a quantitative study can be calculated using a formula based on the degree of error that will be tolerated in statistically testing the “null hypothesis”. In general terms, a null hypothesis states that “there is no statistically significant difference detected in the outcome factor(s) between the intervention and the control group”. Decisions are made about the level of “statistical significance” that will satisfy this statement (given the symbol α) and the acceptable level of “statistical power”. The former is the likelihood of a false positive result – finding a difference when none really exists – and the latter the

likelihood of a false negative – finding no significant difference when there actually is one. The level of significance is commonly set at 0.05 and the statistical power at 0.8 (80%), which means accepting a 5% chance of a false positive (α or type I error) and a 20% chance of a false negative (β or type II error). Both will be reported in a well written article about a well designed study.

In qualitative studies sample sizes are relatively small, with a preference for purposive sampling – choosing subjects because they possess particular knowledge, experience or other attributes. As the aim is to get as much depth as possible, choosing a limited number of subjects who have substantial experience makes sense, although, in some studies, the perspective of subjects with no experience may also be valuable. Ideally all perspectives and possible variables are accounted for, which is called “theoretical sampling”. While qualitative researchers are free to be selective about who they include in their study, they must justify the choices they make, providing a breakdown of the demographic and relevant background characteristics of their sample. Has this information been given? The sample size is not calculated, as in quantitative studies, but is limited by the number of subjects it takes to reach “data saturation” – that is, no new information about the topic is to be gained by further data collection.

Bias and confounders

A bias is a systematic error that has been introduced by the

researcher. For example, using purposive sampling in a quantitative study will introduce “selection” or “sampling bias” – choosing one type of study subject in preference to others. While purposive sampling is preferred in qualitative studies, it is still possible to have a selection bias that could intentionally distort the findings. There are also other forms of bias, such as that due to non-random assignment of subjects or due to incomplete follow-up. Ask yourself – Are there any sources of bias? Have they been acknowledged? How have they been controlled, if at all?

Confounding is a form of bias that is beyond the control of the researcher or has arisen unintentionally, perhaps without them being aware of it. Nevertheless, confounders are likely to influence the results one way or the other and may result in there being multiple explanations for the outcome, other than the study factor. Care should be taken to identify any possible confounders as they decrease construct validity and cast doubt over the results of a study.

Data analysis

In reports of both quantitative and qualitative research, data analysis and interpretation must be transparent and explained in enough detail that the study could be repeated. Ask yourself – Is this so? In quantitative studies, the statistical tests that have been used must be specified and the discerning reader should question whether they are appropriate to the type of data being analysed and for answering the research question. If necessary, expert opinion should be sought from a statistician.

The data in qualitative studies usually consists of words and their meaning, not numbers, and so the challenge is for the researcher to maintain objectivity. Data must be interpreted in context so that it accurately reflects the informants’ perspectives, not that of the researcher. There are various techniques used to ensure that data analysis is disciplined and rigorous. First, where necessary, the researcher declares their role in the research process, acknowledging any preconceptions they may hold. This is referred to as reflexivity. Triangulation is the process of using data from more than one source (e.g. different interest groups) or using multiple data collection methods (e.g. interviews and observation). Respondent or participant validation is where the study conclusions are reviewed by some of the study subjects, to validate the findings. Sound qualitative studies will include such techniques.

Data analysis occurs in parallel with data collection in qualitative research studies; otherwise it would not be possible to know when data saturation had been reached. Analysis informs subsequent data collection, such that the validity of early emergent themes and sub-themes is tested in later interviews, observation or focus groups. Data is analysed using comparative analysis of transcripts or other raw data (inductive analysis), with reference to what is known from the literature and other sources (deductive analysis). Both should be evident in the journal article as part of a logical process of clustering themes and subthemes and into categories and ultimately into a few key concepts.

Results and conclusions

In quantitative studies the results are expressed statistically, relative to the level of significance set for the null-hypothesis to be true, as described earlier. The results will be given as the measured values of the outcome factor(s) (e.g. sample mean) for both the intervention and control groups. Even if there is a measured difference, however, appropriate statistical tests must be performed to determine whether that difference is real or whether it is a

“spurious” finding that resulted by chance when the null-hypothesis is actually true⁹; that there is no difference. The probability that the null-hypothesis is true is calculated as the *P* value. If the *P* value is less than the level of significance (usually < 0.05, as above) there is a low probability that the null-hypothesis is true and it is therefore rejected in favour of the alternative hypothesis that the measured difference is real or “statistically significant”. If the *P* value is greater than the significance level (i.e. > 0.05, for example) there is an unacceptably high probability that the null-hypothesis is true. It cannot be rejected and it is concluded that there is no statistically significant difference. In some studies the significance level (α) is set at only 1% so that the *P* value has to be < 0.01 to reject the null-hypothesis, in which case the evidence of a real difference is stronger and the result is of “higher statistical significance”⁹.

When reporting the results of a quantitative study, the type of statistical tests used must be stated, the *p*-value must be given relative to the significance level ($P < 0.05$ or $P < 0.01$) and a confidence interval should also be given. The confidence interval (CI) is the range of values of the measurements that the researcher is confident includes the true, population value. For example, with a sample mean difference of 59.65 mm, the 95% CI might be 45.69–64.10 mm, meaning that there is a 95% chance that the mean for the population is within that range. The narrower the CI the greater strength of the finding, so it is important to look at the CI as well as the *P* value when considering whether to adopt an intervention into clinical practice.

In qualitative studies, there are no statistical measures against which to assess the results. However, qualitative research is not intended to draw definitive, universal truths but rather to examine, describe and conceptualise complex interactions and provide insight into the way individuals and groups construct their own, subjective reality. The results of the study will consist of direct quotations or other raw data that illustrate the findings, leading to logical conclusions that are traceable back to the original data. Derived themes and sub-themes merge to form a smaller number of clearly represented key concepts that can be displayed graphically as a conceptual model. The results of a well designed, well performed qualitative study will flow logically from the analysis.

Finally, in critically appraising all research papers, there is a need to reflect on how well the conclusions are supported by the results and whether the original research question is answered. Are the conclusions justified? Have any conclusions been overlooked? Has the author extrapolated too much from the findings? Have the limitations of the study been acknowledged? Are new research questions generated?

Summary

Critical appraisal skills can be acquired, provided the reader knows what to look for when reading a research article. This paper summarises the core knowledge necessary for the critical appraisal of both quantitative and qualitative research. While the methodology used in both types of research differs greatly, it is apparent that there are essential commonalities. Both must include a concisely stated research aim that embodies the research question(s) and the study design must be appropriate to address that aim. If the findings are to be applied in practice, the methods and techniques used must result in maximum possible validity and reliability, avoiding methodological pitfalls that degrade the quality of the evidence. Both should be read with a healthy degree of skepticism about the methods, results and conclusions, whether the purpose is to inform the design of one’s own research or the

implementation of new clinical protocols and procedures. We are justified in questioning whether the findings apply to the context of the population that we are interested in.

The author

Tony Smith PhD MSc BSc DipAppSci(MedRad) FIR

References

- 1 Strauss SE, Sackett DL. Getting research findings into practice: using research findings in clinical practice. *Br Med J* 1998; 317: 339–42.
- 2 Trochim WM. Research Methods Knowledge Base. Cornell University © 2006. Available online at: <http://www.socialresearchmethods.net/kb/contents.php> [verified 08/08/09].
- 3 Garfield E. The history and meaning of the journal impact factor. *JAMA* 2006; 295 (1): 90–3.
- 4 Darzins PJ, Smith BJ, Heller RF. How to read a journal article. *Med J Aust* 1992; 157 (6): 389–94.
- 5 Greenhalgh T. How to read a paper: assessing the methodological quality of published papers. *Br Med J* 1997; 315: 305–8.
- 6 Meadows KA. So you want to do research? An introduction to qualitative methods. *Br J Community Nurs* 2003; 8 (10): 464–9.
- 7 Strauss AL, Corbin J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks: Sage Publications; 1990.
- 8 Patton MQ. *Qualitative Research and Evaluation Methods*. 3rd Ed. Thousand Oaks: Sage Publications; 2002.
- 9 Daly LE, Bourke GJ, McGilvray J. *Interpretation and uses of medical statistics*. 4th Ed. Blackwell Scientific Publications. Oxford; 1991.