

Use of Reporting Guidelines in Scientific Writing: PRISMA, CONSORT, STROBE, STARD and Other Resources

Robyn L. Tate¹ and Jacinta Douglas²

¹ Rehabilitation Studies Unit, Northern Clinical School, University of Sydney, Australia

² Human Communication Sciences, La Trobe University, Australia

In this special article we describe a number of reporting guidelines endorsed by the CONSORT (Consolidated Standards of Reporting Trials) group for a range of research designs that commonly appear in scientific journals: systematic reviews, clinical trials with and without randomisation, observational studies, n-of-1 (or single-case experimental design) trials, and diagnostic studies. We also consider reporting guidelines for studies using qualitative methodology. In addition to reporting guidelines, we present method quality rating scales, which aim to measure risk of bias that threatens the internal validity of a study. Advantages of reporting guidelines and method quality rating scales for authors include the provision of a structure by which to improve the clarity and transparency of report writing; for reviewers and readers advantages include a method by which to critically appraise an article. *Brain Impairment* endorses these reporting guidelines and applies them within the review process for submissions to the journal.

Keywords: reporting guidelines, method quality rating scales, CONSORT, clinical trials, observational studies, n-of-1 trials, single-subject designs, qualitative research

The standard of scientific writing of reports addressing health conditions has been criticised. As a case in point, Dijkers and colleagues (2002) surveyed papers in six rehabilitation journals representing medicine, nursing, occupational therapy, physiotherapy and psychology over a 2-year period (1997–1998). Of the 1,039 peer-reviewed articles published in the journals, there were 646 articles describing original, quantitative research and 171 of the reports described intervention research. These 171 papers reported on 314 interventions using a total of 651 outcome measures. Dijkers et al. highlighted deficiencies in these reports: only 31% of papers reported any type of psychometric information on the outcome measures used and less than 50% of reports described the interventions in sufficient detail to allow replication by an experienced researcher or clinician.

To address the issue of inadequate reporting standards, specific reporting guidelines have been

developed for different types of research designs. The purpose of reporting guidelines is a very pragmatic one. They are intended to provide a brief, concise and practical ‘user’s guide’ to publishing reports. The guides include a descriptive checklist which contains a minimum core set of essential items that an author is expected to address when publishing an article. In other words, what needs to be reported and how it should be reported. It is recognised, however, that guidelines have a specific purpose, and in particular, they are not intended to:

- be a ‘how to do it’ manual (e.g., to stipulate that a specific statistical technique is appropriate for a particular dataset). That level of detail can only be dealt with in specialist articles or textbooks
- be evaluative in terms of making recommendations regarding the design and conduct of a study

Address for correspondence: Professor Robyn Tate, MPsychol., PhD, Rehabilitation Studies Unit, PO Box 6, Ryde NSW 1680, Australia. E-mail: rtate@med.usyd.edu.au

- provide a measure of the quality of the study (e.g., a method quality scale assessing internal validity)
- take the place of a reviewer who will evaluate the study from many perspectives (internal and external validity, clinical relevance, contribution to the literature, etc.).

Nonetheless such reporting guidelines have been demonstrated to be of great value in improving the reporting of health outcome research (Hopewell, Dutton, Yu, Chan, & Altman, 2010), and they also assist readers, reviewers and journal editors to critically appraise a report. Reporting guidelines generally publish a short article containing the checklist, as well as a more extensive ‘elaboration and explanation’ paper. The extended document provides rationale for inclusion of the checklist items, along with published extracts of model examples of item reporting.

In addition to reporting guidelines, there are also scales available that have a more specific purpose; namely to evaluate method quality of a report. Not all published articles are of equal quality and method quality rating scales provide one way to rate the adequacy of key features of a report. Method quality scales generally focus on risk of bias or threats to the internal validity of a study. A large number of method quality rating scales is available, although it must be said that some authorities have expressed reservations about their use (see Moher et al., 1995; Liberati et al., 2009, for discussion).

Since its inception, *Brain Impairment* has published papers with research designs covered by many of these reporting guidelines, including systematic reviews, randomised controlled trials (RCT), non-RCTs, single-participant intervention designs, observational studies (including cohort studies using longitudinal/follow-up designs), studies examining diagnostic tests, and qualitative studies. In the remainder of this editorial, we provide brief description of the pertinent guidelines for the common research designs that appear in *Brain Impairment*.

Systematic Reviews and the PRISMA Statement

A distinction is often drawn between the traditional narrative review and the systematic review, with the latter commonly regarded as superior (but see Dijkers (2009) for an interesting analysis of this view). As the name suggests, a systematic review aims to systematically and exhaustively examine the literature on a particular topic, using specific

methodology (designed to minimise risk of bias), which is fully explicated in the report. Results are synthesised and summarised, thereby making such reviews a valuable source of evidence. When there are sufficient data available, meta-analysis can be used to combine the results of independent studies and conduct quantitative analysis.

The QUOROM (Quality Of Reporting Of Meta-analyses) statement was the original reporting guideline for meta-analyses (Moher et al., 1999), but not all systematic reviews conduct meta-analysis and the QUOROM statement has now been replaced by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) statement. The statement is provided in Moher et al. (2009) and the explanation and elaboration document in Liberati et al. (2009). PRISMA is a 27-item checklist (see Table 1) that delineates the criteria an author should address ‘to ensure clear presentation of what was planned, done, and found in a systematic review’ (Liberati et al., p. e2). The 27 items cover all aspects of the report, including title (1 item), abstract (1 item), introduction (2 items), methods (12 items), results (7 items), discussion (3 items), and funding (1 item).

The PRISMA statement makes reference to the PICOS approach, which is a helpful process to articulate the objectives and research question of a review by formulating pertinent and precise questions regarding the *P*opulation of participants with the health condition, *I*nterventions or treatments being evaluated, *C*omparator or control condition (description of which is generally poor), *O*utcomes being measured, and *S*tudy design.

PRISMA is designed for systematic reviews of RCTs, yet not all treatment studies use randomisation. A reporting guideline specific to systematic review of ‘observational’ studies, such as those using data from existing databases, cross-sectional study, case series, case-control, cohort and historical controls, is available: Meta-analysis of Observational Studies in Epidemiology (MOOSE; Stroup et al., 2000).

Although systematic reviews provide Level 1 evidence (see <http://www.cebm.net/index.aspx?o=5653>), the adequacy of reporting of systematic reviews is variable. AMSTAR (Assessment of Multiple Systematic Reviews; Shea, Grimshaw et al., 2007) is an 11-item method quality scale designed for systematic reviews. It was carefully developed and content validity is supported by a statistical item reduction process and expert review. Inter-rater reliability is excellent for the total score ($k_w = .84$, 95% CI: .67–1.00), and moderate to perfect for the individual items ($K_w = .51$ –1.00), and the scale also shows evidence of discriminant validity

TABLE 1
Items from PRISMA Statement*

Section/Topic	Item	Description
Title		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
Abstract		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
Introduction		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
Methods		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Informational sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
Results		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see Item 12).

CONTINUED OVER

TABLE 1 (continued)
Items from PRISMA Statement*

Section/Topic	Item	Description
Results of individual studies	20	For all outcomes considered (benefits or harms), present for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
Discussion		
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy-makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
Funding		
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Note: *Reproduced from Moher et al. (2009).

(Shea, Bouter et al., 2007). Completion time for the AMSTAR is reported as 15 minutes per report.

Randomised Controlled Trials and the CONSORT Statement

The gold standard methodology for group intervention studies is the randomised controlled trial (RCT), which uses the principles of random and concealed allocation to groups, along with blinding (masking) of those receiving and delivering the intervention/s and assessing the outcomes. Randomisation is important because it allows inferences to be made about cause and effect. It eliminates selection bias and thus balances prognostic variables, both those that are known (e.g., age), as well as those that are unknown (perhaps socioeconomic status or exercise history, or previous therapy are predictors of successful outcome), because participants are assigned to groups by chance. The twin concepts of randomisation consist of (1) random allocation to groups, and (2) concealment of the randomisation schedule from the person who is responsible for recruiting participants, so that they cannot predict which treatment the participant will receive or the group to which the participant will be allocated.

The Consolidated Standards of Reporting Trials (CONSORT) statement is the reporting guideline for RCTs, and the updated version was published in 2010, both for the checklist (Schulz, et al., 2010) and the explanation and elaboration document (Moher et al., 2010). CONSORT 2010 is a 25-item checklist, although a number of items have multiple components, making 37 items in total (see Table 2). Like other statements produced by this research group, items cover all aspects of reporting, including title and abstract (2 items), introduction (2 items), methods (17 items), results (10 items), discussion (3 items), other information (3 items).

A distinctive feature of the CONSORT statement is the flow diagram to facilitate the reporting of the design and conduct of the study, and specifically to describe the flow of participants through different stages of the study. Many RCTs are subject to difficulties with participant flow, in terms of participants not receiving the treatment as allocated, attrition or loss to follow-up, or exclusion from analysis. All these factors can be easily and simply recorded by use of a diagram that is demonstrated in Moher et al. (2010) and reproduced in Figure 1.

TABLE 2

Items from CONSORT 2010 Statement*

Section/Topic	Item	Description
Title and abstract	1a	Identification as a randomised trial in the title.
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts [Hopewell et al., 2008]).
Introduction Background and objectives	2a	Scientific background and explanation of rationale.
	2b	Specific objectives or hypotheses.
Methods Trial design	3a	Description of trial design (such as parallel, factorial), including allocation ratio.
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons.
Participants	4a	Eligibility criteria for participants.
	4b	Settings and locations where the data were collected.
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered.
Outcomes	6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed.
	6b	Any changes to trial outcomes after the trial commenced, with reasons.
Sample size	7a	How sample size was determined.
	7b	When applicable, explanation of any interim analyses and stopping guidelines.
Randomisation sequence generation	8a	Method used to generate the random allocation sequence.
	8b	Type of randomisation; details of any restriction (such as blocking and block size).
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned.
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions.
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.
	11b	If relevant, description of the similarity of interventions.
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes.
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses.
Results Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome.
	13b	For each group, losses and exclusions after randomisation, together with reasons.
Recruitment	14a	Dates defining the periods of recruitment and follow-up.
	14b	Why the trial ended or was stopped.
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group.
Number analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

CONTINUED OVER

TABLE 2 (continued)
Items from CONSORT 2010 Statement*

Section/Topic	Item	Description
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory.
Harms	19	All important harms or unintended effects in each group (for specific guidance, see CONSORT for harms [Ioannidis et al., 2004]).
Discussion		
Limitations	20	Trial limitations; addressing sources of potential bias; imprecision; and, if relevant, multiplicity of analyses.
Generalisability	21	Generalisability (external validity, applicability) of the trial findings.
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.
Other information		
Registration	23	Registration number and name of trial registry.
Protocol	24	Where the full trial protocol can be accessed, if available.
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders.

Note: * Reproduced from Schulz et al. (2010).

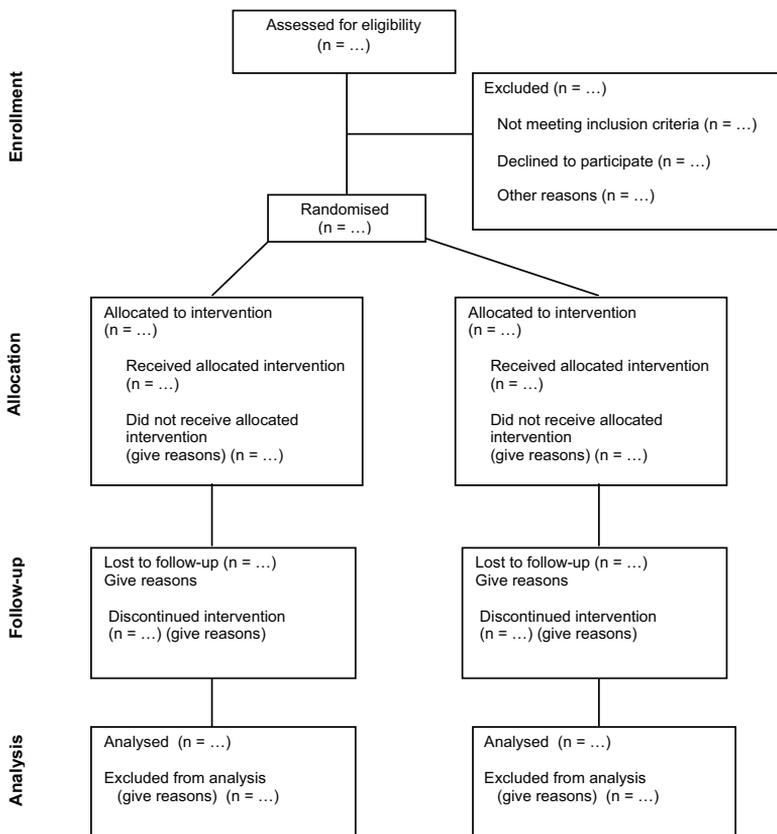


FIGURE 1
Participant flow diagram*
Note: * Reproduced from Moher et al. (2010)

A number of method quality rating scales of RCTs are available. A widely used scale is the PEDro scale, an 11-item scale that uses 10 items to report a method quality score. The PEDro scale has acceptable inter-rater reliability for the total score for consensus ratings (ICC = .68, 95% CI: .57 to .76) and individual items ($k = .50$ to .79; Maher, Sherrington, Herbert, Moseley, & Elkins, 2003).

It is recognised that RCTs are easier to conduct in drug trials than behavioural interventions. Sometimes it is not possible or feasible to conduct either random assignment (e.g., comparing a language therapy to examine its efficacy in anterior versus vs. posterior circulation left hemisphere stroke survivors) and/or blinding (e.g., surgeons need to be aware of the procedure they are performing; similarly therapists need to be aware of the treatment they are delivering). Nevertheless, the elements of randomisation and blinding are considered the most effective ways to minimise the risk of bias and its consequences (e.g., it has been demonstrated that treatment effect is inflated

in biased studies; Schulz et al., 1995). For these reasons, a CONSORT Extension for nonpharmacological studies was developed (Boutron et al., 2008) and is reproduced in Table 3. A corresponding 10-item method quality scale (CLEAR NPT) is also available (Boutron et al., 2005), which is reported to take approximately 10 minutes to complete. Parenthetically, the IMPACT group has recognised the particular challenges in conducting clinical trials and observational studies in the area of traumatic brain injury, and the interested reader is referred to their recommendations for the design of studies with this clinical group (Maas et al., 2010).

Single-Participant Designs and the CENT Statement

One difficulty researchers often encounter with the RCT is that large numbers of participants are generally required to ensure the trial has suffi-

TABLE 3

Items from the CONSORT Extension for Nonpharmacological Trials Statement*

Section	Item	Standard CONSORT description	Extension for nonpharmacologic trials
Title and abstract	1	How participants were allocated to interventions (e.g., 'random allocation', 'randomised', or 'randomly assigned').	In the abstract, description of the experimental treatment, comparator, care providers, centres, and blinding status.
Introduction Background	2	Scientific background and explanation of rationale.	
Methods Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	When applicable, eligibility criteria for centres and those performing the interventions.
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	Precise details of both the experimental treatment and comparator.
	4A		Description of the different components of the interventions and, when applicable, descriptions of the procedure for tailoring the interventions to individual participants.
	4B		Details of how the interventions were standardised. Details of how adherence of care providers with the protocol was assessed or enhanced.
Objectives	4C	Specific objectives and hypotheses.	

CONTINUED OVER

TABLE 3 (continued)

Items from the CONSORT Extension for Nonpharmacological Trials Statement*

Section	Item	Standard CONSORT description	Extension for nonpharmacologic trials
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	When applicable, details of whether and how the clustering by care providers or centres was addressed.
Randomisation sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).	When applicable, how care providers were allocated to each trial group.
Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11A	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment.	Whether or not those administering co-interventions were blinded to group assignment.
	11B†		If blinded, method of blinding and description of the similarity of interventions†.
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.	When applicable, details of whether and how the clustering by care providers or centres was addressed.
Results Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended)—specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome; describe protocol deviations from study as planned, together with reasons.	The number of care providers or centres performing the intervention in each group and the number of patients treated by each care provider or in each centre.
Implementation of intervention	New item		Details of the experimental treatment and comparator as they were implemented.
Recruitment	14	Dates defining the periods of recruitment and follow-up.	

CONTINUED

TABLE 3 (continued)

Items from the CONSORT Extension for Nonpharmacological Trials Statement*

Section	Item	Standard CONSORT description	Extension for nonpharmacologic trials
Baseline data	15	Baseline demographic and clinical characteristics of each group.	When applicable, a description of care providers (case volume, qualification, expertise, and so on) and centres (volume) in each group.
Numbers analysed	16	Number of participants (denominator) in each group included in each analysis and whether analysis was by 'intention-to-treat'; state the results in absolute numbers when feasible (e.g., 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	
Discussion Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.	In addition, take into account the choice of the comparator, lack of or partial blinding, and unequal expertise of care providers or centres in each group.
Generalisability	21	Generalisability (external validity) of the trial findings.	Generalisability (external validity) of the trial findings according to the intervention, comparators, patients, and care providers and centres involved in the trial.
Overall evidence	22	General interpretation of the results in the context of current evidence.	

Note: * Reproduced from Boutron et al. (2008). Additions or modifications to the CONSORT checklist.

CONSORT = Consolidated Standards of Reporting Trials.

† This item anticipates a planned revision in the next version of the standard CONSORT checklist.

cient power to detect a statistically significant difference. This can be a problem for many conditions with infrequent occurrence, and health conditions occurring with acquired brain impairment are one such area — it is unlikely that there will ever be an RCT to determine the effect of an intervention for prosopagnosia! There are also other limitations in the clinical application of results from RCTs, which have been described

by many authorities (see Guyatt et al., 1986; Wilson, 1987).

One research design that can overcome these problems is the n-of-1 trial (also known as the single-case experimental design, SCED). Traditionally, in mainstream medicine, these single-participant designs have either been ignored or (erroneously) dismissed as providing an inferior level of evidence. Due to the efforts of

a number of research teams, however, this state of affairs has been rectified. The Oxford Centre for Evidence-Based Medicine, one of the peak bodies in the field, has revised their levels of evidence and now the n-of-1 trial is introduced as Level 1 evidence for patient treatment decisions, on equal footing with the systematic review of multiple RCTs; a single RCT is classified as Level 2 evidence (Howick et al., 2011).

It needs to be emphasised that the n-of-1 trial and single-case experimental design should not be confused with a simple case report or description; rather, as we have written elsewhere, they 'differ radically from anecdotal, uncontrolled case descriptions. They represent the intensive and prospective study of the individual, using an a priori methodology, which includes systematic observation, manipulation of variables, repeated measurement and data analysis.' (Tate et al., 2008, p. 3). Not only does the single case experimental design (SCED) have methodological rigour, but there is also a range of statistical techniques designed specifically for the them (see Kazdin, 2011; Perdices & Tate, 2009).

An international research team, led by the University of Alberta, Canada, is currently developing reporting guidelines for n-of-1 trials, which will be known as the CONSORT Extension for N-of-1 Trials (CENT) statement (personal communication, S Vohra, March 11, 2011). A complementary team, led by the University of Sydney, Australia, is currently working on a similar guideline for behavioural intervention studies with single-participant designs. When published, the CENT statement will join the statements recommended by *Brain Impairment*.

To our knowledge, there is only a single method quality scale with demonstrated psychometric properties that has been published. The SCED Scale is an 11-item scale, which uses 10 items to provide a method quality score. The SCED Scale has good inter-rater reliability for the total score for consensus ratings (ICC = .88, 95% CI: .78 to .95) and fair-to-perfect reliability for individual items ($k = .48$ to 1.0 ; Tate et al., 2008). Administration time is generally 15 to 20 minutes. The SCED Scale has undergone minor revision including addition of a new item, and version 2 will be available shortly.

Observational Studies and the STROBE Statement

Most reporting guidelines focus on a single methodological design. The statement on Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) is special in

that it addresses multiple designs. The STROBE statement is reported in von Elm et al. (2007) and the explanation and elaboration document in Vandenbroucke et al. (2007). Three types of designs addressed by STROBE are *cohort studies* (using follow-up or longitudinal methods), *case-control designs* (e.g., for studies examining risk factors), and *cross-sectional research* (addressing incidence/prevalence issues). Although STROBE was not developed for *case reports* or *case series*, Vandenbroucke and colleagues note the applicability of a number of STROBE items to such designs. They suggest that authors/reviewers of case reports and case series may find the statement helpful.

The STROBE statement contains a 22-item checklist (see Table 4) addressing title and abstract (1 item), introduction (2 items), methods (9 items), results (5 items), discussion (4 items), and other information (1 item). The combined checklist addressing all three designs is presented in Table 4, but separate checklists for each of the designs are available from the STROBE website (<http://www.strobe-statement.org>).

A method quality rating scale suitable for observational studies (specifically cohort and case-control), as well as randomised trials, was developed by Downs and Black (1998). This method rating scale has a broad base in that it addresses external validity (specifically generalisation) as well as internal validity. The scale contains 26 items in four subscales: Reporting (10 items), Confounding (6 items), Bias (7 items) and External validity (3 items). Inter-rater reliability is moderate for the total score ($r = .75$), but that for the subscales ranges from good (Bias $r = .83$) to poor (External validity $r = -.14$). Administration time averages 20 to 25 minutes (range 10–45 mins).

Diagnostic Test Accuracy and the STARD Statement

The Standards for Reporting of Diagnostic Accuracy (STARD) initiative grew from concerns expressed by the Cochrane Diagnostic and Screening Test Methods Working Group at the Rome Cochrane Colloquium meeting in 1999. During this meeting the Working Group discussed the poor methodological and reporting quality of studies evaluating diagnostic tests and highlighted the importance of rigorous determination of the diagnostic accuracy of tests in the context of patient care. In line with the CONSORT process, the Working Group developed a checklist of items required in a report of a study of diagnostic accuracy. The resultant STARD statement (Bossuyt et al.,

TABLE 4
Items from STROBE Statement*

Section/Topic	Item	Description
Title and abstract	1a	Indicate the study's design with a commonly used term in the title or the abstract.
	1b	Provide in the abstract an informative and balanced summary of what was done and what was found.
Introduction Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported.
	3	State specific objectives, including any prespecified hypotheses.
Methods Study design	4	Present key elements of study design early in the paper.
	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.
Participants	6a	<i>Cohort study</i> — Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. <i>Case-control study</i> — Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. <i>Cross-sectional study</i> — Give the eligibility criteria, and the sources and methods of selection of participants.
	6b	<i>Cohort study</i> — For matched studies, give matching criteria and number of exposed and unexposed. <i>Case-control study</i> — For matched studies, give matching criteria and the number of controls per case.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.
Data sources/ measurement	8 ¹	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.
Bias	9	Describe any efforts to address potential sources of bias.
Study size	10	Explain how the study size was arrived at.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.
Statistical methods	12a	Describe all statistical methods, including those used to control for confounding.
	12b	Describe any methods used to examine subgroups and interactions.
	12c	Explain how missing data were addressed.
	12d	<i>Cohort study</i> — If applicable, explain how loss to follow-up was addressed. <i>Case-control study</i> — If applicable, explain how matching of cases and controls was addressed. <i>Cross-sectional study</i> — If applicable, describe analytical methods taking account of sampling strategy.
	12e	Describe any sensitivity analyses.
Results Participants	13a ¹	Report numbers of individuals at each stage of study — e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.
	13b ¹	Give reasons for nonparticipation at each stage.
	13c ¹	Consider use of a flow diagram.

CONTINUED NEXT PAGE

TABLE 4 (continued)
Items from STROBE Statement*

Section/Topic	Item	Description
Descriptive data	14a ¹	Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.
	14b ¹	Indicate number of participants with missing data for each variable of interest.
	14c ¹	<i>Cohort study</i> — Summarise follow-up time (e.g., average and total amount).
Outcome data	15 ¹	<i>Cohort study</i> — Report numbers of outcome events or summary measures over time <i>Case-control study</i> — Report numbers in each exposure category, or summary measures of exposure. <i>Cross-sectional study</i> — Report numbers of outcome events or summary measures.
	Main results	16a
16b		Report category boundaries when continuous variables were categorised.
16c		If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.
Other analyses	17	Report other analyses done — e.g., analyses of subgroups and interactions, and sensitivity analyses.
Discussion		
Key results	18	Summarise key results with reference to study objectives.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.
Generalisability	21	Discuss the generalisability (external validity) of the study results.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.

Note: * Reproduced from von Elm et al. (2007)

¹ Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

2003a) and its accompanying explanation and elaboration statement (Bossuyt et al., 2003b) were published simultaneously in several medical journals.

Studies of diagnostic accuracy require that the outcomes of a *test* for a target health condition are evaluated by comparison with the outcomes of an established *reference standard* for that health condition. Within the STARD statement, the term *test* is broadly defined and refers to any method of obtaining data about an individual's health condition: 'It includes information from history and physical examination, laboratory tests, imaging tests, function tests and histopathology' (Bossuyt et al., 2003a, pp. 1–2). The term 'health condition' covers any identifiable target condition that triggers a clinical action (e.g., further assessment, initiation or termination of intervention), the reference standard is the best available test/s or method/s for

establishing the presence or absence of the particular health condition, and the test being evaluated is termed the 'index test'. In the context of this framework, accuracy refers to the degree of agreement between the information from the index test and the reference standard in the target condition. Various diagnostic efficiency statistics (e.g., sensitivity and specificity, the odds ratio, the likelihood ratio, the area under receiver operator characteristic (ROC) curve, positive and negative predictive power, kappa and phi) can be applied in the evaluation of diagnostic accuracy. Streiner (2003) and Deeks (2001) provide useful reference summaries of diagnostic efficiency statistics.

The 25-item checklist developed as a result of the STARD initiative covers all aspects of reporting, including title, abstract and keywords (1 item), introduction (1 item), methods (11 items), results

(11 items), and discussion (1 item). A prototypical flowchart was also developed to present detailed information about the method of patient recruitment, the order of test execution and the number of patients present and classified at each stage of a diagnostic accuracy study. The documents are freely available without copyright restrictions and can be downloaded from the STARD website at <http://www.stard-statement.org/>. The checklist is reproduced in Table 5 and the flow diagram is reproduced in Figure 2.

Smidt et al. (2006) investigated inter-assessment, and intra- and inter-observer reproducibility of the STARD checklist. The overall reproducibil-

ity of the quality of reporting on diagnostic studies using the STARD items was high (inter-assessment agreement 85%) and the inter-assessment reliability was acceptable (intercorrelation coefficient, ICC = 0.79; 95% CI: 0.62 to 0.89). However, substantial disagreements were found on specific individual items. Smidt et al. concluded that these disagreements reflected difficulties evaluating the specific items due to poor reporting within the articles rather than differences in interpretation of the items by the reviewers. It should be emphasised that the content of the STARD checklist and flowchart is relevant to most studies examining the validity of

TABLE 5
Items From the STARD Checklist for the Reporting of Diagnostic Accuracy*

Section/Topic	Item	Description
Title/Abstract/ Keywords	1	Identify the article as a study of diagnostic accuracy (recommend MESH heading 'sensitivity and specificity').
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.
Methods		
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard.
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not specify how the participants were further selected.
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	The reference standard and its rationale.
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.
	10	The number training and expertise of the persons executing and reading the index tests and the reference standard.
	11	Whether or not the readers of the index test and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.
Statistical methods	12	Methods for calculating or comparing diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals).
	13	Methods for calculating test reproducibility, if done.
Results		
Participants	14	When study was done, including beginning and ending dates of recruitment.
	15	Clinical and demographic characteristics of the study population (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centres).

CONTINUED OVER

TABLE 5 (continued)

Items From the STARD Checklist for the Reporting of Diagnostic Accuracy*

Section/Topic	Item	Description
Test results	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).
	17	Time interval from the index tests to the reference standard, and any treatment administered between.
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.
	19	A cross-tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.
Estimates	20	Any adverse events from performing the index tests or the reference standard.
	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).
	22	How indeterminate results, missing responses and outliers of the index tests were handled.
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done.
	24	Estimates of test reproducibility, if done.
Discussion	25	Discuss the clinical applicability of the study findings.

Source: *<http://www.stard-statement.org/>

assessment instruments. Thus, the information can be applied to guide the preparation and review of any report concerning the development of an assessment instrument.

Assessment of method quality in systematic reviews of diagnostic accuracy studies is just as essential as it is in any other systematic review. Given the unique design features and efficiency statistics associated with diagnostic accuracy studies, it follows that a tool specific to diagnostic accuracy studies is required to evaluate the methodological quality of such studies. Numerous checklists for assessment of diagnostic accuracy studies have appeared in the literature over the years. However, the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) is the only one that has been developed systematically (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003) and evaluated (Whiting et al., 2006). This tool defines quality in terms of both internal and external validity of a study. It was developed based upon reviews of empirical evidence and a formal consensus procedure with a panel of experts in diagnostic research. The QUADAS checklist contains 14 items covering bias (9 items), variability (2 items) and reporting

(3 items). An 11-item version of the tool is recommended for quality assessment of studies in Cochrane Systematic Reviews of Diagnostic Accuracy (Reitsma et al., 2009). This 11-item Cochrane version of the checklist does not include the 3 items related to reporting in the original publication of the tool (Whiting et al., 2003).

Evaluation of the QUADAS revealed good agreement between reviewers and the final consensus ratings (85% to 91%; Whiting et al., 2006). At the individual item level, two items (item 13: uninterpretable results; item 14: withdrawals) were problematic and the guidelines for scoring these items were modified. QUADAS was found to be relatively quick to complete taking less than 30 minutes by 85% of reviewers. More recently, Mann, Hewitt and Gilbody (2009) examined the utility of QUADAS when assessing diagnostic accuracy studies using psychometric instruments in the clinical context of postnatal depression. The results were broadly congruent with those of the original evaluation (Whiting et al., 2006) with good overall level of agreement (85.7%) and good agreement (> 80%) on the majority of items.

General example

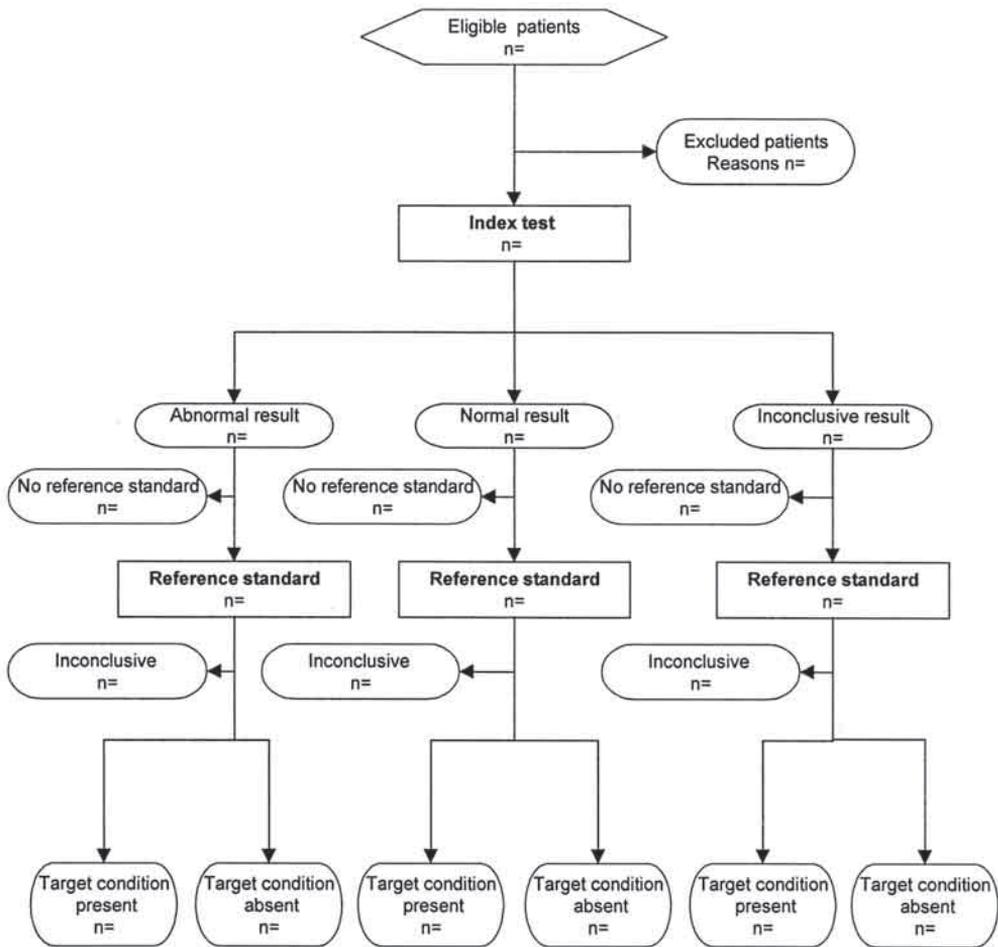


FIGURE 2

Prototypical flow diagram of a diagnostic accuracy study.

Source: <http://www.stard-statement.org/>

Qualitative Studies and Reporting Guidelines

Qualitative research can be defined broadly as naturalistic, social inquiry that focuses on understanding the way people interpret their experiences and their world (Denzin & Lincoln, 1994; Malterud, 2001). Qualitative research has not emerged from a single tradition. Consequently, there are multiple approaches to qualitative research reflecting diverse philosophies. In the context of the healthcare literature, the most fre-

quently encountered approaches are ethnography, grounded theory, phenomenology, case study and action research (Cohen & Crabtree, 2006). Common methods of data collection include interviewing (in-depth interviews and focus groups), participant and nonparticipant observation and collecting documents and artefacts. Along with this range of data collection methods, there is also a range of data analysis techniques; those most frequently used in health research include narrative analysis, qualitative content

analysis, thematic analysis and the constant comparative method of grounded theory.

Opinions regarding the utility of having a set or checklist of criteria for assessing qualitative research are not unified and several perspectives have been expressed in the literature. These perspectives range from the belief that the breadth and diversity of qualitative research prohibits the development of a single set of criteria to the belief that a set of well-defined, appropriate standards and criteria is needed to avoid the risk of quantitative standards and criteria being applied erroneously to qualitative research (See Cohen and Crabtree, 2008 for a discussion of this issue). Nevertheless, there are several sources of guidance for reporting qualitative research. Some of these guidelines have been developed to apply to specific data collection methods (e.g., COREQ a reporting checklist for in-depth interviews and focus groups developed by Tong, Sainsbury & Craig, 2007), while others have been developed to be inclusive of work emerging from the broad diversity of qualitative traditions. Two sets of guidelines developed to support authors and reviewers across the range of qualitative studies are summarised in Table 6. Both these sets of guidelines provide excellent guidance for authors and reviewers across four broad areas: (1) relevance of the study question, (2) appropriateness of the qualitative approach, (3) rigour of data collection and sampling procedures, and (4) interpretation and discussion. The first set was developed by Malterud (2001) and includes 30 items to guide the review of all aspects of manuscripts reporting qualitative studies: aim (3 items), reflexivity (1 item), method and design (2 items), data collection and sampling (5 items), theoretical framework (3 items), analysis (4 items), findings (4 items), discussion (5 items), presentation (2 items), and references (1 item). This guideline can be found in the original publication (Malterud, 2001, p. 485) and is reproduced on the Qualitative Research Guidelines Project website: <http://www.qualres.org/index.html>.

The second qualitative research reporting guideline was developed by Clark (2003) and is referred to by the acronym RATS, derived from the broad domains covered by the guideline: Relevance of study question, Appropriateness of qualitative method, Transparency of procedures, and Soundness of interpretive approach. Clark's guidelines with minor modifications are available through the EQUATOR (Enhancing the Quality and transparency of health research) network library for health research reporting: <http://www.equator-network.org/>. This guideline lists 22 items

across core areas that are expected to be covered in a qualitative report: relevance of study question (2 items), appropriateness of qualitative method (1 item), sampling (2 items), recruitment (2 items), data collection (3 items), role of researchers (1 item), ethics (3 items), analysis (4 items), and discussion and presentation (4 items). Additional information specific to the methodological diversity within qualitative research is provided in the method (item 3), sampling (item 5), and analysis (item 15) areas. This additional information can be particularly useful for the less experienced author or reviewer of qualitative reports.

Assessment of method quality in qualitative research has given rise to substantial discussion in the literature over the past decade (e.g., Cohen & Crabtree, 2008; Dixon-Woods et al., 2007; Mays & Pope, 2000; Noyes, Popay, Pearson, Hannes, & Booth, 2011; Walsh & Downe, 2006) and numerous instruments have been developed to support quality appraisal of qualitative research (Noyes et al. 2011). Spencer, Ritchie, Lewis, and Dillon (2003) developed a framework of quality criteria based on comprehensive review of qualitative methodological literature and existing guidelines as well as expert consensus. Cognisant of the broad church of qualitative research, these authors adopted an 'elemental' approach to quality appraisal whereby the premises underpinning social enquiry informed the criteria covered by the framework. Consequently, this framework is relevant for a range of qualitative methods. The framework consists of 18 appraisal questions covering findings (5 questions), design (1 question), sample (2 questions), data collection (1 question), analysis (4 questions), reporting (2 questions), reflexivity and neutrality (1 question), ethics (1 question), and auditability (1 question). Each question is accompanied by a set of quality indicators to guide the appraisal. This framework can be downloaded from the United Kingdom Government Social Research Service website: http://www.civilservice.gov.uk/Assets/a_quality_framework_tcm6-7314.pdf

Critical appraisal checklists for qualitative research have also been incorporated into software developed by evidence synthesis organisations. For example, the JBI-QARI (Qualitative Assessment and Review Instrument) was prepared for the Joanna Briggs Institute (JBI: <http://www.jbiconnect.org>). JBI-QARI is a web-based database that incorporates a critical appraisal scale, data extraction forms, and a data synthesis function (Pearson, 2003). It facilitates the inclusion of the findings of qualitative research studies as part of a systematic review of evidence.

TABLE 6
Guidelines for Reporting Qualitative Research

		Malterud (2001)		RATS (Clark, 2003)	
Section	Item	Description	Section	Item	Description
Relevance	1	Research question is relevant.	Relevance of study question	1	Research question is interesting.
	2	Aim is sufficiently focused and stated clearly.		2	Research question is relevant.
	3	Title gives clear account of the aim.			
Appropriateness of the approach	4	Researcher's motives, background, perspectives and preliminary hypotheses presented.	Appropriateness of qualitative method	3	Qualitative method chosen is the best approach for the aims: <ul style="list-style-type: none"> • Interviews: experience, perceptions, behaviour, practice, process • Focus groups: group dynamics, convenience, non-sensitive topics • Ethnography: culture, organisation behaviour, interaction • Textual analysis: documents, art, representations, conversations.
	5	Qualitative research methods suit the research question.			
	6	Best method for the question has been chosen.			
Data collection and sampling procedures	7	Strategy for data collection clearly stated (usually purposive or theoretical).	Transparency of procedures	4	Participants selected are the most appropriate to access the type of knowledge sought.
	8	Reasons for choice stated.		5	Sampling strategy is appropriate: <ul style="list-style-type: none"> • Theoretical: based on preconceived or emergent theory • Purposive: diversity of opinion • Volunteer: feasibility, hard-to-reach groups
	9	Best approach for the question has been chosen			
	10	Consequences for chosen strategy are discussed and compared with other options.			
	11	Characteristics of the sample presented in sufficient detail to understand the site and context of the study.			
			Recruitment	6	Recruitment conducted using appropriate methods. Selection bias is not evident.
			Data collection	8	Data collection was systematic and comprehensive.
				9	Characteristics of the study group & setting are clear.
				10	When and why data collection was stopped is stated and justified.
			Role of researchers	11	The researchers' roles and potential impact on the study and results are acknowledged.
			Ethics	12	Informed consent was sought and granted.
				13	Participants' anonymity and confidentiality ensured.
				14	Approval from an appropriate ethics committee was received.

CONTINUED OVER

TABLE 6 (continued)
Guidelines for Reporting Qualitative Research

		Maltzerud (2001)		RATS (Clark, 2003)	
Section	Item	Description	Section	Item	Description
Interpretation and discussion	12	Perspectives and ideas used for data interpretation are presented.	Soundness of interpretive approach	15	The type of analysis is appropriate for the type of study: <ul style="list-style-type: none"> • Thematic: exploratory, descriptive, hypothesis generating • Constant comparison/grounded theory: theory generating, analytical.
	13	Framework is adequate with respect to aim.			
Analysis	14	Role of theoretical framework is accounted for during analysis.		16	Interpretations are presented clearly and supported adequately by the evidence.
	15	Principles and procedures for data organisation and analysis are fully described.			
	16	How categories were identified is stated clearly (from theories or preconception in advance or developed from the data).			
	17	Principles used to organise the presentation of findings are identified.			
Findings	18	Strategies used to validate results are presented (cross-checks for rivaling explanations, member checks, or triangulation).		17 18	Quotes are used appropriately and effectively. Trustworthiness/reliability of the data and interpretations is established (audit trail, triangulation, member checking, independent analyst review of data and themes).
	19	Findings are relevant with respect to the aim.			
	20	Findings provide new insight.			
	21	Presentation well organised and best suited to ensure that findings are drawn from systematic analysis, rather than preconceptions.			
Discussion	22	Quotes are used adequately to support and enrich the researcher's synopsis of the patterns identified by systematic analysis.	Discussion and presentation	19 20 21 22	Findings are sufficiently grounded in a theoretical or conceptual framework. Previous knowledge and how the findings add to it are taken into account. Limitations are considered. The manuscript is well-written and accessible.
	23	Questions about internal validity, external validity, and reflexivity are addressed.			
	24	The design is scrutinised.			
	25	Shortcomings are accounted for and discussed.			
	26	Findings are compared with appropriate theoretical and empirical references.			
	27	Clear consequences of the study are proposed.			
Presentation	28	The report is easy to understand and clearly contextualised.		22	
	29	The voices of the informants and the researcher are distinguishable.			
References	30	Important and specific sources in the field are covered, presented and applied in the text.			

The JBI-QARI critical appraisal instrument consists of 10 items focusing on 3 areas: philosophical and methodological congruity (5 items), explication of researcher biases (2 items), and representation of the participants' perspectives and their relationship to the conclusions (3 items).

Current opinion regarding the value of and need for formal quality appraisal tools in relation to qualitative research continues to be divided. However, there is little doubt that evidence from qualitative studies can make an important contribution to systematic reviews across healthcare practice. Further, few would take exception to the need to establish the quality of evidence before using it to inform clinical practice. Thus, it is not surprising that methods to assist the critical appraisal of qualitative studies continue to appear in the literature. The reader who is interested in following developments in this area is referred to the Cochrane Qualitative Research Methods Group (QRMG): www.joannabriggs.edu.au/cqrmg.

Concluding Remarks

The reporting guidelines described in the foregoing are those that commonly relate to papers published in *Brain Impairment*, and *Brain Impairment* endorses these reporting guidelines. Reviewers and editors will use these standards when assessing submissions to the journal and authors are strongly encouraged to adhere to the items of the relevant reporting guideline when preparing a manuscript.

Although reporting of ethical approval to conduct a study is not included in all these checklists, it is recognised that ethics approval is generally necessary to conduct research on human participants (as well as animals) and *Brain Impairment* requires that authors include a statement about ethics approval to conduct their study.

As editors, we agree with the position of the developers of STROBE: by using the above reporting guidelines for articles submitted to *Brain Impairment* we do not suggest that the guideline structure needs to be followed 'in a rigid format. The checklist items should be addressed in sufficient detail and with clarity somewhere in an article, but the order and format for presenting information depends on author preferences, journal style, and the traditions of the research field' (von Elm et al., 2007, pp. 1455–1456).

Reporting guidelines are recognised by their developers to be evolving documents, and they require revisions and refinement in the light of new information. There are many other reporting guidelines in addition to those described in this special

editorial and the interested reader is referred to EQUATOR Network (http://www.equator-network.org/resource_centre/library-of-health-research-reporting/), established in 2008, which is a useful resource including a library of reporting guidelines.

References

- Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., ... de Vet, H.C.W. (2003a). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Clinical Chemistry*, 49(1), 1–6.
- Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., ... Lijmer, J.G. (2003b). The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clinical Chemistry*, 49(1), 7–18.
- Boutron, I., Moher, D., Altman, D.G., Schulz, K.F., Ravaud, P., for the CONSORT Group. (2008). Extending the CONSORT Statement to randomised trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine*, 148, 295–309.
- Boutron, I., Moher, D., Tugwell, P., Giraudeau, B., Poiraudeau, S., Nizard, R., & Ravaud P. (2005). A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed under consensus. *Journal of Clinical Epidemiology*, 58, 1233–1240.
- Clark, J.P. (2003). How to peer review a qualitative manuscript. In F. Godlee & T. Jefferson (Eds.) *Peer Review in Health Sciences* (2nd ed., pp. 219–235). London: BMJ Books.
- Cohen, D., & Crabtree, B. (2006). Qualitative Research Guidelines Project. Retrieved from <http://www.qualres.org/index.html>.
- Cohen, D., & Crabtree, B. (2008). Evaluative criteria for qualitative research in health care: controversies and recommendations. *Annals of Family Medicine*, 6, 331–339.
- Deeks, J.J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, 323, 157–162.
- Denzin, N.K., & Lincoln, Y.S. (1994). *Handbook of qualitative research*, Thousand Oaks: Sage.
- Dijkers, M. (2009). The value of 'traditional' reviews in the era or systematic reviews. *American Journal of Physical Medicine and Rehabilitation*, 88, 423–430.
- Dijkers, M., Kropp, G.C., Esper, R.M., Yavuzer, G., Cullen, N., & Bakdali, Y. (2002). Quality of intervention research reporting in medical rehabilitation journals. *American Journal of Physical Medicine and Rehabilitation*, 81, 21–33.
- Dixon-Woods, M., Sutton, A., Shaw, R., Miller, T., Smith, J., & Young, B. (2007). Appraising qualitative research for inclusion in systematic reviews: A quantitative and qualitative comparison of three

- methods. *Journal of Health Service Research and Policy*, 12, 42–7.
- Downs, S.H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52, 377–384.
- Guyatt, G., Sackett, D., Taylor, R.W., Chong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapy — randomised trials in individual patients. *The New England Journal of Medicine*, 314(14), 889–892.
- Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D.G. ... CONSORT Group. (2004). CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*, 371, 281–283.
- Hopewell, S., Dutton, S., Yu, L.M., Chan, A.W., & Altman, D.G. (2010). The quality of reports of randomised trials in 2000 and 2006: Comparative study of articles indexed in PubMed. *British Medical Journal*, 340, 1–8.
- Howick, J., Chalmers, I., Glasziou, P., Greenhaigh, T., Heneghan, C., Liberati, A., ... Thornton, H. (2011). The 2011 Oxford CEbm evidence table (introductory document). Oxford Centre for Evidence-Based Medicine. Retrieved from <http://www.cebm.net/index.aspx?o=5653>
- Ioannidis, J.P.A., Evans, S.J.W., Gøtzsche, P.C., O'Neil, R.T., Altman, D.G., Schulz, K., Moher, D., for the CONSORT Group. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT Statement. *Annals of Internal Medicine*, 141(10), 781–788; W–147 to W–151.
- Kazdin, A.E. (2011). *Single-case research designs. Methods for clinical and applied settings* (2nd ed.). Oxford: Oxford University Press.
- Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7), 1–28.
- Maas, A.I.R., Steyerberg, E.W., Marmarou, A., McHugh, G.S., Lingsma, H.R., Butcher, I., ... Murray, G.D. (2010). IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. *Neurotherapeutics*, 7, 127–134.
- Maher, C.G., Sherrington, C., Herbert, R.D., Moseley, A.M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of RCTs. *Physical Therapy*, 83, 713–721.
- Malterud, K. (2001). Qualitative research: standards, challenges and guidelines. *The Lancet*, 358, 483–488.
- Mann, R., Hewitt, C.E., Gilbody, S.M. (2009). Assessing the quality of diagnostic studies using psychometric instruments: Applying QUADAS. *Social Psychiatry and Psychiatric Epidemiology*, 44, 300–307.
- Mays, N., & Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *BMJ*, 320, 50–52.
- Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D.F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials. The QUOROM statement: Quality of reporting of meta-analyses. *Lancet*, 354, 1896–1900.
- Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., ... Altman, D.G. (2010). CONSORT 2010 Explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, 63(8), e1–e37.
- Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomised controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, 62–73.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), e1000097.
- Noyes, J., Popay, J., Pearson, A., Hannes, K., & Booth, A. (2011). Qualitative research and Cochrane reviews. In J.P.T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions version 5.1.0* (updated March 2011). The Cochrane Collaboration. Retrieved from www.cochrane-handbook.org.
- Pearson, A. (2003). Balancing the evidence: Incorporating the synthesis of qualitative data into systematic reviews. *JBIR Reports*, 1, 3.
- Perdices, M., & Tate, R.L. (2009). Single-subject designs as a tool for evidence-based clinical practice: are they unrecognized and undervalued? *Neuropsychological Rehabilitation*, 19(6), 904–927.
- Reitsma, J.B., Rutjes, A.W. S., Whiting, P., Vlassov, V.V., Leeflang, M.M.G., Deeks, J.J. (2009). Assessing methodological quality. In J.J. Deeks, P.M. Bossuyt & C. Gatsonis (Eds.), *Cochrane handbook of systematic reviews of diagnostic test accuracy* (Version 1.0.0). The Cochrane Collaboration, 2009. Retrieved from <http://srdta.cochrane.org>
- Shea, B.J., Bouter, L.M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., & Grimshaw, J.M. (2007). External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS ONE*, 12, e1350.
- Shea, B.J., Grimshaw, J.M., Wells, G.A., Boers, M., Andersson, N., Hamel, C., ... Bouter, L.M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7(10).
- Shultz, K.F., Chalmers, I., Hayes, R.J., & Altman, D.G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, 273, 408–412.

- Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, 152(11), 1–11.
- Smidt, N., Rutjes, A.W.S., van der Windt, D.A.W.M., Ostelo, R.W.J.G., Bossuyt, P.M., Reitsma, J.B., ... De Vet, H.C.W. (2006). Reproducibility of the STARD checklist: An instrument to assess the quality of reporting of diagnostic accuracy studies. *BMC Medical Research Methodology*, 6(12). doi: 10.1186/1471-2288-6-12.
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in qualitative evaluation: A framework for assessing research evidence*. London: Government Chief Social Researcher's Office.
- Streiner, D. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, 81(3), 210–219.
- Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Wilkinson, G.D., Rennie, D., ... Thacker, S.B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-Analysis of Observational Studies in Epidemiology (MOOSE) Group. *Journal of the American Medical Association*, 283(15), 2008–2012.
- Tate, R.L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation*, 18(4), 385–401.
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357.
- Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., ... Egger, M. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *Annals of Internal Medicine*, 147(8), W163–W194.
- von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., for the STROBE Initiative. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, 370, 1453–1457.
- Walsh, D., & Downe, S. (2006). Appraising the quality of qualitative research. *Midwifery*, 22, 108–119.
- Whiting, P., Rutjes, A.W. S., Reitsma, J.B., Bossuyt, P.M. M., & Kleijnen, J. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 3(25). Retrieved from: <http://www.biomedcentral.com/1471-2288/3/25>.
- Whiting, P.F., Weswood, M.E., Rutjes, A.W. S., Reitsma, J.B., Bossuyt, P.N. M., & Kleijnen, J. (2006). Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology*, 6(9). doi: 10.1186/1471-2288-6-9.
- Wilson, B. (1987). Single-case experimental designs in neuropsychological rehabilitation. *Journal of Clinical and Experimental Neuropsychology*, 9(5), 527–544.