

Research article

Open Access

A systematic review of the content of critical appraisal tools

Persis Katrak¹, Andrea E Bialocerkowski², Nicola Massy-Westropp¹, VS Saravana Kumar¹ and Karen A Grimmer*¹

Address: ¹Centre for Allied Health Evidence: A Collaborating Centre of the Joanna Briggs Institute, City East Campus, University of South Australia, North Terrace, Adelaide, 5000, Australia and ²School of Physiotherapy, The University of Melbourne, Melbourne, 3010, Australia

Email: Persis Katrak - pktrak@hotmail.com; Andrea E Bialocerkowski - aebial@unimelb.edu.au; Nicola Massy-Westropp - mwestropp@picknowl.com.au; VS Saravana Kumar - saravana.kumar@unisa.edu.au; Karen A Grimmer* - karen.grimmer@unisa.edu.au

* Corresponding author

Published: 16 September 2004

Received: 10 May 2004

BMC Medical Research Methodology 2004, 4:22 doi:10.1186/1471-2288-4-22

Accepted: 16 September 2004

This article is available from: <http://www.biomedcentral.com/1471-2288/4/22>

© 2004 Katrak et al; licensee BioMed Central Ltd.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Consumers of research (researchers, administrators, educators and clinicians) frequently use standard critical appraisal tools to evaluate the quality of published research reports. However, there is no consensus regarding the most appropriate critical appraisal tool for allied health research. We summarized the content, intent, construction and psychometric properties of published, currently available critical appraisal tools to identify common elements and their relevance to allied health research.

Methods: A systematic review was undertaken of 121 published critical appraisal tools sourced from 108 papers located on electronic databases and the Internet. The tools were classified according to the study design for which they were intended. Their items were then classified into one of 12 criteria based on their intent. Commonly occurring items were identified. The empirical basis for construction of the tool, the method by which overall quality of the study was established, the psychometric properties of the critical appraisal tools and whether guidelines were provided for their use were also recorded.

Results: Eighty-seven percent of critical appraisal tools were specific to a research design, with most tools having been developed for experimental studies. There was considerable variability in items contained in the critical appraisal tools. Twelve percent of available tools were developed using specified empirical research. Forty-nine percent of the critical appraisal tools summarized the quality appraisal into a numeric summary score. Few critical appraisal tools had documented evidence of validity of their items, or reliability of use. Guidelines regarding administration of the tools were provided in 43% of cases.

Conclusions: There was considerable variability in intent, components, construction and psychometric properties of published critical appraisal tools for research reports. There is no 'gold standard' critical appraisal tool for any study design, nor is there any widely accepted generic tool that can be applied equally well across study types. No tool was specific to allied health research requirements. Thus interpretation of critical appraisal of research reports currently needs to be considered in light of the properties and intent of the critical appraisal tool chosen for the task.

Background

Consumers of research (clinicians, researchers, educators, administrators) frequently use standard critical appraisal tools to evaluate the quality and utility of published research reports [1]. Critical appraisal tools provide analytical evaluations of the quality of the study, in particular the methods applied to minimise biases in a research project [2]. As these factors potentially influence study results, and the way that the study findings are interpreted, this information is vital for consumers of research to ascertain whether the results of the study can be believed, and transferred appropriately into other environments, such as policy, further research studies, education or clinical practice. Hence, choosing an appropriate critical appraisal tool is an important component of evidence-based practice.

Although the importance of critical appraisal tools has been acknowledged [1,3-5] there appears to be no consensus regarding the 'gold standard' tool for any medical evidence. In addition, it seems that consumers of research are faced with a large number of critical appraisal tools from which to choose. This is evidenced by the recent report by the Agency for Health Research Quality in which 93 critical appraisal tools for quantitative studies were identified [6]. Such choice may pose problems for research consumers, as dissimilar findings may well be the result when different critical appraisal tools are used to evaluate the same research report [6].

Critical appraisal tools can be broadly classified into those that are research design-specific and those that are generic. Design-specific tools contain items that address methodological issues that are unique to the research design [5,7]. This precludes comparison however of the quality of different study designs [8]. To attempt to overcome this limitation, generic critical appraisal tools have been developed, in an attempt to enhance the ability of research consumers to synthesise evidence from a range of quantitative and or qualitative study designs (for instance [9]). There is no evidence that generic critical appraisal tools and design-specific tools provide a comparative evaluation of research designs.

Moreover, there appears to be little consensus regarding the most appropriate items that should be contained within any critical appraisal tool. This paper is concerned primarily with critical appraisal tools that address the unique properties of allied health care and research [10]. This approach was taken because of the unique nature of allied health contacts with patients, and because evidence-based practice is an emerging area in allied health [10]. The availability of so many critical appraisal tools (for instance [6]) may well prove daunting for allied health practitioners who are learning to critically appraise

research in their area of interest. For the purposes of this evaluation, allied health is defined as encompassing "...all occasions of service to non admitted patients where services are provided at units/clinics providing treatment/counseling to patients. These include units primarily concerned with physiotherapy, speech therapy, family planning, dietary advice, optometry occupational therapy..." [11].

The unique nature of allied health practice needs to be considered in allied health research. Allied health research thus differs from most medical research, with respect to:

- the paradigm underpinning comprehensive and clinically-reasoned descriptions of diagnosis (including validity and reliability). An example of this is in research into low back pain, where instead of diagnosis being made on location and chronicity of pain (as is common) [12], it would be made on the spinal structure and the nature of the dysfunction underpinning the symptoms, which is arrived at by a staged and replicable clinical reasoning process [10,13].
- the frequent use of multiple interventions within the one contact with the patient (an occasion of service), each of which requires appropriate description in terms of relationship to the diagnosis, nature, intensity, frequency, type of instruction provided to the patient, and the order in which the interventions were applied [13]
- the timeframe and frequency of contact with the patient (as many allied health disciplines treat patients in episodes of care that contain multiple occasions of service, and which can span many weeks, or even years in the case of chronic problems [14])
- measures of outcome, including appropriate methods and timeframes of measuring change in impairment, function, disability and handicap that address the needs of different stakeholders (patients, therapists, funders etc) [10,12,13].

Methods

Search strategy

In supplementary data [see additional file 1].

Data organization and extraction

Two independent researchers (PK, NMW) participated in all aspects of this review, and they compared and discussed their findings with respect to inclusion of critical appraisal tools, their intent, components, data extraction and item classification, construction and psychometric properties. Disagreements were resolved by discussion with a third member of the team (KG).

Data extraction consisted of a four-staged process. First, identical replica critical appraisal tools were identified and removed prior to analysis. The remaining critical appraisal tools were then classified according to the study design for which they were intended to be used [1,2]. The scientific manner in which the tools had been constructed was classified as whether an empirical research approach has been used, and if so, which type of research had been undertaken. Finally, the items contained in each critical appraisal tool were extracted and classified into one of eleven groups, which were based on the criteria described by Clarke and Oxman [4] as:

- Study aims and justification
- Methodology used, which encompassed method of identification of relevant studies and adherence to study protocol;
- Sample selection, which ranged from inclusion and exclusion criteria, to homogeneity of groups;
- Method of randomization and allocation blinding;
- Attrition: response and drop out rates;
- Blinding of the clinician, assessor, patient and statistician as well as the method of blinding;
- Outcome measure characteristics;
- Intervention or exposure details;
- Method of data analyses;
- Potential sources of bias; and
- Issues of external validity, which ranged from application of evidence to other settings to the relationship between benefits, cost and harm.

An additional group, "miscellaneous", was used to describe items that could not be classified into any of the groups listed above.

Data synthesis

Data was synthesized using MS Excel spread sheets as well as narrative format by describing the number of critical appraisal tools per study design and the type of items they contained. Descriptions were made of the method by which the overall quality of the study was determined, evidence regarding the psychometric properties of the tools (validity and reliability) and whether guidelines were provided for use of the critical appraisal tool.

Results

One hundred and ninety-three research reports that potentially provided a description of a critical appraisal tool (or process) were identified from the search strategy. Fifty-six of these papers were unavailable for review due to outdated Internet links, or inability to source the relevant journal through Australian university and Government library databases. Of the 127 papers retrieved, 19 were excluded from this review, as they did not provide a description of the critical appraisal tool used, or were published in languages other than English. As a result, 108 papers were reviewed, which yielded 121 different critical appraisal tools [1-5,7,9,15-102,116].

Empirical basis for tool construction

We identified 14 instruments (12% all tools) which were reported as having been constructed using a specified empirical approach [20,29,30,32,35,40,49,51,70-72,79,103,116]. The empirical research reflected descriptive and/or qualitative approaches, these being critical review of existing tools [40,72], Delphi techniques to identify then refine data items [32,51,71], questionnaires and other forms of written surveys to identify and refine data items [70,79,103], facilitated structured consensus meetings [20,29,30,35,40,49,70,72,79,116], and pilot validation testing [20,40,72,103,116]. In all the studies which reported developing critical appraisal tools using a consensus approach, a range of stakeholder input was sought, reflecting researchers and clinicians in a range of health disciplines, students, educators and consumers. There were a further 31 papers which cited other studies as the source of the tool used in the review, but which provided no information on why individual items had been chosen, or whether (or how) they had been modified. Moreover, for 21 of these tools, the cited sources of the critical appraisal tool did not report the empirical basis on which the tool had been constructed.

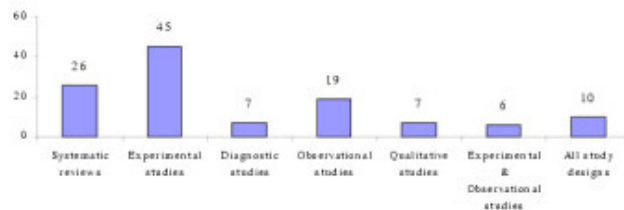


Figure 1
Number of critical appraisal tools per study design [1,2]

Table 1: Summary of tools sourced in this review.

Research design focus of critical appraisal tools		Critical appraisal tools with summary scores	
Secondary studies	Systematic reviews/meta-analyses [2-5,15-36,116]	All study designs [1,18,41,96-102,116]	Summary score [18,41,96,97,116]
Primary studies	Experimental studies [2-4,19,25-27,34,37-73]		No summary score [1,98-102]
	Diagnostic studies [19,74-79]	Experimental studies	Summary score [19,37-59] No summary score [2-4,25,27,28,34,60-73]
	Observational studies [2,3,7,19,25,66,72,80-86]	Diagnostic studies	Summary score [16,74-77] No summary score [78,79]
	Qualitative studies [9,26,66,87-90]		
	Experimental & Observational studies [9,91-102]	Qualitative studies	Summary score [87] No summary score [9,26,66,88-90]
		Experimental and observational studies	Summary score [91-93] No summary score [9,94,95]

Critical appraisal tools per study design

Seventy-eight percent (N = 94) of the critical appraisal tools were developed for use on primary research [1-5,7,9,18,19,25-27,34,37-41], while the remainder (N = 26) were for secondary research (systematic reviews and meta-analyses) [2-5,15-36,116]. Eighty-seven percent (N = 104) of all critical appraisal tools were design-specific [2-5,7,9,15-90], with over one third (N = 45) developed for experimental studies (randomized controlled trials, clinical trials) [2-4,25-27,34,37-73]. Sixteen critical appraisal tools were generic. Of these, six were developed for use on both experimental and observational studies [9,91-95], whereas 11 were purported to be useful for any qualitative and quantitative research design [1,18,41,96-102,116] (see Figure 1, Table 1).

Critical appraisal items

One thousand, four hundred and seventy five items were extracted from these critical appraisal tools. After grouping like items together, 173 different item types were identified, with the most frequently reported items being focused towards assessing the external validity of the study (N = 35) and method of data analyses (N = 28) (Table 2). The most frequently reported items across all critical appraisal tools were:

- Eligibility criteria (inclusion/exclusion criteria) (N = 63)
- Appropriate statistical analyses (N = 47)
- Random allocation of subjects (N = 43)
- Consideration of outcome measures used (N = 43)

- Sample size justification/power calculations (N = 39)
- Study design reported (N = 36)
- Assessor blinding (N = 36)

Design-specific critical appraisal tools*Systematic reviews*

Eighty-seven different items were extracted from the 26 critical appraisal tools, which were designed to evaluate the quality of systematic reviews. These critical appraisal tools frequently contained items regarding data analyses and issues of external validity (Tables 2 and 3).

Items assessing data analyses were focused to the methods used to summarize the results, assessment of sensitivity of results and whether heterogeneity was considered, whereas the nature of reporting of the main results, interpretation of them and their generalizability were frequently used to assess the external validity of the study findings. Moreover, systematic review critical appraisal tools tended to contain items such as identification of relevant studies, search strategy used, number of studies included and protocol adherence, that would not be relevant for other study designs. Blinding and randomisation procedures were rarely included in these critical appraisal tools.

Experimental studies

One hundred and twenty thirteen different items were extracted from the 45 experimental critical appraisal tools. These items most frequently assessed aspects of data analyses and blinding (Tables 1 and 2). Data analyses items were focused on whether appropriate statistical analysis

Table 2: The type and number of component items contained in critical appraisal tools per study design.

Type of items	Design-specific critical appraisal tool components					Generic critical appraisal tool components		Total
	Systematic reviews	Experimental studies	Diagnostic studies	Observational studies	Qualitative studies	Exp & Obs ^a studies	All study designs	
Study aims and justification	35	27	5	18	17	4	11	117
Methodology used	38	1	0	0	0	0	1	40
Sample selection	30	62	12	37	10	10	14	175
Randomization	2	65	1	5	0	6	5	84
Attrition	4	59	3	23	0	8	8	105
Blinding	1	77	5	8	0	5	7	103
Outcome measure characteristics	41	46	3	33	2	9	19	153
Intervention	7	42	3	13	0	5	12	82
Data analyses	83	91	14	54	12	14	27	295
Bias	24	14	2	5	0	3	6	54
External validity	72	50	12	30	27	9	27	227
Miscellaneous	11	12	7	5	7	2	6	50
Total	348	546	67	331	75	75	143	1485

Table 3: The type and number of guidelines accompanying critical appraisal tools per study design

Type of critical appraisal tool	Type of guideline				Total	Total number of critical appraisal tools
	Handbook/published paper		Accompanying explanation			
	Number of tools	References	Number of tools	References		
Systematic reviews	9	[2,4,15,20,25,28,29,331,36,116]	3	[16,26,27]	12	26
Experimental studies	10	[2,4,25,37,41,50,64-66,69]	6	[26,40,49,51,57,59]	16	45
Diagnostic studies	3	[74,75,76]	1	[79]	4	7
Observational studies	9	[2,25,66,80,84-87]	1	[83]	10	19
Qualitative studies	4	[9,87,89,90]	1	[26]	5	7
Experimental & Observational studies	2	[9,95]	1	[91]	3	6
All study designs	1	[100]	1	[102]	2	10
Total	38		14		52	120

was performed, whether a sample size justification or power calculation was provided and whether side effects of the intervention were recorded and analysed. Blinding was focused on whether the participant, clinician and assessor were blinded to the intervention.

Diagnostic studies

Forty-seven different items were extracted from the seven diagnostic critical appraisal tools. These items frequently addressed issues involving data analyses, external validity of results and sample selection that were specific to diagnostic studies (whether the diagnostic criteria were defined, definition of the "gold" standard, the calculation of sensitivity and specificity) (Tables 1 and 2).

Observational studies

Seventy-four different items were extracted from the 19 critical appraisal tools for observational studies. These items primarily focused on aspects of data analyses (see Tables 1 and 2, such as whether confounders were considered in the analysis, whether a sample size justification or power calculation was provided and whether appropriate statistical analyses were preformed).

Qualitative studies

Thirty-six different items were extracted from the seven qualitative study critical appraisal tools. The majority of these items assessed issues regarding external validity, methods of data analyses and the aims and justification of the study (Tables 1 and 2). Specifically, items were focused to whether the study question was clearly stated, whether data analyses were clearly described and appropriate, and application of the study findings to the clinical setting. Qualitative critical appraisal tools did not contain items regarding sample selection, randomization, blinding, intervention or bias, perhaps because these issues are not relevant to the qualitative paradigm.

Generic critical appraisal tools

Experimental and observational studies

Forty-two different items were extracted from the six critical appraisal tools that could be used to evaluate experimental and observational studies. These tools most frequently contained items that addressed aspects of sample selection (such as inclusion/exclusion criteria of participants, homogeneity of participants at baseline) and data analyses (such as whether appropriate statistical analyses were performed, whether a justification of the sample size or power calculation were provided).

All study designs

Seventy-eight different items were contained in the ten critical appraisal tools that could be used for all study designs (quantitative and qualitative). The majority of these items focused on whether appropriate data analyses

were undertaken (such as whether confounders were considered in the analysis, whether a sample size justification or power calculation was provided and whether appropriate statistical analyses were preformed) and external validity issues (generalization of results to the population, value of the research findings) (see Tables 1 and 2).

Allied health critical appraisal tools

We found no critical appraisal instrument specific to allied health research, despite finding at least seven critical appraisal instruments associated with allied health topics (mostly physiotherapy management of orthopedic conditions) [37,39,52,58,59,65]. One critical appraisal development group proposed two instruments [9], specific to quantitative and qualitative research respectively. The core elements of allied health research quality (specific diagnosis criteria, intervention descriptions, nature of patient contact and appropriate outcome measures) were not addressed in any one tool sourced for this evaluation. We identified 152 different ways of considering quality reporting of outcome measures in the 121 critical appraisal tools, and 81 ways of considering description of interventions. Very few tools which were not specifically targeted to diagnostic studies (less than 10% of the remaining tools) addressed diagnostic criteria. The critical appraisal instrument that seemed most related to allied health research quality [39] sought comprehensive evaluation of elements of intervention and outcome, however this instrument was relevant only to physiotherapeutic orthopedic experimental research.

Overall study quality

Forty-nine percent (N = 58) of critical appraisal tools summarised the results of the quality appraisal into a single numeric summary score [5,7,15-25,37-59,74-77,80-83,87,91-93,96,97] (Figure 2). This was achieved by one of two methods:

- An equal weighting system, where one point was allocated to each item fulfilled; or
- A weighted system, where fulfilled items were allocated various points depending on their perceived importance.

However, there was no justification provided for any of the scoring systems used. In the remaining critical appraisal tools (N = 62), a single numerical summary score was not provided [1-4,9,25-36,60-73,78,79,84-90,94,95,98-102]. This left the research consumer to summarize the results of the appraisal in a narrative manner, without the assistance of a standard approach.

Psychometric properties of critical appraisal tools

Few critical appraisal tools had documented evidence of their validity and reliability. Face validity was established

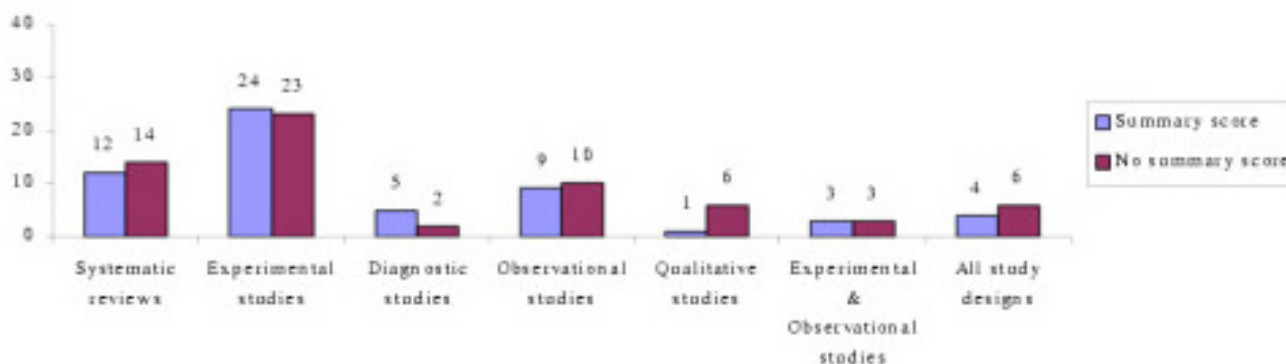


Figure 2
Number of critical appraisal tools with, and without, summary quality scores

in nine critical appraisal tools, seven of which were developed for use on experimental studies [38,40,45,49,51,63,70] and two for systematic reviews [32,103]. Intra-rater reliability was established for only one critical appraisal tool as part of its empirical development process [40], whereas inter-rater reliability was reported for two systematic review tools [20,36] (for one of these as part of the developmental process [20]) and seven experimental critical appraisal tools [38,40,45,51,55,56,63] (for two of these as part of the developmental process [40,51]).

Critical appraisal tool guidelines

Forty-three percent (N = 52) of critical appraisal tools had guidelines that informed the user of the interpretation of each item contained within them (Table 2). These guidelines were most frequently in the form of a handbook or published paper (N = 31) [2,4,9,15,20,25,28,29,31,36,37,41,50,64-67,69,80,84-87,89,90,95,100,116], whereas in 14 critical appraisal tools explanations accompanied each item [16,26,27,40,49,51,57,59,79,83,91,102].

Discussion

Our search strategy identified a large number of published critical appraisal tools that are currently available to critically appraise research reports. There was a distinct lack of information on tool development processes in most cases. Many of the tools were reported to be modifications of other published tools, or reflected specialty concerns in specific clinical or research areas, without attempts to justify inclusion criteria. Less than 10 of these tools were relevant to evaluation of the quality of allied health research, and none of these were based on an empirical research approach. We are concerned that although our search was

systematic and extensive [104,105], our broad key words and our lack of ready access to 29% of potentially useful papers (N = 56) potentially constrained us from identifying all published critical appraisal tools. However, consumers of research seeking critical appraisal instruments are not likely to seek instruments from outdated Internet links and unobtainable journals, thus we believe that we identified the most readily available instruments. Thus, despite the limitations on sourcing all possible tools, we believe that this paper presents a useful synthesis of the readily available critical appraisal tools.

The majority of the critical appraisal tools were developed for a specific research design (87%), with most designed for use on experimental studies (38% of all critical appraisal tools sourced). This finding is not surprising as, according to the medical model, experimental studies sit at or near the top of the hierarchy of evidence [2,8]. In recent years, allied health researchers have strived to apply the medical model of research to their own discipline by conducting experimental research, often by using the randomized controlled trial design [106]. This trend may be the reason for the development of experimental critical appraisal tools reported in allied health-specific research topics [37,39,52,58,59,65].

We also found a considerable number of critical appraisal tools for systematic reviews (N = 26), which reflects the trend to synthesize research evidence to make it relevant for clinicians [105,107]. Systematic review critical appraisal tools contained unique items (such as identification of relevant studies, search strategy used, number of studies included, protocol adherence) compared with tools used for primary studies, a reflection of the secondary nature of data synthesis and analysis.

In contrast, we identified very few qualitative study critical appraisal tools, despite the presence of many journal-specific guidelines that outline important methodological aspects required in a manuscript submitted for publication [108-110]. This finding may reflect the more traditional, quantitative focus of allied health research [111]. Alternatively, qualitative researchers may view the robustness of their research findings in different terms compared with quantitative researchers [112,113]. Hence the use of critical appraisal tools may be less appropriate for the qualitative paradigm. This requires further consideration.

Of the small number of generic critical appraisal tools, we found few that could be usefully applied (to any health research, and specifically to the allied health literature), because of the generalist nature of their items, variable interpretation (and applicability) of items across research designs, and/or lack of summary scores. Whilst these types of tools potentially facilitate the synthesis of evidence across allied health research designs for clinicians, their lack of specificity in asking the 'hard' questions about research quality related to research design also potentially precludes their adoption for allied health evidence-based practice. At present, the gold standard study design when synthesizing evidence is the randomized controlled trial [4], which underpins our finding that experimental critical appraisal tools predominated in the allied health literature [37,39,52,58,59,65]. However, as more systematic literature reviews are undertaken on allied health topics, it may become more accepted that evidence in the form of other research design types requires acknowledgement, evaluation and synthesis. This may result in the development of more appropriate and clinically useful allied health critical appraisal tools.

A major finding of our study was the volume and variation in available critical appraisal tools. We found no gold standard critical appraisal tool for any type of study design. Therefore, consumers of research are faced with frustrating decisions when attempting to select the most appropriate tool for their needs. Variable quality evaluations may be produced when different critical appraisal tools are used on the same literature [6]. Thus, interpretation of critical analysis must be carefully considered in light of the critical appraisal tool used.

The variability in the content of critical appraisal tools could be accounted for by the lack of any empirical basis of tool construction, established validity of item construction, and the lack of a gold standard against which to compare new critical tools. As such, consumers of research cannot be certain that the content of published critical appraisal tools reflect the most important aspects of the quality of studies that they assess [114]. Moreover, there

was little evidence of intra- or inter-rater reliability of the critical appraisal tools. Coupled with the lack of protocols for use, this may mean that critical appraisers could interpret instrument items in different ways over repeated occasions of use. This may produce variable results [123].

Conclusions

Based on the findings of this evaluation, we recommend that consumers of research should carefully select critical appraisal tools for their needs. The selected tools should have published evidence of the empirical basis for their construction, validity of items and reliability of interpretation, as well as guidelines for use, so that the tools can be applied and interpreted in a standardized manner. Our findings highlight the need for consensus to be reached regarding the important and core items for critical appraisal tools that will produce a more standardized environment for critical appraisal of research evidence. As a consequence, allied health research will specifically benefit from having critical appraisal tools that reflect best practice research approaches which embed specific research requirements of allied health disciplines.

Competing interests

No competing interests.

Authors' contributions

PK Sourced critical appraisal tools

Categorized the content and psychometric properties of critical appraisal tools

AEB Synthesis of findings

Drafted manuscript

NMW Sourced critical appraisal tools

Categorized the content and psychometric properties of critical appraisal tools

VSK Sourced critical appraisal tools

Categorized the content and psychometric properties of critical appraisal tools

KAG Study conception and design

Assisted with critiquing critical appraisal tools and categorization of the content and psychometric properties of critical appraisal tools

Drafted and reviewed manuscript

Addressed reviewer's comments and re-submitted the article

Additional material

Additional File 1

Search Strategy.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2288-4-22-S1.doc>]

References

- National Health and Medical Research Council: *How to Review the Evidence: Systematic Identification and Review of the Scientific Literature*. Canberra 2000.
- National Health and Medical Research Council: *How to Use the Evidence: Assessment and Application of Scientific Evidence*. Canberra 2000.
- Joanna Briggs Institute [<http://www.joannabriggs.edu.au>]
- Clarke M, Oxman AD: *Cochrane Reviewer's Handbook 4.2.0* Oxford: The Cochrane Collaboration; 2003.
- Crombie IK: *The Pocket Guide to Critical Appraisal: A Handbook for Health Care Professionals* London: BMJ Publishing Group; 1996.
- Agency for Healthcare Research and Quality: *Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47, Publication No. 02-E016*. Rockville 2002.
- Elwood JM: *Critical Appraisal of Epidemiological Studies and Clinical Trials* 2nd edition. Oxford: Oxford University Press; 1998.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB: *Evidence Based Medicine. How to Practice and Teach EBM* London: Churchill Livingstone; 2000.
- Critical literature reviews** [http://www.cotfcanada.org/corf_critical.htm]
- Bialocerkowski AE, Grimmer KA, Milanese SF, Kumar S: **Application of current research evidence to clinical physiotherapy practice.** *J Allied Health Res Dec* in press.
- The National Health Data Dictionary - Version 10.** . <http://www.aihw.gov.au/publications/hwi/nhdd12/nhdd12-v1.pdf> and <http://www.aihw.gov.au/publications/hwi/nhdd12/nhdd12-v2.pdf>
- Grimmer K, Bowman P, Roper J: **Episodes of allied health outpatient care: an investigation of service delivery in acute public hospital settings.** *Disability and Rehabilitation* 2000, **22(1/2)**:80-87.
- Grimmer K, Milanese S, Bialocerkowski A: **Clinical guidelines for low back pain: A physiotherapy perspective.** *Physiotherapy Canada* 2003, **55(4)**:1-9.
- Grimmer KA, Milanese S, Bialocerkowski AE, Kumar S: **Producing and implementing evidence in clinical practice: the therapists' dilemma.** *Physiotherapy* 2004 in press.
- Greenhalgh T: **How to read a paper: papers that summarize other papers (systematic reviews and meta-analysis).** *BMJ* 1997, **315**:672-675.
- Auperin A, Pignon J, Poynard T: **Review article: critical review of meta-analysis of randomised clinical trials in hepatogastroenterology.** *Alimentary Pharmacol Therapeutics* 1997, **11**:215-225.
- Barnes DE, Bero LA: **Why review articles on the health effects of passive smoking reach different conclusions.** *J Am Med Assoc* 1998, **279**:1566-1570.
- Beck CT: **Use of meta-analysis as a teaching strategy in nursing research courses.** *J Nurs Educat* 1997, **36**:87-90.
- Carruthers SG, Larochelle P, Haynes RB, Petrasovits A, Schiffrin EL: **Report of the Canadian Hypertension Society Consensus Conference: I. Introduction.** *Can Med Assoc J* 1993, **149**:289-293.
- Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchinson BG, Milner RA, Streiner DL: **Agreement among reviewers of review articles.** *J Clin Epidemiol* 1991, **44**:91-98.
- Sacks HS, Reitman D, Pagano D, Kupelnick B: **Meta-analysis: an update.** *Mount Sinai Journal of Medicine* 1996, **63**:216-224.
- Smith AF: **An analysis of review articles published in four anaesthesia journals.** *Can J Anaesth* 1997, **44**:405-409.
- L'Abbe KA, Detsky AS, O'Rourke K: **Meta-analysis in clinical research.** *Ann Intern Med* 1987, **107**:224-233.
- Mulrow CD, Antonio S: **The medical review article: state of the science.** *Ann Intern Med* 1987, **106**:485-488.
- Continuing Professional Development: A Manual for SIGN Guideline Developers** [<http://www.sign.ac.uk>]
- Learning and Development Public Health Resources Unit** [<http://www.phru.nhs.uk/>]
- FOCUS Critical Appraisal Tool** [<http://www.focusproject.org.uk/>]
- Cook DJ, Sackett DL, Spitzer WO: **Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on meta-analysis.** *J Clin Epidemiol* 1995, **48**:167-171.
- Cranney A, Tugwell P, Shea B, Wells G: **Implications of OMER-ACT outcomes in arthritis and osteoporosis for Cochrane metaanalysis.** *J Rheumatol* 1997, **24**:1206-1207.
- Guyatt GH, Sackett DL, Sinclair JC, Hoyward R, Cook DJ, Cook RJ: **User's guide to the medical literature. IX. A method for grading health care recommendations.** *J Am Med Assoc* 1995, **274**:1800-1804.
- Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EAF, Milson ME, Rasooli Iris, Frank JW, Riben PD, Mathias RG: **An approach to the development of practice guidelines for community health interventions.** *Can J Public Health* 1994, **85**:S8-13.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: **Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses.** *Lancet* 1999, **354**:1896-1900.
- Oxman AD, Cook DJ, Guyatt GH: **Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group.** *J Am Med Assoc* 1994, **272**:1367-1371.
- Pogue J, Yusuf S: **Overcoming the limitations of current meta-analysis of randomised controlled trials.** *Lancet* 1998, **351**:47-52.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB: **Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group.** *J Am Med Assoc* 2000, **283**:2008-2012.
- Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mostellar F: **Guidelines for meta-analyses evaluating diagnostic tests.** *Ann Intern Med* 1994, **120**:667-676.
- Moseley AM, Herbert RD, Sherrington C, Maher CG: **Evidence for physiotherapy practice: A survey of the Physiotherapy Evidence Database. Physiotherapy Evidence Database (PEDro).** *Australian Journal of Physiotherapy* 2002, **48**:43-50.
- Cho MK, Bero LA: **Instruments for assessing the quality of drug studies published in the medical literature.** *J Am Med Assoc* 1994, **272**:101-104.
- De Vet HCV, De Bie RA, Van der Heijden GJ, Verhagen AP, Sijpkens P, Kipschild PG: **Systematic reviews on the basis of methodological criteria.** *Physiotherapy* 1997, **83**:284-289.
- Downs SH, Black N: **The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions.** *J Epidemiol Community Health* 1998, **52**:377-384.
- Evans M, Pollock AV: **A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection.** *Br J Surg* 1985, **72**:256-260.
- Fahey T, Hyde C, Milne R, Thorogood M: **The type and quality of randomized controlled trials (RCTs) published in UK public health journals.** *J Public Health Med* 1995, **17**:469-474.
- Gotzsche PC: **Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis.** *Control Clin Trials* 1989, **10**:31-56.
- Imperiale TF, McCullough AJ: **Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials.** *Ann Int Med* 1990, **113**:299-307.
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ: **Assessing the quality of reports of randomized clinical trials: is blinding necessary?** *Control Clin Trials* 1996, **17**:1-12.
- Khan KS, Daya S, Collins JA, Walter SD: **Empirical evidence of bias in infertility research: overestimation of treatment**

- effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril* 1996, **65**:939-945.
47. Kleijnen J, Knipschild P, ter Riet G: **Clinical trials of homoeopathy.** *BMJ* 1991, **302**:316-323.
 48. Liberati A, Himmel HN, Chalmers TC: **A quality assessment of randomized control trials of primary treatment of breast cancer.** *J Clin Oncol* 1986, **4**:942-951.
 49. Moher D, Schulz KF, Altman DG, for the CONSORT Group: **The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials.** *J Am Med Assoc* 2001, **285**:1987-1991.
 50. Reisch JS, Tyson JE, Mize SG: **Aid to the evaluation of therapeutic studies.** *Pediatrics* 1989, **84**:815-827.
 51. Sindhu F, Carpenter L, Seers K: **Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique.** *J Advanced Nurs* 1997, **25**:1262-1268.
 52. Van der Heijden GJ, Van der Windt DA, Kleijnen J, Koes BW, Bouter LM: **Steroid injections for shoulder disorders: a systematic review of randomized clinical trials.** *Br J Gen Pract* 1996, **46**:309-316.
 53. Van Tulder MW, Koes BW, Bouter LM: **Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions.** *Spine* 1997, **22**:2128-2156.
 54. Garbutt JC, West SL, Carey TS, Lohr KN, Crews FT: *Pharmacotherapy for Alcohol Dependence. Evidence Report/Technology Assessment No. 3, AHCPR Publication No. 99-E004.* Rockville 1999.
 55. Oremus M, Wolfson C, Perrault A, Demers L, Momoli F, Moride Y: **Inter-rater reliability of the modified Jadad quality scale for systematic reviews of Alzheimer's disease drug trials.** *Dement Geriatr Cognit Disord* 2001, **12**:232-236.
 56. Clark O, Castro AA, Filho JV, Djubelgovic B: **Interrater agreement of Jadad's scale.** *Annual Cochrane Colloquium Abstracts October 2001 Lyon* [<http://www.biomedcentral.com/abstracts/COCHRANE/1/op031>].
 57. Jonas W, Anderson RL, Crawford CC, Lyons JS: **A systematic review of the quality of homeopathic clinical trials.** *BMC Alternative Medicine* 2001, **1**:12.
 58. Van Tulder M, Malmivaara A, Esmail R, Koes B: **Exercises therapy for low back pain: a systematic review within the framework of the Cochrane Collaboration back review group.** *Spine* 2000, **25**:2784-2796.
 59. Van Tulder MW, Ostelo R, Vlaeyen JWS, Linton SJ, Morley SJ, Assendelft WJJ: **Behavioral treatment for chronic low back pain: a systematic review within the framework of the cochrane back.** *Spine* 2000, **25**:2688-2699.
 60. Aronson N, Seidenfeld J, Samson DJ, Aronson N, Albertson PC, Bayoumi AM, Bennett C, Brown A, Garber ABA, Gere M, Hasselblad V, Wilt T, Ziegler MPH, Pharm D: *Relative Effectiveness and Cost Effectiveness of Methods of Androgen Suppression in the Treatment of Advanced Prostate Cancer. Evidence Report/Technology Assessment No. 4, AHCPR Publication No.99-E0012.* Rockville 1999.
 61. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A: **A method for assessing the quality of a randomized control trial.** *Control Clin Trials* 1981, **2**:31-49.
 62. der Simonian R, Charette LJ, McPeck B, Mosteller F: **Reporting on methods in clinical trials.** *New Eng J Med* 1982, **306**:1332-1337.
 63. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA: **Incorporating variations in the quality of individual randomized trials into meta-analysis.** *J Clin Epidemiol* 1992, **45**:255-265.
 64. Goudas L, Carr DB, Bloch R, Balk E, Ioannidis JPA, Terrin MN: *Management of Cancer Pain. Evidence Report/Technology Assessment No. 35 (Contract 290-97-0019 to the New England Medical Center), AHCPR Publication No. 99-E004.* Rockville 2000.
 65. Guyatt GH, Sackett DL, Cook DJ: **Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group.** *J Am Med Assoc* 1993, **270**:2598-2601.
 66. Khan KS, Ter Riet G, Glanville J, Sowden AJ, Kleijnen J: *Undertaking Systematic Reviews of Research on Effectiveness: Centre of Reviews and Dissemination's Guidance for Carrying Out or Commissioning Reviews* York 2000.
 67. McNamara R, Bass EB, Marlene R, Miller J: *Management of New Onset Atrial Fibrillation. Evidence Report/Technology Assessment No.12, AHRQ Publication No. 01-E026.* Rockville 2001.
 68. Prendiville W, Elbourne D, Chalmers I: **The effects of routine oxytocic administration in the management of the third stage of labour: an overview of the evidence from controlled trials.** *Br J Obstet Gynaecol* 1988, **95**:3-16.
 69. Schulz KF, Chalmers I, Hayes RJ, Altman DG: **Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.** *J Am Med Assoc* 1995, **273**:408-412.
 70. The Standards of Reporting Trials Group: **A proposal for structured reporting of randomized controlled trials.** *J Am Med Assoc* 1994, **272**:1926-1931.
 71. Verhagen AP, de Vet HC, de Bie RA, Kessels AGH, Boers M, Bouter LM, Knipschild PG: **The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus.** *J Clin Epidemiol* 1998, **51**:1235-1241.
 72. Zaza S, Wright-De Agüero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH, Sosin DM, Anderson L, Carande-Kullis VG, Teutsch SM, Pappaioanou M: **Data collection instrument and procedure for systematic reviews in the guide to community preventive services. Task force on community preventive services.** *Am J Prevent Med* 2000, **18**:44-74.
 73. Haynes BB, Wilczynski N, McKibbon A, Walker CJ, Sinclair J: **Developing optimal search strategies for detecting clinically sound studies in MEDLINE.** *J Am Informatics Assoc* 1994, **1**:447-458.
 74. Greenhalgh T: **How to read a paper: papers that report diagnostic or screening tests.** *BMJ* 1997, **315**:540-543.
 75. Arroll B, Schechter MT, Sheps SB: **The assessment of diagnostic tests: a comparison of medical literature in 1982 and 1985.** *J Gen Int Med* 1988, **3**:443-447.
 76. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *J Am Med Assoc* 1999, **282**:1061-1066.
 77. Sheps SB, Schechter MT: **The assessment of diagnostic tests. A survey of current medical research.** *J Am Med Assoc* 1984, **252**:2418-2422.
 78. McCrory DC, Matchar DB, Bastian L, Dutta S, Hasselblad V, Hickey J, Myers MSE, Nanda K: *Evaluation of Cervical Cytology. Evidence Report/Technology Assessment No. 5, AHCPR Publication No.99-E010.* Rockville 1999.
 79. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, DeVet HCW: **Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative.** *Clin Chem* 2003, **49**:1-6.
 80. Greenhalgh T: **How to Read a Paper: Assessing the methodological quality of published papers.** *BMJ* 1997, **315**:305-308.
 81. Angelillo I, Villari P: **Residential exposure to electromagnetic fields and childhood leukaemia: a meta-analysis.** *Bull World Health Org* 1999, **77**:906-915.
 82. Ariens G, Mechelen W, Bongers P, Bouter L, Van der Wal G: **Physical risk factors for neck pain.** *Scand J Work Environ Health* 2000, **26**:7-19.
 83. Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM: **Physical load during work and leisure time as risk factors for back pain.** *Scand J Work Environ Health* 1999, **25**:387-403.
 84. Laupacis A, Wells G, Richardson WS, Tugwell P: **Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group.** *J Am Med Assoc* 1994, **272**:234-237.
 85. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V: **Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group.** *J Am Med Assoc* 1994, **271**:1615-1619.
 86. Carey TS, Boden SD: **A critical guide to case series reports.** *Spine* 2003, **28**:1631-1634.
 87. Greenhalgh T, Taylor R: **How to read a paper: papers that go beyond numbers (qualitative research).** *BMJ* 1997, **315**:740-743.
 88. Hoddinott P, Pill R: **A review of recently published qualitative research in general practice. More methodological questions than answers?** *Fam Pract* 1997, **14**:313-319.
 89. Mays N, Pope C: **Quality research in health care: Assessing quality in qualitative research.** *BMJ* 2000, **320**:50-52.
 90. Mays N, Pope C: **Rigour and qualitative research.** *BMJ* 1995, **311**:109-112.

91. Colditz GA, Miller JN, Mosteller F: **How study design affects outcomes in comparisons of therapy. I: Medical.** *Stats Med* 1989, **8**:441-454.
92. Turlik MA, Kushner D: **Levels of evidence of articles in podiatric medical journals.** *J Am Pod Med Assoc* 2000, **90**:300-302.
93. Borghouts JAJ, Koes BW, Bouter LM: **The clinical course and prognostic factors of non-specific neck pain: a systematic review.** *Pain* 1998, **77**:1-13.
94. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P, Abenhaim L, Hardy J, Sampalis J, Pinfold SP, Morgan PP: **Links between passive smoking and disease: a best-evidence synthesis. A report of the working group on passive smoking.** *Clin Invest Med* 1990, **13**:17-46.
95. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F: **Systematic reviews of trials and other studies.** *Health Tech Assess* 1998, **2**:1-276.
96. Chestnut RM, Carney N, Maynard H, Patterson P, Mann NC, Helfand M: *Rehabilitation for Traumatic Brain Injury. Evidence Report/Technology Assessment No. 2, Agency for Health Care Research and Quality Publication No. 99-E006.* Rockville 1999.
97. Lohr KN, Carey TS: **Assessing best evidence: issues in grading the quality of studies for systematic reviews.** *Joint Commission J Qual Improvement* 1999, **25**:470-479.
98. Greer N, Mosser G, Logan G, Halaas GW: **A practical approach to evidence grading.** *Joint Commission J Qual Improvement* 2000, **26**:700-712.
99. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D: **Current methods of the U.S. Preventive Services Task Force: a review of the process.** *Am J Prevent Med* 2001, **20**:21-35.
100. Anonymous: **How to read clinical journals: IV. To determine etiology or causation.** *Can Med Assoc J* 1981, **124**:985-990.
101. Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S: **Systematic review of cost effectiveness studies of telemedicine interventions.** *BMJ* 2002, **324**:1434-1437.
102. Forrest JL, Miller SA: **Evidence-based decision making in action: Part 2-evaluating and applying the clinical evidence.** *J Contemp Dental Pract* 2002, **4**:42-52.
103. Oxman AD, Guyatt GH: **Validation of an index of the quality of review articles.** *J Clin Epidemiol* 1991, **44**:1271-1278.
104. Jones T, Evans D: **Conducting a systematic review.** *Aust Crit Care* 2000, **13**:66-71.
105. Papadopoulos M, Rheeder P: **How to do a systematic literature review.** *South African J Physiother* 2000, **56**:3-6.
106. Selker LG: **Clinical research in Allied Health.** *J Allied Health* 1994, **23**:201-228.
107. Stevens KR: **Systematic reviews: the heart of evidence-based practice.** *AACN Clin Issues* 2001, **12**:529-538.
108. Devers KJ, Frankel RM: **Getting qualitative research published.** *Ed Health* 2001, **14**:109-117.
109. Canadian Journal of Public Health: **Review guidelines for qualitative research papers submitted for consideration to the Canadian Journal of Public Health.** *Can J Pub Health* 2000, **91**:12.
110. Malterud K: **Shared understanding of the qualitative research process: guidelines for the medical researcher.** *Fam Pract* 1993, **10**:201-206.
111. Higgs J, Titchen A: **Research and knowledge.** *Physiotherapy* 1998, **84**:72-80.
112. Maggs-Rapport F: **Best research practice: in pursuit of methodological rigour.** *J Advan Nurs* 2001, **35**:373-383.
113. Cutcliffe JR, McKenna HP: **Establishing the credibility of qualitative research findings: the plot thickens.** *J Advan Nurs* 1999, **30**:374-380.
114. Andresen EM: **Criteria for assessing the tools of disability outcomes research.** *Arch Phys Med Rehab* 2000, **81**:S15-S20.
115. Beatie P: **Measurement of health outcomes in the clinical setting: applications to physiotherapy.** *Phys Theory Pract* 2001, **17**:173-185.
116. Charnock DF, (Ed): **The DISCERN Handbook: Quality criteria for consumer health information on treatment choices.** Radcliffe Medical Press; 1998.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/4/22/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

