Public Health (2008) **122**, 92–98

ELSEVIER

PUBLIC HEALTH

JOURNAL OF THE ROYAL INSTITUTE OF PUBLIC HEALTH

www.elsevierhealth.com/journals/pubh

Original Research

Critical appraisal for public health: A new checklist

Richard F. Heller*, Arpana Verma, Islay Gemmell, Roger Harrison, Judy Hart, Richard Edwards

Evidence for Population Health Unit, Division of Epidemiology and Health Sciences, University of Manchester, Manchester M13 9PT, UK

Received 11 October 2006; received in revised form 30 March 2007; accepted 10 April 2007
Available online 4 September 2007

KEYWORDS
Critical appraisal;
Public health;
Evidence for population health

Summary *Objectives:* There have been a number of attempts to develop critical appraisal tools, but few have had a public health focus. This paper describes a new checklist with public health aspects.
Study design: Review of previous appraisal instruments and pilot test of new checklist.
Methods: Criteria of particular reference to public health practice were added to well-established appraisal criteria. The checklist was piloted with 21 public health professionals, research staff or postgraduate students.
Results: The checklist is organized using the 'ask', 'collect', 'understand' and 'use' categories of the Population Health Evidence Cycle. Readers are asked to assess validity, completeness and transferability of the data as they relate to: the study question; key aspects of the methodology; possible public health implications of the key results; and the implications for implementation in their own public health practice. Of the 21 public health professionals that piloted the checklist, 20 said that they found the checklist useful and 18 would use it or recommend it in the future. Participants were prepared to commit to the majority of the questions, and there was good agreement with a consensus of 'correct' answers.
Conclusions: The public health critical appraisal checklist adds public health aspects that were missing from previous critical appraisal tools.

Introduction

Critical appraisal is a key skill required by health professionals and policy makers. There have been a number of attempts to develop critical appraisal tools, but few have had a public health focus. For example, neither the Users' Guide to Evidence-based Practice, originally published as a series in the *Journal of the American Medical Association*,[1] nor the UK Critical Appraisal Skills Programme,[2] nor the book by Greenhalgh, originally published

*Corresponding author. Tel.: +44 131 555 4493.
E-mail address: Dick.Heller@manchester.ac.uk (R.F. Heller).

0033-3506/$ - see front matter © 2007 The Royal Institute of Public Health. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.puhe.2007.04.012

as a series of papers in the *British Medical Journal*,[3] have guidelines that relate to public health interventions. A number of tools have been published on easily accessible websites with an emphasis on public health and health promotion interventions.[4–7] One[5] includes a reference to the data-collection instrument used by the US Guide to Preventive Services,[8] and another includes summary points for methodology quality for primary studies.[6] However, much of the emphasis is on tools to assess reviews rather than individual studies, and only one includes public health issues.[8] Most guidelines suggest that a different set of questions should be asked for each study type. However, the authors' experience over many years with a single checklist to cover key aspects of study design is very positive,[9] and a generic instrument may be attractive for those wishing to keep a structure for critical appraisal in their minds, for example, for the purposes of an examination.

Critical appraisal of papers relating to public health issues may require a different focus from papers relating purely to clinical topics. For example, there may be particular questions that will need to be answered in deciding whether to introduce public health interventions that relate to the context in which it is to be implemented, resource implications and appropriateness of the intervention. Public health practitioners appraising the policy relevance of studies exploring the link between risk factors and health outcomes need to take into account the potential population impact of preventive measures.[10] A critical appraisal tool for public health practitioners needs to include these considerations.

Evidence-based medicine is often criticized for focusing on the assessment of efficacy and neglecting the importance of 'implementation',[11] and a public health appraisal tool should place particular importance on how the results can be implemented. The tool also needs to be user friendly at all levels of experience to be of maximum use to public health teams, which include practitioners with varied levels of prior training.

This paper describes the development, characteristics and piloting with public health practitioners of a new framework for the appraisal of evidence of relevance to public health, through a critical appraisal checklist to help review papers and reports of public health research. The checklist has been designed to be used in conjunction with an evidence hierarchy that is relevant to public health rather than to clinical practice.[10,12]

Methods

Development of the public health critical appraisal checklist

The appraisal method for study methodologies presented in this paper relates to well-established criteria of study methodology, as relevant to the type of study being reported. For this, previous critical appraisal criteria have been adapted. Various explorations of public health evidence were referred to whilst deciding which additional criteria are of particular reference to public health.

Petticrew and Roberts[13] adapted a framework from Muir Gray, and suggested that one should first identify the research question being asked, and then identify the most appropriate research design to answer this question. Victora et al.[14] and Habicht et al.[15] argued that the main objective of an evaluation is to influence decisions, and thus the nature and complexity of the evaluation should depend on the decision to be made once the evaluation has been performed. They provide two axes along which an evaluation might be performed,'What do you want to measure?' and 'How sure do you want to be?', and have suggested a taxonomy of design – probability (randomized controlled trial), plausibility (non-random observational comparisons involving a comparison group) and adequacy (observation of trends). Rychetnik et al.[16] pointed out that the study design is not sufficient to decide on whether the results of a study are adequate markers of quality to be of use in public health decision making. They define credibility, completeness and transferability as being of importance beyond the study design itself. Kirkwood et al.[17] drew on experience of research in developing countries to suggest design issues of relevance to community-based interventions. Many of the themes identified by the above authors have been adapted and incorporated into the present checklist. Specifically, the questions were oriented to emphasize the relevance of the study to the research question being asked[13–15] and to the policy issue being addressed[15] in the population of interest.[17] It has also been made clear in the checklist that the questions are designed to assess completeness and transferability (as well as quality).[16] The Population Health Evidence Cycle ('ask', 'collect', 'understand' and 'use')[18] was used as an organizing framework to guide the major steps in the appraisal process.

Piloting the checklist

The usefulness of the checklist in critical appraisal exercises was explored in two settings; a public

health conference in the north–west of England, and a departmental journal club attended by research staff and postgraduate students. Individuals were presented with up to three journal articles and asked their views on the perceived usefulness of the checklist. Concordance was examined between their critical appraisal responses on the checklist and a 'correct' answer developed by discussion between the authors.

The final version of the checklist presented here (see Appendix) is modified based on the results of the pilot study, feedback from these practitioners and further discussion among the authors.

Results

The checklist

Readers are asked to appraise the validity, completeness and transferability of the data presented as they relate to:

- the study question ('ask');
- the study design, sampling, exposures, outcomes, confounders and other aspects of internal validity relevant to the study type ('collect');
- the interpretation and population relevance of the results ('understand'); and
- the implications for implementation in their own population and public health practice ('use').

A glossary of terms is included at the end of the checklist in case any of the terms are unfamiliar.

It is suggested that users should work systematically through the checklist, first seeing if they can find relevant information in the paper, and then considering if the methods used threaten either the validity, completeness or transferability of the study in such a way as to limit its usefulness.

Pilot experience with use of the checklist

All 21 participants appraised one paper, 11 participants had time to appraise two papers, and six participants appraised three papers. Thus, the maximum number of possible responses was 38. Twenty of the 21 participants said that they found the checklist useful and 18 stated that they would use it themselves in the future or recommend it to others.

Table 1 shows the number of participants who responded to each question, and whether their responses agreed with an external consensus derived by the authors. Participants were prepared to commit to the majority of questions, and there was a high level of agreement (65–96%) with the 'correct' answers derived by the authors of this paper.

Discussion

The public health critical appraisal checklist includes public health aspects that are missing from previous critical appraisal tools. As mentioned

Table 1 Responses after appraisal of papers by 21 people*.

Item	Number answering	Agree with 'correct' answer
Research question: is the study type appropriate for the research question?	33	27
Sampling: is the sampling method adequate?	32	26
Exposures: is the measurement of exposures adequate?	28	18
Outcomes: is the measurement of outcomes appropriate?	33	29
Is the choice of outcome measures appropriate?	25	22
Confounders: has confounding been dealt with adequately?	27	26
Other issues of internal validity: are you happy with other issues of internal validity?	20	16
Have the results been interpreted properly?		
Statistical tests	30	28
Public health impact	24	18
Do you think that the study is useful enough to impact on health policy in some way?	28	22
Will you use the results of this study?	27	22

*All were asked to appraise one paper, 11 appraised two papers, and six appraised a third paper: maximum number of responses was 38.

above, most critical appraisal tools are not designed to incorporate the critical appraisal of public health aspects of individual studies. This includes the majority of tools available for reviewing public health interventions. The major exception to this is the data-collection instrument used by the US Guide to Preventive Services,[8] which is too detailed and extensive to allow for easy application by practitioners. A systematic review of the content of critical appraisal tools found none that related specifically to allied health research, and suggested the need to interpret critical appraisal according to the properties and intent of the chosen tool.[19] Hence, the need for a tool that is relevant to a public health context.

The use of the Population Health Evidence Cycle ('ask', 'collect', 'understand', 'use') allows a simple and reproducible set of organizing criteria.

It is suggested that this tool should be used to obtain an initial appraisal, but more detailed critical appraisal tools should be used to fill in any gaps in the detailed examination of methodology, particularly where details of the internal validity of studies are concerned. The checklist does not include a system to derive a quality score, as quality scores are of dubious validity and reproducibility.[20–22] The checklist could also be used to identify key dimensions of validity which should be met in study design and methodology as part of the process of formulating grades of recommendation,[23] as well as for those designed especially for public health interventions.[12] The checklist could be particularly useful for studies considering public health policy issues, and where the populations or the interventions proposed may not be similar to those where the policy is to be applied.

The checklist is designed to examine individual studies and reports; however, policy and practice decisions are rarely based on the findings of a single study.[24] An appraisal of the policy implications of an individual study should be made in the light of consistency with other available evidence, preferably in a systematic review. The checklist therefore asks specifically how the results fit into the wider evidence base. It also asks how the findings would apply to the reviewer's own population of interest, since organizational context is an important consideration for the implementation of research findings.

The pilot examination of the use of the checklist does not demonstrate any improvement in performance in comparison with other, or no, tools. Nor were explicit criteria used to define whether the respondents found the checklist useful. However, a range of individuals working in public health were prepared to commit to the majority of the questions, and there was good agreement with a consensus of 'correct' answers. Further work would be required to establish any improvement in performance associated with use of the checklist.

Conclusions

The checklist incorporates public health aspects into previous critical appraisal tools. It was piloted among people working in public health, who said that it was useful and would recommend its use. It is the authors' hope that it may find a place as an addition to other critical appraisal checklists that do not include a public health perspective.

Ethical approval

Not applicable.

Funding

None.

Competing interests

None.

Appendix. Public health critical appraisal checklist

Table A.1

Table A.1 Public health critical appraisal checklist.

Item on Population Health Evidence Cycle	Can you find this in the paper?	Questions to assess the validity, completeness and transferability of the study.	Answers
Ask	What is the research question/hypothesis?	Is the research question and/or hypothesis stated clearly?	The research question/ hypothesis is: Yes/no/why
	Relevance of research question/hypothesis	Is this question relevant to my, and/or other, population(s) and to the policy decision to be made?	

Table A.1 (*continued*)

Item on Population Health Evidence Cycle	Can you find this in the paper?	Questions to assess the validity, completeness and transferability of the study.	Answers
Collect	Study design	What is the study type?	
		Is the study type appropriate for the research question?	Yes/no/why
		Is there a comparison group?	Yes/no
	Sampling	Are the sampling frame and sampling method appropriate?	Yes/no/why
		Was sample size/power calculated and appropriate?	Yes/no
		Is the sample representative of the population being studied?	Yes/no/why
		Are exclusion criteria appropriate?	Yes/no/why
		Is the response rate adequate?	Yes/no/why
		Can you generalize from the population being studied? (external validity)	Yes/no/why
		Is this sample relevant to my population?	Yes/no/why
	Exposures	Observations/risk factors: how are the exposures measured? Is there bias in the measurement?	
		Intervention features (for an intervention study): is the intervention described adequately?	Yes/no
		Are these observations, risk factors or interventions relevant in my population?	Yes/no
	Outcomes	What are the outcome factors?	
		How are they measured? Is there bias in the measurement?	
		Are these outcome measures appropriate? Will they help answer the research question? Do they cover the interests of the stakeholders? Are they relevant to my population?	Yes/no
	Confounders	What important confounders are considered, and how are they addressed?	
		Has confounding been dealt with adequately?	Yes/no/how
		Are there other confounders that should have been addressed?	Yes/no/which
Understand	Have the results been interpreted appropriately?	Are statistical tests appropriate and correct?	Yes/no
		What are the main results and are they presented in an understandable way?	Yes/no
		Have measures of absolute risk as well as relative risk been included?	Yes/no
		[For any intervention study] Have the resource and cost implications of implementing the intervention and cost-effectiveness of the intervention been described?	Yes/no
		Has the impact on the population been presented?	Yes/no
		Is the study ethical?	Yes/no
		What conclusions did the authors reach about the study question?	
		What do you think are the main strengths and weaknesses of the study, and hence the level of evidence it provides?	
Use	Can and should the results of the study influence your public health practice?	How do these results fit into the wider evidence base?	
		Has implementation of the study results been discussed?	
		What are the public health and policy implications of the findings – in general and for your population?	
		Would you use the results of this study?	

Table A.1 (*continued*)

Item on Population Health Evidence Cycle	Can you find this in the paper?	Questions to assess the validity, completeness and transferability of the study.	Answers
			Yes, without more information Maybe, need more evidence – if so, what? No, can ignore the study
	Other issues of internal validity*	In a cross-sectional study, is the item-specific response rate adequate? In a case–control study, are the controls representative of the source population for the cases, are exposures and population representative of your population of interest? In a randomized controlled trial, is the method of random allocation well described and appropriate; if cluster design, is this included in sample size and analysis; is it performed on a representative and relevant population with appropriate exposure/ intervention? In a quasi-experimental design, are appropriate comparison groups included? Is there baseline comparability, and is a relationship between acceptance of intervention, risk factor change and outcome demonstrated? In a longitudinal study, how many subjects reached final follow-up, was the propensity for exposure to the risk factor or intervention examined? In a time series, is there comparison with other time periods or populations without the intervention? In an ecologic study, are the findings supported by individual level data?	

*There are other issues relevant to the study design, which should also be considered, with reference to other sources.

Glossary of terms (based on Last's Dictionary of Epidemiology)

- *Study type*: Qualitative or quantitative. Observation: ecological, cross-sectional (survey), case–control, cohort. Intervention: trial, randomized controlled trial (individuals or groups can be randomized), quasi-experimental, time series (before/after);
- *Sample*: A sample is a group of individuals that is a subset of a population and has been selected from the population in some way;
- *Bias*: Deviation of results or inferences from the truth, or process leading to such deviation. Any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth. (There are three major sources of bias: selection, measurement and confounding.);
- *Exposure variable*: In a study of causation, the potential causative factor is the exposure; in a trial, the intervention is the exposure;
- *Outcome factor*: This is the study endpoint, either the condition that has been caused by (in a causation study), or been prevented by (in a trial), the exposure variable. In a cross-sectional (observational) study, it is sometimes difficult to distinguish which is the exposure and which is the outcome as they are measured at the same time;

- Confounding variable (confounder): A confounder is a variable that is an independent determinant of the outcome of interest and is unequally distributed among the exposed and non-exposed (it is related to both the exposure and the outcome);
- *Validity*: The extent to which a variable or intervention measures what it is supposed to measure or performs what it is supposed to perform. (a) The internal validity of a study refers to the integrity of the experimental design. (b) The external validity of a study refers to the appropriateness by which its results can be applied to non-study patients or populations.

References

1. Evidence Based Medicine Working Group. Users' guide to evidence-based practice. Centre for Health Evidence; 2005. Available at: 〈http://www.cche.net/usersguides/main.asp〉 Edmonton, Alberta. (accessed 20 May 2007).
2. *Critical appraisal skills programme*. Public Health Resource Unit; Oxford 2005. Available at: 〈http://www.phru.nhs.uk/casp/casp.htm〉 (accessed 20 May 2007).
3. Greenhalgh T. *How to read a paper*: the basics of evidence based medicine. *BMJ* Books; 2000.
4. Cochrane Health Promotion and Public Health Field, Victoria, Australia 2006. Available at: http://www.vichealth.vic.gov.au/cochrane/training/index.htm. (Accessed 20 May 2007).
5. Community Guide Branch. *Guide to community preventive services*: methods. Atlanta: Centers for Disease Control and Prevention; 2006.
6. Health-evidence ca. *Promoting evidence based decision making*. Hamilton, Ontorio, 2003. Available at: 〈http://health-evidence.ca/pdf/ValidityTool.pdf〉 (accessed 20 May 2007).
7. Health-evidence ca. Effective Public Health Practice Project. City of Hamilton, 2006. Available at: 〈http://health-evidence.ca/home.aspx〉 (accessed 20 May 2007).
8. Zaza S, Wright-De Aguero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH. et al. *Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. Am J Prev Med* 2000;**18**(Suppl):44–74.
9. Darzins PJ, Smith BJ, Heller RF. How to read a journal article. *Med J Aust* 1992;**157**:389–94.
10. Heller RF. *Evidence for population health*. Oxford: Oxford University Press; 2005.
11. Heller RF, Page JH. A population perspective to evidence based medicine: "evidence for population health". *J Epidemiol Community Health* 2002;**56**:45–7.
12. Weightman A, Ellis S, Cullum A, Sander L, Turley R. *Grading evidence and recommendations for public health interventions: developing and piloting a framework*. London: Health Development Agency; 2005. Available at: 〈http://www.publichealth.nice.org.uk/page.aspx?o=503421〉 (accessed 20 May 2007).
13. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health* 2003;**57**:527–9.
14. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;**94**:400–5.
15. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999; **28**:10–8.
16. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;**56**:119–27.
17. Kirkwood BR, Cousens SN, Victora CG, de Zoysa I. Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. *Trop Med Int Health* 1997;**2**:1022–9.
18. Heller RF, Heller TD, Pattison S. *Putting the public back into public health. Part II. How can public health be accountable to the public? Public Health* 2003;**117**:66–71.
19. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 2004;**4**:22. Available at: 〈http://www.biomedcentral.com/1471-2288/4/22〉 (accessed 20 May 2007).
20. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**:1054–60.
21. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005;**5**:19.
22. The Cochrane Colaboration. *Cochrane handbook for systematic reviews of interventions* 4.2.5 [updated May 2005]. Assessment of study quality. Available at: 〈http://www.cochrane.dk/cochrane/handbook/hbook.htm〉 (accessed 20 May 2007).
23. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490.
24. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;**328**:39–41.