

THEORY AND METHODS

Evidence, hierarchies, and typologies: horses for courses

M Petticrew, H Roberts

Debate is ongoing about the nature and use of evidence in public health decision making, and there seems to be an emerging consensus that the “hierarchy of evidence” may be difficult to apply in other settings. It may be unhelpful however to simply abandon the hierarchy without having a framework or guide to replace it. One such framework is discussed. This is based around a matrix, and emphasises the need to match research questions to specific types of research. This emphasis on methodological appropriateness, and on typologies rather than hierarchies of evidence may be helpful in organising and appraising public health evidence.

Is water fluoridation effective in reducing dental caries in children? Do children learn better in small classes? Can young offenders be “scared straight” through tough penal measures? Can the steep social class gradient in fire related child deaths be reduced by installing smoke alarms?

Anyone faced with making a decision about the effectiveness of an intervention, whether a social intervention, such as the provision of some form of social service, or a clinical intervention, or a decision about the provision of a therapeutic intervention, is faced with a formidable task. The research findings to help answer the question may well exist, but locating that research, assessing its evidential “weight” and relevance, and incorporating it with other existing information is often difficult. One commonly used aid to clinical decision making is the “hierarchy of evidence” (box 1), which lists a range of study designs ranked in order of decreasing internal validity. This tool was developed initially by the Canadian Task Force on the Periodic Health Examination to help decide on priorities when searching for studies to answer clinical questions, and was subsequently adopted by the US Preventive Services Task Force. It has been further developed to include methods for assessing the strength of evidence for public health decision making, and now asks not only “Does it work?” but also “Is it worth it?”.¹⁻⁴ However the “hierarchy of evidence” remains a source of debate, and as McQueen, and Rychetnik and colleagues have recently reminded us,^{5,6} the very use of the term evidence is often contentious when applied to health promotion and public health. Even in medicine the hierarchy of evidence is not without critics, with a recent editorial on the hierarchy of evidence asking whether “the Emperor has no clothes”.⁷ In a recent issue of *Journal of Epidemiology and Community Health* Rychetnik *et al* moved the debate forward by seeking to broaden the scope of the

J Epidemiol Community Health 2003;**57**:527–529

criteria that are used to appraise public health interventions.⁶ This provided a valuable guide to the other types of public health knowledge that are needed to guide interventions, and also outlined the role of different types of research based information; particularly observational and qualitative data. At its heart is a recognition that the hierarchy of evidence is a difficult construct to apply in evidence based medicine, and even more so in public health, and the paper points to the continuing debate about the appropriateness of relying on study design as a marker for the credibility of evidence. Our paper further pursues this issue of the hierarchy of evidence, and advocates its revision on two main grounds. It also suggests a greater emphasis on methodological appropriateness rather than study design.

Box 1 An example of the “hierarchy of evidence”^{17 18}

- 1 Systematic reviews and meta-analyses
- 2 Randomised controlled trials with definitive results
- 3 Randomised controlled trials with non-definitive results
- 4 Cohort studies
- 5 Case-control studies
- 6 Cross sectional surveys
- 7 Case reports

The first of our grounds for contesting the hierarchy is empirical. There is evidence now from a number of recent systematic reviews to contest the view that the hierarchy is “fixed”, with RCTs always occupying the top rungs of the methodological ladder, and observational studies occupying the lower rungs, because of their tendency to produce inflated estimates of the effects of interventions. To bolster this argument several recent studies have assembled data to show that this pattern is not always followed.⁸⁻¹⁰ One of these studies compared observational studies and RCTs

Key points

- The concept of a “hierarchy of evidence” is often problematic when appraising the evidence for social or public health interventions.
- The promotion of typologies rather than hierarchies may be more useful than hierarchies in conceptualising the strengths and weaknesses of different methodological approaches
- A matrix based approach, which emphasises the need to match research questions to specific types of research may prove more useful

See end of article for authors' affiliations

Correspondence to: Mark Petticrew, MRC Social and Public Health Sciences Unit, University of Glasgow, Glasgow G12 8RZ, UK; mark@msoc.mrc.gla.ac.uk

Accepted for publication 13 November 2002

Table 1 An example of a typology of evidence (example refers to social interventions in children) (adapted from Muir Gray²⁴)

Research question	Qualitative research	Survey	Case-control studies	Cohort studies	RCTs	Quasi-experimental studies	Non experimental evaluations	Systematic reviews
Effectiveness Does this work? Does doing this work better than doing that?				+	++	+		+++
Process of service delivery How does it work?	++	+					+	+++
Salience Does it matter?	++	++						+++
Safety Will it do more good than harm?	+		+	+	++	+	+	+++
Acceptability Will children/parents be willing to or want to take up the service offered?	++	+			+	+	+	+++
Cost effectiveness Is it worth buying this service?					++			+++
Appropriateness Is this the right service for these children?	++	++						++
Satisfaction with the service Are users, providers, and other stakeholders satisfied with the service?	++	++	+	+				+

of a range of treatments including calcium channel blockers for coronary artery disease, appendicectomy, and treatments for subfertility, and found that in most cases the estimates of effectiveness were similar.⁷⁻¹⁰ This view is however contradicted by other research showing that non-randomisation does indeed significantly inflate effect size estimates.¹¹ This debate about the relative merits of observational and experimental studies is longstanding (early systematic reviews had compared effect size estimates of the effectiveness of psychotherapy, for example) and the empirical basis is still underdeveloped.⁷ What is clear however is that in certain circumstances the positions at the top of the hierarchy can be reversed; while RCTs remain the gold standard for evaluating effectiveness, methodologically unsound RCTs for example do not invariably “trump” sound observational studies. The hierarchical order also depends on the question asked. For assessment of effectiveness, the hierarchy is generally appropriate, but as Rychetnik *et al* point out, the levels of the hierarchy are about the narrow concept of study design, and not the broader concept of evidence. There has also been some evolution in the original hierarchy of evidence, and the quality of individual studies now receives greater emphasis than was originally the case.¹² For example, Liberati and colleagues have identified nine scales that are currently used to assess levels of evidence. These vary in complexity and the extent to which they assess the methodological quality of the individual studies.¹

The second argument against the use of a hierarchy is that it disregards the issue of methodological aptness—that is, the fact that different types of research question are best answered by different types of study. There is now a considerable body of methodological literature in the social sciences, particularly from qualitative researchers, on the aptness of particular study designs to answering particular research questions, and as Sackett and Wennberg make clear, focusing on the question being asked is more important than squabbling over the “best” method.¹³⁻¹⁸ End point users, policy makers, and practitioners in particular ask many questions about interventions that are not just about effectiveness. This possibility is sometimes obscured by the existence of a single hierarchy, and

the possibility that in certain circumstances the hierarchy may even be inverted, placing for example qualitative research methods on the top rung, is not widely appreciated. The hierarchy also obscures the synergistic relation between RCTs and qualitative research, and (particularly in the case of social and public health interventions) the fact that both sorts of research are often required in tandem; robust evidence of outcomes comes from randomised controlled trials but evidence of the process by which those outcomes were achieved, the quality of implementation of the intervention, and the context in which it occurred is likely to come from qualitative and other data. The use of RCTs and qualitative methods is therefore less of a choice between extremes than the hierarchy implies, and effective implementation of an intervention ideally requires both sorts of information.¹⁹

A related problem lies in the stark use of the term “evidence”. It is not uncommon for discussion papers to use the terms “evidence,” “evidence based”, and “hierarchies of evidence,” while avoiding any discussion what sort of evidence they are advocating (or rejecting). For epidemiological questions relating to “real world” risk factors that are not amenable to randomisation (for example, does smoking cause cancer?) a particular sort of data is required, with prospective cohort studies at the top of the hierarchy. Qualitative studies, expert opinion, and surveys on the other hand are likely to have crucial lessons for those wanting to understand the process of implementing an intervention, what can go wrong, and what the unexpected adverse effects might be when an implementation is rolled out to a larger population. A different sort of hierarchy is again implied. Overall, information on both outcomes and processes are of value. Knowing that an intervention works is no guarantee that it will be used, no matter how obvious or simple it is to implement. For example, it is nearly 150 years since Semmelweis’ trial showed that handwashing reduces infection, yet healthcare workers’ compliance with handwashing remains poor.²⁰ Even the most simple, cost effective, and logical intervention fails if people will not carry it out.²¹

With increasing interest in the effectiveness of social interventions and the development of UK and international

initiatives in this area (<http://campbell.gse.upenn.edu/> and <http://www.evidencenetwork.org/>) a single hierarchy of methods has become increasingly unhelpful, and at present certainly misrepresents the interplay between the question being asked and the type of research required most suited to answering it. For this reason, a matrix, or a typology, may be a useful construct. Different research methods are, after all, more or less good at answering different kinds of research question. A randomised controlled trial, well conducted, can tell us which kind of smoke alarm is most likely to be functioning 18 months after installation, but it cannot tell us what the best way is to work effectively with housing managers on making sure smoke alarms are installed effectively and cost effectively, while ensuring that the households of the most vulnerable tenants are included. The obstacles and levers for the uptake of research findings are also likely to be understood through methods different from those usually found at the top of the hierarchy.^{16 22 23} It may therefore be most useful to think of how you can best use the wide range of evidence available—and particularly to consider what types of study are most suitable for answering particular types of question.

TYPOLOGICAL TRIAGE

One example of such an approach is suggested by Muir Gray, who suggests the use of a typology rather than a hierarchy to indicate schematically the relative contributions that different kinds of methods can make to different kinds of research questions.²⁴ This simple matrix was originally designed to help health care decision makers determine the appropriateness of different research methodologies for evaluating different outcomes, and was intended to be applied to health care interventions. However it also has a wider applicability (table 1).

It can be seen from this table that different research methods are at, or close to the top of different hierarchies, depending on the questions asked. Using this example of the contribution of different kinds of research, and in a spirit of methodological pluralism, we therefore suggest that the promotion of typologies rather than hierarchies may be more useful than hierarchies in conceptualising the strengths and weaknesses of different methodological approaches.

CONCLUSION

“Horses for courses” is not a dramatic theoretical insight, but the energy dissipated in debates on methodological primacy could be better used were this aphorism to be accepted. There are a number of important areas where this released energy could be used, key among which is further work on the synthesis of non-trial data (both quantitative and qualitative). Much information about the health and other impacts of community interventions falls into this category, yet it is not helpful to reiterate that the best evidence is lacking. The immediate methodological challenges (as Rychetnik *et al* emphasise), are to determine how complete the evidence needs to be before recommendations can be made, and how much weight should be given to non-experimental data when making decisions about provision of services, or about policies.

ACKNOWLEDGEMENTS

David McQueen, Lucie Rychetnik.

Authors' affiliations

M Petticrew, MRC Social and Public Health Sciences Unit, University of Glasgow, UK

H Roberts, City University, London, UK

Funding: MP and HR receive funding as part of the ESRC “Evidence Network”. MP is funded by the Chief Scientist Office of the Scottish Executive Department of Health, and part of this work was completed while he was visiting fellow at VicHealth, Melbourne.

Competing interests: none declared.

REFERENCES

- 1 **Liberati A**, Buzzeiti, Grilli R *et al*. Which guidelines can we trust? *West J Med* 2001;**174**:262.
- 2 **Truman BI**, Smith-Akin CK, Hinman AR, *et al*. Developing the Guide to Community Preventive Services—overview and rationale. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;**18** (suppl 1):18–26.
- 3 **Novick LF**. The Guide to Community Preventive Services. A public health imperative. *Am J Prev Med* 2001;**21** (suppl 4):13–15.
- 4 **Saha S**, Hoerger TJ, Pignone MP, *et al*. Cost Work Group, Third US Preventive Services Task Force. The art and science of incorporating cost effectiveness into evidence-based recommendations for clinical preventive services. *Am J Prev Med* 2001;**20** (suppl 3):36–43.
- 5 **McQueen D**. The evidence debate. *J Epidemiol Community Health* 2002;**56**:83–4.
- 6 **Rychetnik L**, Frommer M, Hawe P, *et al*. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;**56**:119–27.
- 7 **Bigby M**. Challenges to the hierarchy of evidence: does the emperor have no clothes? *Arch Dermatol* 2001;**137**:345–6.
- 8 **Concato J**, Shah N, Horwitz RJ. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.
- 9 **McKee M**, Britton A, Black N, *et al*. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;**319**:312–15.
- 10 **Benson K**, Hartz AJ. Comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**:1878–86.
- 11 **Schulz KF**, Chalmers I, Hayes RJ, *et al*. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.
- 12 **Harbour R**, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;**323**:334–6.
- 13 **Sackett DL**, Wennberg JE. Choosing the best research design for each question. *BMJ* 1997;**315**:1636.
- 14 **Hammersley M**. *The dilemma of qualitative method: Herbert Blumer and the Chicago Tradition*. London: Routledge, 1989.
- 15 **Roberts H**. Qualitative research methods in interventions in injury. *Arch Dis Child* 1997;**76**:487–9.
- 16 **Davies P**. Contributions from qualitative research. In: Davies H, Nutley S, Smith P, eds. *What works? Evidence-based policy and practice in public services*. Bristol: The Policy Press, 2000.
- 17 **Brannen J**. *Mixing methods: qualitative and quantitative research*. Aldershot: Avebury, 1992.
- 18 **Popay J**, Williams G. Qualitative research and evidence-based healthcare. *J R Soc Med* 1998;**91** (suppl 35):32–7.
- 19 **MRC**. *Increasing the prevalence of functioning smoke alarms in disadvantaged inner city housing: a randomised controlled trial (ongoing)*. MRC Serial number G9901222. (<http://www.controlled-trials.com/mrct/>).
- 20 **Teare L**, Cookson B, Stone S. Hand hygiene. *BMJ* 2001;**323**:411–12.
- 21 **Sheldon TA**, Guyatt GH, Haines A. Getting research findings into practice: When to act on the evidence. *BMJ* 1998;**317**:139–42.
- 22 **Halladay M**, Bero L. Implementing evidence-based practice in health care. *Public Money and Management* 2000;**20**:43–50.
- 23 **Barnardo's Research & Development Team**. *What works? Making connections: linking research and practice*. Ilford: Barnardo's, 2000.
- 24 **Muir Gray JM**. *Evidence-based healthcare*. London: Churchill Livingstone, 1996.