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.

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

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Systematic reviews and meta-analyses</td>
</tr>
<tr>
<td>2 Randomised controlled trials with definitive results</td>
</tr>
<tr>
<td>3 Randomised controlled trials with non-definitive results</td>
</tr>
<tr>
<td>4 Cohort studies</td>
</tr>
<tr>
<td>5 Case-control studies</td>
</tr>
<tr>
<td>6 Cross sectional surveys</td>
</tr>
<tr>
<td>7 Case reports</td>
</tr>
</tbody>
</table>

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.19 One of these studies compared observational studies and RCTs criteria that are used to appraise public health interventions.” 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.

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.
being asked is more important than squabbling over the
Sackett and Wennberg make clear, focusing on the question
designs to answering particular research questions, and as
from qualitative researchers, on the aptness of particular study
methodological literature in the social sciences, particularly
by different types of study. There is now a considerable body of
fact that different types of research question are best answered
it disregards the issue of methodological aptness—that is, the
methodological quality of the individual studies.
These vary in complexity and the extent to which they assess
nine scales that are currently used to assess levels of evidence.

The second argument against the use of a hierarchy is 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.6,7 This view is however contra-
dicted by other research showing that non-randomisation
does indeed significantly inflate effect size estimates.8 This
debate about the relative merits of observational and
experimental studies is longstanding (early systematic re-
views had compared effect size estimates of the effectiveness
of psychotherapy, for example) and the empirical basis is still
underdeveloped.9 What is clear however is that in certain cir-
cumstances 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 hier-
archical order also depends on the question asked. For assess-
ment 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.7 For example, Liberati and colleagues have identified
ten criteria 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
different fact that 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.1,9 End point users, policy makers, and practi-
tioners 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 hier-
archy 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 out-
comes 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 there-
fore less of a choice between extremes than the hierarchy
implies, and effective implementation of an intervention ide-
ally requires both sorts of information.10

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’ com-
pliance with handwashing remains poor.11 Even the most sim-
ple, cost effective, and logical intervention fails if people will
not carry it out.21

With increasing interest in the effectiveness of social inter-
ventions and the development of UK and international

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

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. 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. 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**


7 Bigby M. Challenges to the hierarchy of evidence: does the emperor have no clothes? Arch Dermatol 2001; 137:345–6.


19 MRC. Increasing the prevalence of functioning smoke alarms in disadvantaged inner city housing in a randomised controlled trial (ongoing). MRC Serial number G9901222. [http://www.controlled-trials.com/mrc/].


Evidence, hierarchies, and typologies: horses for courses

M Petticrew and H Roberts

J Epidemiol Community Health 2003 57: 527-529
doi: 10.1136/jech.57.7.527

Updated information and services can be found at:
http://jech.bmj.com/content/57/7/527

These include:

References
This article cites 17 articles, 8 of which you can access for free at:
http://jech.bmj.com/content/57/7/527#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/