THEORY AND METHODS

Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go?

David Ogilvie, Matt Egan, Val Hamilton, Mark Petticrew

J Epidemiol Community Health 2005;**59**:886–892. doi: 10.1136/jech.2005.034199

See end of article for authors' affiliations

Correspondence to: Dr D Ogilvie, MRC Social and Public Health Sciences Unit, University of Glasgow, 4 Lilybank Gardens, Glasgow G12 8RZ, UK; d.ogilvie@msoc.mrc.gla.ac.uk

Accepted for publication 19 April 2005

Study objective: There is little guidance on how to select the best available evidence of health effects of social interventions. The aim of this paper was to assess the implications of setting particular inclusion criteria for evidence synthesis.

Design: Analysis of all relevant studies for one systematic review, followed by sensitivity analysis of the effects of selecting studies based on a two dimensional hierarchy of study design and study population.

Setting: Case study of a systematic review of the effectiveness of interventions in promoting a population shift from using cars towards walking and cycling.

Main results: The distribution of available evidence was skewed. Population level interventions were less likely than individual level interventions to have been studied using the most rigorous study designs; nearly all of the population level evidence would have been missed if only randomised controlled trials had been included. Examining the studies that were excluded did not change the overall conclusions about effectiveness, but did identify additional categories of intervention such as health walks and parking charges that merit further research, and provided evidence to challenge assumptions about the actual effects of progressive urban transport policies.

Conclusions: Unthinking adherence to a hierarchy of study design as a means of selecting studies may reduce the value of evidence synthesis and reinforce an ''inverse evidence law'' whereby the least is known about the effects of interventions most likely to influence whole populations. Producing generalisable estimates of effect sizes is only one possible objective of evidence synthesis. Mapping the available evidence and uncertainty about effects may also be important.

Despite increasing calls for systematic reviews of health effects of social interventions, there is little methodological research or even guidance on how such reviews should be done. We have lifted the lid on the ''private life'' of the input side of one such systematic review to expose some of our methodological processes and decisions to critical analysis.[1] In a companion paper, we set the scene and examined one phase of the review, the search for evidence.[2] In this paper, we examine another phase of the review: the selection of evidence for inclusion. We investigate the effect of varying our inclusion criteria on the findings and overall value of the review.

SELECTING EVIDENCE FOR INCLUSION

Researchers designing systematic reviews of intervention studies are advised to specify their research questions in terms of four facets: the intervention, the population receiving the intervention, the outcome of interest, and the study designs deemed worthy of inclusion.[3] This approach is undoubtedly helpful for structuring research questions and protocols, but we aimed to synthesise population level evidence in a cross disciplinary field where comparatively little empirical intervention research has been done. A broad understanding of population health and its wider determinants implied a need to frame our primary research question rather differently. We were not asking, for example, ''What is the evidence that traffic calming leads to a change in travel behaviour?'', but rather ''What interventions, of any kind, lead to such a change?'' In other words, we focused on the outcome of interest and were open to the possibility that any kind of intervention might contribute towards achieving it. This is an example of addressing a ''broad'' review question—acknowledged as a valid, but often difficult, type of review to carry out.[4] Broad questions are also often appropriate in other types of evidence synthesis, such as that used in health impact assessment.[5]

Many published systematic reviews have only considered evidence from randomised controlled trials (RCTs). The motivation is to minimise bias, but the Cochrane handbook recognises that this can compromise the relevance of a review, and asks (but does not answer) the question ''How far is it possible to achieve a higher level of relevance by including evidence other than that derived from RCTs without violating the central principle: minimising bias?''[4] There are already precedents for varying the inclusion criteria for study design according to the nature of the available evidence. For example, although some reviews published by the Cochrane Tobacco Addiction Group are restricted to RCTs, those on community or population level interventions include other study designs including, in some cases, uncontrolled before and after studies.[6 7]

It is increasingly recognised that the usual approach to selecting studies based on a ''hierarchy of evidence'' may rely too heavily on study design as a marker of validity or utility.[8–10] This may favour interventions most amenable to certain types of study design, particularly those with a medical rather than a social focus and those that target individual people rather than populations.[8] This type of bias has been described as ''methodological imperialism'' that could distort, rather than strengthen, the evidence base.[11]

The relative lack of methodological research on how to deal with evidence from studies other than RCTs may make researchers feel vulnerable at key decision points in the process of synthesising evidence.[12] In this paper we describe how we selected studies for inclusion. We then analyse the utility of the different types of studies identified, report a

Criteria for relevance

Studies were selected as relevant studies if:

- they described an intervention aimed at promoting, or likely to be associated with, a shift from using cars towards physically active modes of transport, applied to an urban population in a developed country, and
- they reported data on the choice of mode of transport in the population before and during or after the intervention.

sensitivity analysis of the effects of excluding certain types of evidence, and reflect on what systematic reviews in this field can be expected to contribute to the evidence base.

METHODS

We have reported details of our methods previously.[13] Briefly, we designed a wide search strategy defined entirely in terms of the outcome of interest. We screened the titles and abstracts, examined the full text of any documents that appeared relevant, and finally identified 69 relevant studies that met our preliminary criteria (see box).

We carried out full data extraction and critical appraisal on all of these relevant studies, and therefore formed an overview of the full range of study populations, interventions, and study designs available in the field as well as the range of outcome metrics used and effect sizes identified.

It became clear that both the types of study design and the nature of the study populations varied widely. Some studies had used comparatively robust methods to measure, for example, changes in vehicle flows along certain roads, but these studies could tell us nothing about the people using those vehicles or about their non-vehicular (walking) trips. Similarly, we found studies showing how the distribution of transport mode choice had changed among weekend shoppers interviewed in a city centre street, but these studies could tell us nothing about where the shoppers had come from or whether their overall travel behaviour had changed.

We also found particular difficulty in deciding what to do with articles—typically book chapters—about ''successful'' towns or cities in which trends in travel patterns were linked post hoc to a variety of interventions, often part of a complex integrated urban policy that included land use planning, public transport improvements, widespread traffic restraint, cycle routes, pedestrianisation, and related measures. These articles did not seem to report the results of specific studies of specific interventions as such, so we characterised them as ''case studies'' in which authors had reported trends of interest to us, but had not presented data in a way that enabled us to assess the strength of the causal assertions being made.

These findings led us to devise a simple matrix, or two dimensional hierarchy, of study utility (table 1). We categorised studies not only on the study design (a marker of internal validity) but also on the study population, which we took as our primary marker of external validity—in other words, a marker of how useful the study would be for answering our question about changes in population health and health determinants. We plotted the distribution of all relevant studies in this matrix and used it to specify our final inclusion criteria. We further assessed and summarised the internal validity of included studies using 10 methodological criteria.[13]

When our review was complete, we also conducted a sensitivity analysis to examine what the content and findings of the review would have been if we had taken one of two extreme approaches to inclusion—either (a) by restricting the review to randomised controlled trials, or (b) by including all relevant studies. This sensitivity analysis was intended to answer two questions: were the conclusions of our review sensitive to the inclusion criteria, and could we have reached our conclusions more efficiently?

RESULTS

Two dimensional hierarchy

We examined the distribution of studies in the matrix (fig 1) and chose final thresholds for inclusion. These were, of course, still somewhat arbitrary but were based on having reviewed all available relevant studies in detail.

We first excluded studies whose design was neither prospective nor controlled (n = 29). We then excluded studies whose populations did not represent a local population or subset thereof (n = 9). This left 31 studies (represented by the dark columns in the figure). Nine of these were subsequently excluded on the grounds that they contained inadequate information about methods, results or both, leaving 22 studies finally included in the review.

Sensitivity analysis

Effect of including only RCTs

We found only three RCTs. If we had included only these studies, we would have benefited from reviewing a small set of studies that were well written and comparatively easy to appraise. These were also the only studies that contained robust data on direct health outcomes. However, we would only have been able to include evidence about two small categories of intervention: targeted behaviour change programmes for commuters, and school travel coordinators. We would not have identified any evidence about, or perhaps even the existence of, any population wide health promotion activities, ''environmental'' engineering or transport service developments, or financial incentives, and we would not have

Table 1 Two dimensional hierarchy of study utility

Study design	Study population
Randomised controlled trial	Households or local residents
Controlled panel study (repeated measures on the same participants)	Subset of local population (drivers, commuters, or school pupils)
Controlled repeated cross sectional study	Participants in a targeted intervention already selected from one of the groups above
Controlled retrospective study	Passers by at, or visitors to, a study location
Uncontrolled panel study (repeated measures on the same participants)	Patients*
Uncontrolled repeated cross sectional study	Vehicles
Uncontrolled retrospective study	Population not clear
Case study of trends in mode share	
Design not clear	

*The effectiveness of interventions given in a clinical setting was outside the scope of the review.

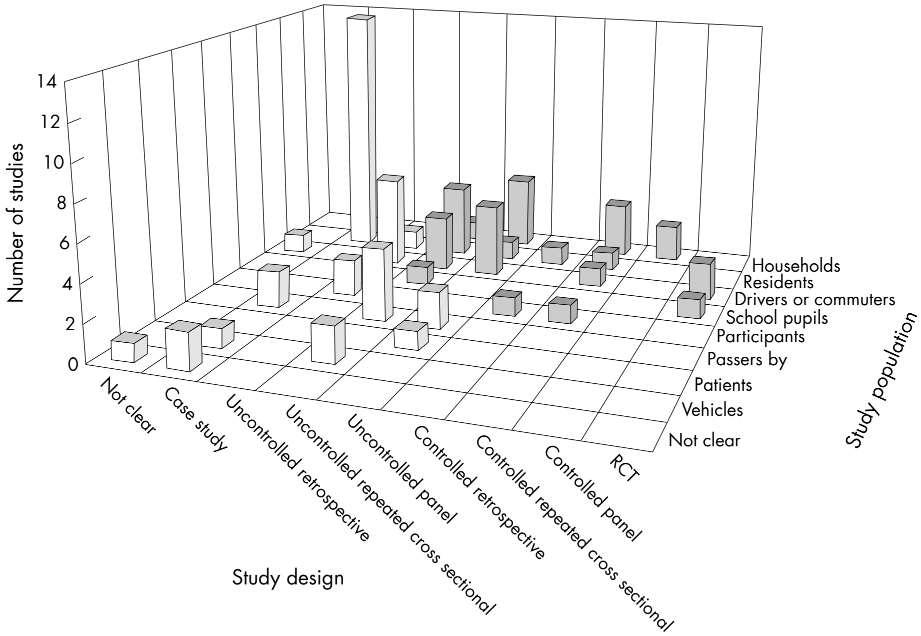

Figure 1 Distribution of studies by design and population.

identified any of the studies that indicated possible unexpected or inequitable effects of interventions.[13]

Evidence provided by excluded studies

We identified several types of evidence provided by studies we did exclude, which are summarised in table 2, grouped by type of intervention.

A larger taxonomy of interventions of interest

Some specific types of intervention were only represented in excluded studies: health walks, parking charges, and fuel rationing. Most of these studies indicated potential for a positive effect, albeit based on designs with important methodological weaknesses with respect to our review question. These types of intervention therefore merit more detailed consideration by researchers and policymakers.

Evidence about some interventions consistent with the stronger evidence already included in the review

We had found the strongest evidence of positive effects in the area of targeted behaviour change programmes (based on six studies of four interventions).[13] Two excluded studies of targeted programmes also identified potential for positive effects, as did two excluded studies of workplace schemes involving free bikes. We also found a large number of excluded studies of engineering measures whose findings were broadly consistent with our primary finding of little evidence of positive effects, and single excluded studies of road user charging and alternative transport services that did not contradict our primary findings.

Evidence about one category of intervention that could contradict our primary findings

We excluded two studies of publicity campaigns for sustainable transport that both claimed a substantial positive effect. Neither study was reported in sufficient detail for our purposes (for example, there were no details of sampling method, response rate, survey instrument, and so on), we could not find any more detailed reports, and authors did not reply to a request for more information. It is therefore possible that evidence exists to contradict our primary finding of little evidence of effectiveness for publicity campaigns, although it seems unlikely that such evidence would be strong.

Evidence to challenge assumptions about ''successful'' cities

Even if it were possible to attribute the observed trends in travel patterns in ''case study'' cities to part or all of their multifaceted urban transport policies, a positive change (in our terms) was only actually reported in three of the 13 cities, and in two of these that positive change was only seen for trips into the city centre and not for residents' trips overall. Where modal shifts were reported, these were more likely to be, for example, an increase in public transport at the expense of all other modes including walking and cycling.

What is already known on this subject?

- We need better syntheses of evidence about the effects of interventions to influence the wider determinants of health
- Some have questioned whether selecting evidence according to a rigid, unidimensional hierarchy based on study design—for example, only including randomised controlled trials—is appropriate in this field
- We lack an accepted, evidence based methodology for selecting useful evidence for inclusion in evidence synthesis.

DISCUSSION

Hierarchies of evidence for public health

We reported previously that the most robust evidence of effectiveness was concentrated around interventions targeted on motivated groups of volunteers.[13] Our subsequent analysis shows that this ''evidence bias'' may reflect, at least partly, an ''evaluative bias'': other types of intervention (especially those applied to whole populations or areas) have tended to be evaluated using less rigorous methods. For those interested in improving population health, the most useful evidence is likely to come from population level studies with designs of high internal validity—those located in the far

Table 2 Evidence from excluded studies

Category of intervention	Included studies	Summary of findings from included studies*	Excluded studies	Primary reasons for exclusion†	Summary of findings from excluded studies*	Comparison with included studies
Targeted behaviour change programmes	One randomised trial, three controlled prospective studies of the same intervention in different settings, and two uncontrolled prospective studies	Modal shift in motivated subgroups (five out of six studies)	One uncontrolled prospective study of advising patients to take more exercise	Population	Both identified potential for modal shift	Consistent
			One pilot study of targeted information for households and commuters	Information		
Agents of change and publicity campaigns	One randomised controlled trial of school travel coordinators	Small modal shift in only one relatively weak study out of four	One case study of a travel management association	Design	"Inability to achieve any significant shift in travel behaviour"	Consistent
	One controlled prospective and two uncontrolled prospective studies of publicity campaigns		Two studies of sustainable transport campaigns, one uncontrolled prospective and one controlled retrospective	Information	Both suggested substantial modal shift	Contradictory, but of unknown validity
Health walks	No included studies	—	Two uncontrolled retrospective studies of participants	Design	About a quarter of participants claimed to have changed their travel behaviour	Additional category of intervention with potential for positive effect—requires further research
Cycling promotion	One uncontrolled prospective study (also included in the group of targeted behaviour change programmes above)	Intensive targeted programme including free bike produced a modal shift	Two uncontrolled retrospective studies of free workplace bikes plus local infrastructural improvements	Design	Both identified the potential for modal shift	Consistent
			One uncontrolled prospective study of promoting cycling for shopping	Population	No evidence of modal shift	Additional category of intervention with no evidence of positive effect
Networks of routes for cyclists and pedestrians	One controlled prospective study and two uncontrolled prospective studies	Increase in cycling mode share (only) in only one of three studies	Five studies, all either case studies reported with scant detail or based on vehicle counts	Design (4) Population (1)	One study based on vehicle counts reported increases in cycling, but no data on walking; the others showed no evidence of modal shift	Consistent
Traffic restraint	Three uncontrolled prospective studies: one of city centre traffic restraint, one of bypasses and one of 20 mph (30 km/h) zones	Small modal shift in only one of three studies	One uncontrolled retrospective study of neighbourhood traffic calming schemes	Design	Small proportions of residents claimed to have changed their travel behaviour	Consistent overall
			Seven studies of a variety of urban traffic restraint schemes, either using unclear or case study designs or based on vehicle counts or shoppers	Design (3) Population (4)	Most showed no clear evidence of modal shift; where a modal shift was demonstrated, it was only among weekend city centre shoppers	
			Two uncontrolled prospective, or partially-prospective, studies of drivers' trips across road bridges	Population	Small proportions claimed to have shifted mode after the bridge was closed	
			Four studies (one controlled prospective and three uncontrolled prospective) which were reported with scant detail or which conflated bike and moped trips	Information	Impossible to assess	
Road user charging	One uncontrolled prospective study	No evidence of modal shift	One pilot study of the effect of an in-car charging unit on commuting journeys	Information	"Another peak-period alternative for two of the sample was cycling"	Consistent— no clear quantification of modal shift
Parking charges	No included studies	–	One uncontrolled prospective study of town centre parking charges	Information	Identified potential for modal shift among commuters, but not among residents	Additional category of intervention with potential for positive effect — requires further research
			One uncontrolled retrospective study of parking charges at government workplaces	Design	No overall modal shift	

Table 2 Continued

Category of intervention	Included studies	Summary of findings from included studies*	Excluded studies	Primary reasons for exclusion†	Summary of findings from excluded studies*	Comparison with included studies
Fuel rationing	No included studies	–	One uncontrolled retrospective study of changes to car trips after introduction of fuel rationing	Design	Identified potential for substantial modal shift	Additional category of intervention with potential for positive effect — requires further research
Providing alternative services	One uncontrolled prospective study of a new railway station One controlled prospective study of a car sharing club One controlled retrospective study of neighbourhood telecommuting centres	Modal shift in only one of three studies	One uncontrolled retrospective study of introducing a passenger only commuter ferry	Design	Not designed to collect relevant outcome data	Consistent — no clear quantification of modal shift
Multifaceted urban transport policies	No included studies	–	Thirteen case studies of trends in mode share in cities where multifaceted transport policies had been enacted over a long period	Design	Three cities reported a modal shift for trips to the city centre only, but in two of these cities there was no modal shift in residents' trips overall. All other cities reported zero or negative shifts	Additional category of intervention with potential for positive effect — requires further research

*These are the reviewers' summaries, not those of the authors of the primary studies, except where shown in inverted commas—in which case the quoted data constitute the relevant outcome data reported by the authors and do not require further summary. †Studies were excluded because the study *design* did not meet final inclusion criteria, because the study *population* did not meet the final inclusion criteria, or because they fulfilled these criteria but provided inadequate *information* about methods, results, or both. Some studies failed more than one criterion.

What does this study add?

- Relying on randomised controlled trials would have seriously compromised the scope and value of our evidence synthesis
- Relevant, population level evidence is dispersed across a wide range of types of study; mapping all of this evidence is a useful exercise in its own right and may be an important part of the process of selecting the most useful evidence for final inclusion
- Filtering out studies for exclusion without examining them in detail may deprive both reviewers and users of important evidence and insights.

right hand corner of our matrix. In reality, however, the distribution of the available evidence is skewed. Many genuinely population or area level interventions have been studied using comparatively weak study designs, or not studied at all in our terms, and the "gold standard" randomised controlled trial methodology has only once been applied to an area level intervention in this field. In other words, we know least about the effects of those interventions that are most likely to influence the wider determinants of health—a problem described elsewhere as an evidence deficit, or "inverse evidence law".[14 15]

Our findings therefore support concerns raised elsewhere that rigid or simplistic adherence to a hierarchy of study design as the primary marker of study utility may be unhelpful, particularly in the fields of health promotion and public health.[8–11] For example, the interventions studied in RCTs represent only a small subset of all those that could be or have been advocated. We support the use of RCTs where possible, but many interventions of interest in public health cannot be studied in this way for scientific, political, or practical reasons.[16 17] Extending the inclusion criteria as far as we did enabled us to review evidence about a much larger range of interventions and identify some pointers towards potential unexpected effects.[13] Having re-examined the evidence contained in the studies we did exclude, we do not think that we unwittingly censored any convincing evidence of effectiveness. However, we did identify some interventions that could have positive effects and should be the subject of further research. We also identified other studies, notably the case studies of cities frequently cited as examples of good practice in transport policy, in which there was no actual evidence of success in promoting walking and cycling as an alternative to using cars.

Best available evidence

The Cochrane handbook acknowledges a place for systematic reviews that address broad questions, but warns of potential difficulties with synthesising and interpreting data from a large set of heterogeneous studies. Identifying all relevant studies is part of what distinguishes a systematic review from a traditional narrative review.[4 18] We developed our inclusion criteria iteratively by searching widely, fully appraising all relevant studies, and thereby forming an overview of all available evidence before deciding what should be included.[19] Others have also acknowledged that it may not always be possible to specify inclusion criteria in advance[12] and that the definition of "relevant" studies may emerge through an extended process of searching, scanning, production of criteria, and further searching.[20]

Our approach reflects the principle described by Slavin as "best evidence synthesis", in other words, not allowing a desire for the "best" evidence to stand in the way of using the best available evidence.[21] In a review of the effectiveness of strategies for transferring patient information, Badger *et al* framed the reviewer's task as to review and evaluate ''such research as is available''. This did not mean they abandoned the need for critical appraisal; rather, they made informed judgments about the utility of different studies in the light of the whole range of studies available.[22]

What is evidence synthesis for?

The answer to the question ''How low should you go?'' depends on what researchers think evidence synthesis is for and what evidence is available in a given topic area. Evidence synthesis is often undertaken with the objective of pooling results to produce generalisable estimates of effect size, preferably (in some circles) using the formal technique of meta-analysis. We found that the ''best available evidence'' in our topic area did not permit us to do this. Is such an objective necessary for a systematic review of intervention studies? A recent editorial highlighted disagreement between authors and peer reviewers over whether the topic of a systematic review of community based interventions was sufficiently coherent or precise to permit generalisation, and argued that learning in public health is best promoted by the critical sharing of evidence, not by censoring suboptimal evidence.[23] Systematic reviews may contribute to public health decision making in various ways.[24] Hammersley has argued that ''synthesis'' may mean different things to different people, identifying one particular use of the word common among qualitative researchers but not systematic reviewers: producing a mosaic or map in which the distinctive, complementary contributions from different studies are combined to produce a ''bigger picture''.[25] This meaning, which is in sharp contrast with the pooling of data from homogeneous studies in a meta-analysis, perhaps reflects more closely what our review achieved. One aspect of this ''bigger picture'' is the articulation of uncertainty—about the effectiveness of interventions, about the research undertaken on them, and about their potential for unexpected or inequitable effects.[13] Alderson has argued that we should not be embarrassed to admit uncertainty, but should admit it so that the evidence base can then be strengthened.[26] We do not, of course, suggest that reviewers should incorporate the results of less robust studies uncritically in their synthesis of evidence of effectiveness, because doing so can significantly change the resulting recommendations about what interventions are labelled ''effective''.[27] However, our sensitivity analysis shows that our excluded but relevant studies could make an additional valuable contribution to the larger mosaic, even though we seemed to have been justified in excluding them from the primary synthesis of evidence of effectiveness. Indeed, the preliminary mapping of all available evidence has been an explicit part of the process of some systematic reviews.[28]

Is the systematic review a fraud?

Handbooks and protocols for systematic reviews, and the reports of their findings, can often given the impression of a linear, rational research process driven by a set of decisions made a priori. But the further a review strays from the world of the placebo controlled drug trial, the less tenable this idea becomes. In this respect, a report of a systematic review is no different from any other scientific publication: it can give a misleading narrative of the research process.[29] The evidence never speaks for itself, but is always open to interpretation, and there are elements of the review process that entail judgment and cannot be made entirely transparent or replicable.[25 30] Designing and conducting systematic reviews of the health effects of interventions to influence the wider determinants of health is a difficult task for which a standard methodology—whether for searching, study selection, or any other part of the process—has not yet emerged. The methods we have adopted, and our decision to scrutinise them, are open to challenge. None the less, we suggest that it is preferable to reach conclusions, however tentative, that are based on the best available evidence rather than simply stating that no evidence is available.[16]

Authors' affiliations

D Ogilvie, M Egan, M Petticrew, MRC Social and Public Health Sciences Unit, University of Glasgow, UK
V Hamilton, Development and Alumni Office, University of Glasgow

Funding: the review was funded by the Chief Scientist Office of the Scottish Executive Health Department and by the ESRC Evidence Network. DO is now funded by a Medical Research Council fellowship. The funding sources played no part in the design, analysis, interpretation, or writing up of the study or in the decision to publish.

Competing interests: none known.

Ethical approval: not required.

A list of references to the studies excluded from the systematic review is available on request from the first author.

REFERENCES

1 **Roberts I**, Schierhout G. The private life of systematic reviews. *BMJ* 1997;**315**:686–7.
2 **Ogilvie D**, Hamilton V, Egan M, *et al*. Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *J Epidemiol Community Health* 2005;**59**:804–8.
3 **NHS Centre for Reviews and Dissemination**. *Undertaking systematic reviews of research on effectiveness: CRD report no 4*. York: NHS Centre for Reviews and Dissemination, 2001.
4 **Alderson P**, Green S, Higgins J. Cochrane reviewers' handbook 4.2.2 [updated March 2004]. *Cochrane Library*. Issue 1. Chichester: Wiley, 2004.
5 **Mindell J**, Boaz A, Joffe M, *et al*. Enhancing the evidence base for health impact assessment. *J Epidemiol Community Health* 2004;**58**:546–51.
6 **Serra C**, Cabezas C, Bonfill X, *et al*. Interventions for preventing tobacco smoking in public places. *Cochrane Library*. Issue 1. Chichester: Wiley, 2004.
7 **Stead L**, Lancaster T. Interventions for preventing tobacco sales to minors. *Cochrane Library*. Issue 1. Chichester: Wiley, 2004.
8 **Rychetnik L**, Frommer M, Hawe P, *et al*. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;**56**:119–27.
9 **Petticrew M**, Roberts H. Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health* 2003;**57**:527–9.
10 **Glasziou P**, Vandenbroucke J, Chalmers I. Assessing the quality of research. *BMJ* 2004;**328**:39–41.
11 **Green J**, Tones K. Towards a secure evidence base for health promotion. *J Public Health Med* 1999;**21**:133–9.
12 **Dickersin K**. Systematic reviews in epidemiology: why are we so far behind? *Int J Epidemiol* 2002;**31**:6–12.
13 **Ogilvie D**, Egan M, Hamilton V, *et al*. Promoting walking and cycling as an alternative to using cars: systematic review. *BMJ* 2004;**329**:763–6.
14 **Nutbeam D**. How does evidence influence public health policy? Tackling health inequalities in England. *Health Promot J Aust* 2003;**14**:154–8.
15 **Petticrew M**, Whitehead M, Macintyre S, *et al*. Evidence for public health policy on inequalities: 1. The reality according to policymakers. *J Epidemiol Community Health* 2004;**58**:811–16.
16 **Thomson H**, Hoskins R, Petticrew M, *et al*. Evaluating the health effects of social interventions. *BMJ* 2004;**328**:282–5.
17 **Barreto M**. Efficacy, effectiveness, and the evaluation of public health interventions. *J Epidemiol Community Health* 2005;**59**:345–6.
18 **Last J**, ed. *A dictionary of epidemiology*. Oxford: Oxford University Press, 2001.
19 **Hawker S**, Payne S, Kerr C, *et al*. Appraising the evidence: reviewing disparate data systematically. *Qual Health Res* 2002;**12**:1284–99.
20 **McNally R**, Alborz A. Developing methods for systematic reviewing in health services delivery and organization: an example from a review of access to health care for people with learning disabilities. Part 1. Identifying the literature. *Health Info Libr J* 2004;**21**:182–92.
21 **Slavin R**. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995;**48**:9–18.
22 **Badger D**, Nursten J, Williams P, *et al*. Should all literature reviews be systematic? *Eval Res Educ* 2000;**14**:220–30.
23 **Moller J**. Reconsidering community based interventions. *Inj Prev* 2004;**10**:2–3.
24 **Petticrew M**. Presumed innocent: why we need systematic reviews of social policies. *Am J Prev Med* 2003;**24**:2–3.
25 **Hammersley M**. *Systematic or unsystematic, is that the question? Some reflections on the science, art and politics of reviewing research evidence*. London: Health Development Agency Public Health Steering Group, 2002,

http://www.hda.nhs.uk/evidence/sys_unsys_phesg_hammersley.pdf (accessed 18 Jan 2005).
26 **Alderson P**, Roberts I. Should journals publish systematic reviews that find no evidence to guide practice? Examples from injury research. *BMJ* 2000;**320**:376–7.
27 **Peersman G**, Harden A, Oliver S, *et al.* Discrepancies in findings from effectiveness reviews: the case of health promotion interventions to change cholesterol levels. *Health Educ J* 1999;**58**:192–202.
28 **Rees R**, Harden A, Shepherd J, *et al. Young people and physical activity: a systematic review of research on barriers and facilitators.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, 2001.
29 **Medawar P**. Is the scientific paper a fraud? In: Edge D, ed. *Experiment: a series of scientific case histories.* London: British Broadcasting Corporation, 1964.
30 **Kaptchuk T**. Effect of interpretive bias on research evidence. *BMJ* 2003;**326**:1453–5.