

## Perspective

# Systematic reviews – do they ‘work’ in informing decision-making around health inequalities?

MARK PETTICREW\*

*Public and Environmental Health Research Unit, London School of Hygiene and Tropical Medicine, London, UK*

HELEN ROBERTS

*Social Science Research Unit, Institute of Education, University of London, UK*

## Introduction

The question of ‘what works’ is a fundamental one not only for politicians and policy makers who need to devise or implement policies on everything from reducing juvenile crime to increasing the national wealth, but it is also fundamental for citizens on the receiving end of interventions. The observation that some things work better than others (and other things work not at all) is commonplace. So is scepticism among the public and professionals about grand claims for the effectiveness of policies, particularly given our understanding that modest interventions normally have modest effects. Whilst research can help in informing decisions about what works, conflicting research findings, and simple information overload often simply cloud the issue. Literature reviews may be designed to solve (or at least address) the problems of information management, but these reviews may themselves conflict. Take, for example, literature reviews of the effectiveness of mentoring in young people to reduce anti-social behaviour. The findings of reviews may conflict not just because of differences in inclusion criteria but because authors appraise and synthesize information on the outcomes differently (for example, not differentiating between more and less objective sources of outcome data, which vary in the extent to which they are prone to bias). Moreover, the outcomes themselves – stated satisfaction with the service, higher self-esteem, or a reduction in offending for instance – themselves differ (Roberts *et al.*, 2004).

\*Correspondence to: Mark Petticrew, Public and Environmental Health Research Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. Email: Mark.Petticrew@lshtm.ac.uk

Thanks to Kristin Liabo and Ian Shemilt for suggesting relevant literature. Mark Petticrew was funded by the Chief Scientist Office of the Scottish Executive Department of Health and the ESRC when much of the work described in this paper was completed. The paper was also informed by discussions with colleagues in Calgary in 2006 at the International Collaboration on Complex Interventions funded by the Canadian Institutes of Health Research.

In clinical decision making, such uncertainty about the true effects of interventions (as well as evidence of considerable variations in practice) acted as an important stimulus to the adoption of systematic reviews as a way of marshalling robust scientific evidence and bringing it to bear on clinical decisions. While this approach, led by the Cochrane Collaboration, met with early criticism, particularly on the grounds of ‘cookbook medicine’ (the unthinking and uncritical application of research findings to individual patients) – an accusation rejected by proponents (Sackett *et al.*, 1996) – broad acceptance of the general principles of systematic reviews meant that they were quickly and widely adopted to synthesize evidence on the effects of health technologies, and their extension to policies and practices informed by disciplines other than medicine quickly followed.

The *systematic review* is a method of critically appraising, summarizing, and attempting to reconcile the evidence on a particular problem (Jadad *et al.*, 1997; Petticrew, 2000; Petticrew and Roberts, 2006). One of the values of a systematic review is that it can provide a synthesis of studies in a particular field of work, which no policy maker or practitioner, however diligent, could hope to read themselves. None of this is remotely new of course and it has been good scientific practice for at least a century to start a new study by summarizing previous research (e.g., Nichols, 1891).

Systematic reviews offer a challenge to the role of the expert, and to much received wisdom. Much current theory and practice are based on assumptions about what works, about what is appropriate, and on past practices. Over time, these become crystallized and resistant to challenge. New practitioners may question this received wisdom, and new researchers in a field may doubt the theory, or the strength of evidence offered by certain studies, but often no direct challenge is possible. Moreover, much social, and other, research is conducted within ‘schools’ – schools of thought, which direct the type of science conducted, and to some extent can control the outcomes – by controlling how, when (and if) studies with particular approaches or reporting particular findings are published. ‘Invisible colleges’ prefer orthodoxy; researchers work within a paradigm (or, more likely, several paradigms) in which there is unwritten consensus about what is to be researched. Again, this is not new (‘Every learner hath a deference more or less to authority, especially the young learners ... and things early admitted by repetition become familiar. And this familiarity at length passeth for evidence’: From ‘A defence of free-thinking in mathematics’ by Bishop George Berkley, 1685–1753). Systematic reviews, however, allow challenge to the paradigm – a challenge permitted by close examination of the underpinning evidence.

Systematic review methods, originally developed by social scientists, have percolated widely and can be seen in such diverse fields as conservation, software engineering, and religion (see for example: <http://www.cebc.bangor.ac.uk/index.htm>; <http://portal.acm.org/citation.cfm?id=1134500>; and <http://www.informaworld.com/smpp/content~content=a713626400~db=all~order=page>).

Systematic reviews in public health have also become commonplace and it is widely accepted that they are necessary to support decision making; systematic review methods underpin the approach taken by NICE for example in producing its public health guidance (see <http://www.nice.org.uk/page.aspx?o=295452>; also Birch and Gafni, 2007).

***Public health evidence: rare, dispersed, and different***

There are, however, considerable technical and conceptual challenges facing public health systematic reviewers. Some of these are well known, and relate to the *under-populated*, *dispersed*, and *different* nature of the public health evidence base (Millward *et al.*, 2003). It is *under-populated* because there are few outcome evaluations of public health interventions and fewer still that examine the distributive effects of interventions across different social groups – and can thus shed light on the effective means of reducing health inequalities. Moreover, the evidence that exists is much more widely *dispersed* than evidence of the effectiveness of clinical interventions. Setting aside for a moment the effects of publication and related biases, reports of evaluations of the health outcomes of clinical health care interventions can be (mostly) located relatively easily through searches of biomedical databases. Perversely, wider, more exhaustive searches may even in some cases introduce bias, by drawing in weaker evidence (Egger *et al.*, 2003). Search strategies with known sensitivity and specificity to identify the most robust evidence have been developed (and published) to allow studies of the relevant trials to be located (Glanville *et al.*, 2006). In the case of healthcare interventions, the interventions themselves can generally be easily specified to allow them to be used as search terms.

This is not often the case with public health systematic reviews. If we accept that the main determinants of public health and health inequalities are social, then the interventions of most interest to public health reviewers are social interventions: social projects, programmes, and policies. Everything outside the health sector (as well as some things within it) are potential public health interventions. Interventions within the housing, transport, environment, employment, justice and many other sectors are thus potentially eligible for inclusion in such reviews. Evaluations of the outcomes of such interventions are, however, uncommon, and studies of their health effects even more so. This is not surprising, as (with only a very few exceptions, mainly relating to disability) housing is not provided, nor transport policies implemented for health reasons. Assessments of the distributive effects of such policies are even rarer.

Even locating relevant evidence by means of electronic searches is challenging. For example, studies of transport policies may report on the health or social outcomes of such interventions, but this may not be reflected in the titles and abstracts of journal articles. Relevant studies may thus be easily missed. In some databases, older studies may lack abstracts entirely and this further increases the risk of missing relevant evidence. It may even introduce an

additional source of bias. For this reason, handsearching of journals and books is perhaps even more important when conducting systematic reviews in public health than in other fields.

More importantly the public health evidence base to be synthesized is *different*, because although there is an undoubted absence of very many randomized controlled trials of complex social interventions such as housing improvement, where there are very few such studies (Thomson *et al.*, 2001) or new roadbuilding, where there are – and probably will always be – none (Egan *et al.*, 2003), it is nonetheless untrue to suggest that there is ‘no’ evidence at all. Instead, there is a very great deal of non-experimental evidence, some of which explores relationships between risk factors (such as aspects of the physical or social environment) and health or other wellbeing-related outcomes (such as levels of physical activity; or perceptions of safety). There are also studies which exploit natural variations in the provision of some service – or which use, for example, time-series data to assess the effects of a new policy. In the latter case, there falls many evaluations of population tobacco control policies which restrict smoking in public places. In another early example, the effects of the 1974 fuel crisis on mortality rates were examined by comparing data from the first quarter of 1974 with the first quarters of 1970–1973. As the fuel crisis resulted in reduced exhaust emissions and reductions in highway speed limits, all-cause mortality as well as rates of death from cardiovascular diseases and chronic lung disease fell (Brown *et al.*, 1974) – a drop which the authors felt might be attributed to the fuel crisis.

The public health evidence base also consists of econometric studies that are used for example to model the effects of taxation on health behaviours, but such studies rarely make it into systematic reviews. This is partly because they do not fit easily into existing hierarchies of evidence, but also because their statistical methods tend to lack transparency for reviewers. The critical appraisal of study quality that is such an important feature of the systematic review process is therefore difficult to apply to such studies and so a potentially valuable part of the public health evidence base tends to be overlooked.

The contribution of observational studies similarly tends to be downplayed even though much public health evidence is observational in nature, and reviews which synthesize a range of data – for example experimental and observational data – are still relatively uncommon. The exception is the integration of quantitative and qualitative research where there is rapidly growing interest in developing methods of synthesizing qualitative and quantitative evidence (Lucas *et al.*, 2007), including Bayesian and interpretive approaches (see, for example, the recent book by Pope *et al.*, 2007).

Perhaps one of the main problems facing those conducting systematic reviews with the aim of informing policies for reducing inequalities is that reducing inequalities is so rarely the intended purpose of most intervention studies. Evaluations aimed at improving the health of the poor tend not to seek specifically

to reduce inequalities, or to measure the reduction in the gap or the gradient as outcomes; it has been taken as read that interventions targeted at the poorest areas or communities or individuals will naturally support the wider policy goal of reducing inequality. Of course, it is now widely known that this is not the case, and interventions may well improve the health of the poor but at the same time contribute to increasing inequalities (Macintyre, 2003).

### *Intervention-selection bias in public health reviews*

What is less widely appreciated is that by including only intervention studies and experimental studies into systematic reviews, one may inadvertently introduce an ‘intervention-selection’ bias. For example, by including only RCTs, one is indirectly selecting mainly individual-level interventions, which are more easily randomized. One then risks overlooking possible upstream solutions – among them the upstream determinants of inequalities. This is illustrated by a recent review of interventions to promote walking, in which the interventions ranged from the individual level – such as brief advice and pedometers – up to the community level – such as environmental improvements – and transport – such as car-sharing clubs (Ogilvie *et al.*, 2007). The authors suggested that the available evidence (particularly from the most robust study designs) was skewed in favour of studies of interventions that seem easier to evaluate, or perhaps easier to randomize – typically the individually focused interventions, such as brief advice and pedometers.

A similar finding emerged from a different systematic review of interventions to promote modal shift in travel (that is, from cars to walking and cycling). Here, only three RCTs were found, and these studies were the only ones reporting robust data on health outcomes. However, including only these studies would have meant only including two small categories of intervention: targeted behaviour change programmes for commuters, and school travel co-ordinators. The review would not have identified evidence about, or perhaps even the existence of, any population-wide health promotion activities, “environmental” engineering or transport service developments, or financial incentives, and would not have identified any of the studies that indicated possible unexpected or inequitable effects of interventions. This finding supported the contention that there is an “inverse evidence law” whereby the least is known about the effects of interventions most likely to influence whole populations, and should be borne in mind when reviewing evidence on the social determinants of health inequalities (Ogilvie *et al.*, 2005). Narrowly defined exclusion criteria in public health reviews may mean that systematic reviews ‘work’ less well than they could.

The above discussion has concentrated on methodological issues when assessing how systematic reviews ‘work’, and on some of the things that may interfere with their workings. At this stage, our interim answer is that public health systematic reviews can be made to work. They require flexibility and thought (like any piece of research), but there is no inherent reason why they cannot

be made to answer useful and important public health questions, and in some cases challenge conventional wisdom. A systematic review by DiGiuseppi *et al.* (2001) of the effectiveness of interventions to promote residential smoke alarms offers one example, where it was found that programmes to promote smoke alarms have only modest effects on smoke alarm ownership and function, and no demonstrated beneficial effects on either fires or fire-related injuries. Counselling as part of child health surveillance was found to be more effective in terms of its effect on smoke alarm ownership and function (though its effects on injuries were found to be evaluated). The mentoring study referred to below cast doubt on claims of all-round effectiveness, and perhaps most notably, the ‘scared straight’ systematic review suggests that visits to prisons with tough regimes by young offenders may have the reverse of the desired effect (Petrosino *et al.*, 2002).

Systematic reviews, then, make a claim to tell us ‘what works’. They are also often described as scientific hypothesis-testing tools. This next section considers in more detail whether reviews really ‘work’ in these two respects.

#### *Can systematic reviews tell us ‘what works’?*

To answer this, we need first to think about how we know whether something works, or not. In the field of health, the question ‘what works?’ is normally used as a shorthand for asking whether an intervention is effective with respect to some outcome (or set of outcomes). If we are interested in tackling health inequalities through the social determinants of health, then we are most interested in changes in the distribution of health status resulting from the implementation of social policies. However, assessing whether those policies work in health terms or in terms of health inequalities is difficult and often impossible, because, as indicated above, improving health is not the core business of most social interventions, and health outcomes (and the distribution of health outcomes) are rarely assessed. If they are, they are not often quantified. By extension, systematic reviews of interventions will for the indefinite future find it particularly difficult to wrestle with questions of inequalities unless they synthesize evidence on the intermediate determinants of health inequalities – such as employment, education, and other intermediate outcomes – not just on health inequalities themselves. Focussing solely on the latter may risk producing reviews that reiterate that we don’t know much yet about the means of tackling health inequalities, and therefore risk undervaluing those interventions that aim to produce a different and highly complex set of social outcomes.

Byford and Sefton illustrate this by comparing the outcomes of health care interventions – often localized, physical, and amenable to quantitative measurement – with the outcomes of social care interventions – generally multiple, subjective, and ‘holistic’ (Byford and Sefton, 2003). For many social interventions, this complexity is multiplied as the outcomes that are valued also differ widely among stakeholders. In the case of housing regeneration programmes, public

health researchers are most interested in specific health outcomes, often respiratory health and mental health, and injuries among tenants; housing providers by contrast may be most interested in stock protection, turnover and occupancy rates, and amenity (e.g., warmth). Tenants themselves, however, may value space and privacy. Identifying a single primary outcome on which to power a study to evaluate such complex, interrelated outcomes is difficult, and possibly meaningless. Synthesizing evidence across such evaluative studies needs to take account of (i.e., include, and describe) the range of positive and negative effects associated with the ‘intervention’ (itself highly complex, and varying in its implementation). Reviewers often deal with complex issues and interventions by focussing down; they tend to be ‘splitters’ rather than ‘lumpers’. The latter tend to lump concepts into broad categories, including, for instance, variation in the types of interventions, outcomes, or study designs. However, the more one splits and the more narrowly the review becomes focussed, the less value and meaning the findings are likely to have for anticipated users.

One particular criticism of reviews from the point of view of practitioners is their readability/digestibility. A method to improve these aspects of reviews, used in the ESRC-funded *What Works for Children?* project ([www.whatworks-forchildren.org.uk](http://www.whatworks-forchildren.org.uk)), designed to bring research evidence and social care practitioners closer together, was the development of ‘evidence nuggets’. These summaries of research were based on findings from systematic reviews and focussed more on key messages from the evidence (see Box 1) and implementation (see Box 2) than on the methodology of included studies. They were peer-reviewed, and referenced (although the references have been omitted from the boxes below) so that practitioners who wanted to could check the evidence could. Where it was possible to find literature on the resource implications of an effective service using the evidence, it was included. Feedback from practitioners suggested that accessibility was more highly valued than detail (though not by everyone). As Brocklehurst and Liabo (2004) suggest, such products, written in plain English, have the potential to promote a more democratic partnership between researchers, professionals, and possibly even users of services, in jointly planning care.

*Can systematic reviews tell us ‘what works’? The hypothesis-testing function of systematic reviews*

For scientists, a significant selling point for systematic reviews is the possibility that they can be used for hypothesis testing. Early influential papers on systematic reviews in healthcare emphasized this point (e.g., see Petticrew and Roberts, 2006). However, when one applies systematic review methods to complex interventions the relative absence of experimental studies, and the complex and holistic nature of the outcomes means that ‘hard’ tests of hypotheses about the effectiveness of social interventions are probably few and far between.

**Box 1****Extract from 'Evidence Nugget' on Traffic Calming****Key messages**

- Child pedestrian injury arising from road accidents is a leading cause of accidental death.
- Children in poor neighbourhoods are five times more likely to be injured by a car than those in affluent areas.
- Area-wide traffic calming is designed to control traffic in urban residential areas and has been shown to be effective in reducing child accidents.
- Introducing an area-wide traffic-calming scheme is likely to be an effective measure in reducing inequalities in child health.

(Liabo, 2004)

**Box 2****Extract from 'Evidence Nugget' on the implementation of cognitive behaviour therapy in managing behavioural problems and conduct disorder in pre-adolescence****What are the policy and practice implications?**

In the UK, it is common policy that children with emotional or behavioural difficulties should be retained within mainstream schools with behaviour management plans in place wherever possible. This makes schools potential contact and treatment points for children. Consider your target population when deciding on the most appropriate intervention. Children with conduct disorder often have lower than average verbal intelligence with a short attention span, and it may be appropriate to tailor the CBT to include less discussion and be more action-orientated. The treatment should take into account the diagnosis and age of the child. In many cases, and for children with more than one diagnosis, multiple treatments are needed, requiring cross-agency collaboration across health, social services, education, juvenile justice, and voluntary sector agencies.

Practitioners with knowledge of CBT theory and practice will certainly need to be included in the development of a programme. The UK Council for Psychotherapy offers information on seeking an accredited psychotherapist or on training relevant staff, specific to your chosen intervention. Contact details for this and other organizations can be found in the 'Contacts' section at the end of this nugget.

The application of CBT requires knowledge of social learning principles and a variety of different skills. These skills can be readily taught but this does take time. There is evidence that proper training in the psychological therapies enhances clinical efficacy.

### **What are the resource implications?**

The costs of CBT will vary depending on the programme chosen, and whether it is integrated into existing services or targeted at a particular high-risk group. Initial assessment is important and may be costly when using a CBT approach because implementation will be tailored to the needs of each individual child.

A recent (2004) estimate of the cost of employing a Clinical Psychologist, based on a mid-point salary and including on-costs and overheads, is £30 per hour overall and £69 per client contact hour. Other professionals or non-professionals may be able to deliver this intervention, but appropriate training will be an important and significant budgeting consideration. The research quoted here mainly used graduate students.

Analysis of CBT in other contexts have shown it to be a cost-effective intervention, in particular in relation to youth offending. Our literature searches did not find any cost-analyses of CBT for disruptive behaviour in children.  
[http://www.whatworksforchildren.org.uk/nugget\\_summaries.htm#cbt](http://www.whatworksforchildren.org.uk/nugget_summaries.htm#cbt)

### *The multiple functions of systematic reviews*

In any case, the assessment of ‘what works’ is only one of the many tasks that systematic reviews can fulfill and we often undersell the others. Pope *et al.* (2007) differentiate between the knowledge-support and decision-support objectives of reviews, in which ‘knowledge-support’ reviews confine themselves to synthesizing research evidence, whereas ‘decision-support’ reviews aim to include further analytical tasks required to reach a specific decision in a specific context, and may include non-research evidence to help them to do so. Widening the objectives of systematic reviews beyond scientific hypothesis testing is important as it allows the greater use of a wider range of evidence and maximizes the value that may be gained from the enormous time and resources generally consumed by systematic review(er)s. Thus, while knowing whether something works in tackling inequalities is vitally important, it is also important to use the review for other legitimate and important tasks, including stocktaking and reality checking.

### *Stocktaking systematic reviews*

Systematic reviews have a simple stocktaking function – sometimes referred to as ‘mapping’ the evidence base. Simply knowing with some degree of accuracy

what has been done previously, and where it can be located, is important, if only to identify (and possibly shift) researchers' current priorities. This has been done with some regularity in public health in recent years, though stock-taking, such as painting the Forth Bridge, is never-ending and regular updates are needed. The Centre for Reviews and Dissemination (CRD) at the University of York for example published several versions of its Wider Public Health report (<http://www.york.ac.uk/inst/crd/wph.htm>), identifying reviews of relevance to the wider public health agenda. Individual systematic reviews relevant to a public health focus have also been carried out as part of the work of the Cochrane Collaboration, and the Campbell Collaboration, particularly as part of the work of the Social Welfare Group (<http://www.campbellcollaboration.org/SWCG/index.asp>). The EPPI-Centre <http://eppi.ioe.ac.uk/cms/>, the ESRC Centre for Evidence Based Public Health Policy (<http://www.msoc-mrc.gla.ac.uk/Evidence/Evidence.html>), and the Public Health Research Consortium (<http://www.york.ac.uk/phrc/>), and a new Cochrane Public Health Review Group (<http://ph.cochrane.org/en/index.html>) are among many others doing this work.

### *Mining*

Systematic reviews are intended to act as drivers of future research by identifying gaps where future studies are needed. However, they are also important sources of methodological and other information to be used when planning new studies. Our own systematic reviews on the health effects of housing improvement and new road building directly informed the development of new primary studies assessing the outcomes of the provision of new social housing and the building of a new motorway, respectively (Ogilvie *et al.*, 2006). The specific contribution of these reviews – aside from identifying the partial state of the evidence in these fields, and thus indicating the need for new research – was to help specify and refine the research questions, to identify the primary and secondary outcomes, and to estimate the possible range and size of positive and negative effects. They also allowed us to identify the measurement tools used to assess outcomes in previous studies (in the case of intervention studies of housing and health studies, commonly the SF-36 and the GHQ) to ensure that any new data we collected were collected in a manner consistent with previous research and to thus permit some degree of cumulation of evidence.

'Theoretical mining' of existing research is also possible, in which existing research is reviewed in order to exploit theories of health behaviour, which in turn can help in the development of new interventions (Munro *et al.*, 2007).

This sort of methodological mining of previous research, however, requires close coupling between primary and secondary research. Ultimately, the relationship between the reviews and primary studies is a somewhat circular one; the next update of the housing systematic review mentioned above will include some of our own new primary research.

Finally, the reality-checking function of systematic reviews is also important as they, like other forms of research, provide a parallel commentary on current policy or practice – often implicitly or explicitly challenging the evidence base underlying specific decisions.

### *Economics and systematic reviews*

What has been much less common in public health systematic reviews is the incorporation of an economic perspective. There may be several reasons for this. The simplest is that there are so few economic evaluations of social interventions that reviewers have not found many relevant economic evaluations to include. One other reason is that there has been relatively little guidance from major systematic review centres as to ‘how to do it’ in practice. Outside of the field of health care, systematic reviews have often therefore confined themselves to answering questions about effectiveness alone, and (much less often) questions about aetiology. This is changing, however, and the Campbell and Cochrane Economics Methods Group (CCEMG) has recently produced a significant document outlining the rationale for including economic issues within Campbell and Cochrane reviews, and guidance on the methods which reviewers may adopt to do so (Shemilt *et al.*, in press). For non-economists the absence of consensus about what constitutes ‘good quality’ among economic studies will prove a challenge (though the same can equally be said about incorporating qualitative or observational research in systematic reviews). Guidance on the critical appraisal of economic evaluations is not new, and appraisal tools already exist (Byford *et al.*, in press; Chiou *et al.*, 2003; Centre for Reviews and Dissemination, 2001; Drummond *et al.*, 1996). Closer working with economists on review teams may help foster the production of more meaningful and generalizable systematic reviews. A necessary precursor to this of course is the involvement of economists in the primary research – that is, in outcome evaluations of social interventions, and this is still all too uncommon.

One other important development is the establishment of a Cochrane and Campbell Health Equity Group to take forward systematic reviews of the effects on equity of social and health care policies (see website at <http://www.equity.cochrane.org/en/index.html>). Most Cochrane and other systematic reviews currently only produce an estimate of the ‘overall’ effects of the intervention of interest – the Campbell and Cochrane Equity Field aims to work with the 51 Cochrane Review Groups to expand their methods to include a description of equity components that are currently missing into all their reviews. The Equity Field will also encourage the production of Campbell and Cochrane reviews on interventions that are primarily focused on reducing socioeconomic inequalities in health and/or improving the health of the disadvantaged.

It is also developing methods. Several reviews have tested a method of collecting data on health inequality using an acronym called ‘PROGRESS’ for measuring disadvantage, developed by Hilary Brown and Tim Evans (WHO).

PROGRESS stands for Place of residence, Race/ethnicity, Occupation/unemployed, Gender, Religion, Education, Socioeconomic Status (income or composite measures) and social capital, and can be used as a guide to identifying relevant data in primary studies (Tugwell *et al.*, 2006). One of the early reviews from the group assesses the effects on health and health inequalities of school feeding in developed and developing countries (Kristjansson *et al.*, 2007). This review found that school meals have some small benefits (on weight and height, and educational outcomes) for disadvantaged children. The review wrestled with all the methodological challenges outlined earlier in this paper, including the need to incorporate a range of study designs, and multiple outcomes. Data from randomized controlled trials, non-randomized controlled clinical trials, controlled before and after studies, and interrupted time series studies were all included, and data on contextual and other factors which may affect the success of school feeding were synthesized (Greenhalgh *et al.*, 2007). It has been suggested that such information – that is, going beyond simple measures of effectiveness – is essential for policy makers and practitioners if they are to be able to use systematic reviews to implement public health programmes. Policies that work in one context – for example, in one cultural setting, or at one point in time, or in one particular country – may work very differently or not at all in another.

## Conclusions

So do systematic reviews ‘work’? We of course believe they do, or at least that they can, but accept that they are a tool, not an answer. They need to be embedded in an intelligent assessment of context, and with a close eye to problems of implementation. For example, while home visiting schemes have shown considerable promise, a group that has worked extensively in this area cautions against its use in circumstances of severe domestic violence (Eckenrode *et al.*, 2000). And in trying to understand how best to use reviews in reducing health inequalities, we also need to know a good deal more about important sub-populations – the needs of people with learning disabilities, the needs of people living in rural communities, and the needs of young fathers for instance. Even where reviews do provide an answer, or a partial answer, they do not substitute for social value judgments, such as whether we prioritize some populations, or age groups, or ethnicities – though an understanding the effectiveness of interventions within such groups might help to inform social value judgments. This is a topic which has been discussed by the NICE Citizens Council (NICE, 2007), who were asked which of two broad strategies would be more appropriate for NICE to follow: whether it is more appropriate for NICE to issue guidance that concentrates resources on improving the health of the whole population (which may mean improvement for all groups) even if there is a risk of widening the gap between the socio-economic groups; or whether it is

appropriate for NICE to issue guidance that concentrates resources on trying to improve the health of the most disadvantaged members of our society, thus narrowing the gap between the least and most disadvantaged, even if this has only a modest impact on the health of the population as a whole. This in turn is one of the factors that will inform the development of a new version of a paper on social value judgements, shortly to go out for consultation.

The epistemological/philosophical and methodological challenges in this field are considerable, and there is still some cynicism about the ability of systematic reviews to deliver meaningful messages (other than simply underlining the need for more research); but such reviews, performed thoughtfully, and with the input of users, help redress the undue emphasis still placed on single studies, and form a rational, and, to users, acceptable basis for evaluating complex social interventions, and for supporting public health decision making (LaPelle *et al.*, 2006; Lavis *et al.*, 2005, Dobbins *et al.*, 2007).

## References

- Birch, S. and A. Gafni (2007), 'Economists' dream or nightmare? Maximizing health gains from available resources using the NICE guidelines', *Health Economics, Policy and Law*, 2: 193–202.
- Brocklehurst, N. and K. Liabo (2004), 'Evidence nuggets: promoting evidence based practice', *Community Practitioner*, 77: 292–296.
- Brown, S.M., M.G. Marmot, S.T. Sacks, and L.W. Kwok (1975), 'Effect on mortality of the 1974 fuel crisis', *Nature*, 257(5524): 306–330.
- Byford, S., M. Drummond, E. Eisenstein, M. Knapp, J. Mallender, D. McDaid, M. Mugford, I. Shemilt, L. Vale, and D. Walker (in press), 'Incorporating economics evidence', in J. Higgins and S. Green (eds), *Cochrane Handbook for Systematic Reviews of Interventions*, Chichester: John Wiley & Sons.
- Centre for Reviews and Dissemination (2001), 'Improving access to cost-effectiveness information for health care decision making: the NHS Economic Evaluation Database', CRD, University of York.
- Chiou, C., J.W. Hay, J.F. Wallace, B.S. Bloom, P.J. Neumann, S.D. Sullivan, H. Yu, E.B. Keeler, J.M. Henning, and J.J. Ofman (2003), 'Development and validation of a grading system for the quality of cost-effectiveness studies', *Medical Care*, 41: 32–44.
- DiGiuseppi, C., C. Goss, and J.P.T. Higgins (2001), 'Interventions for promoting smoke alarm ownership and function', *Cochrane Database of Systematic Reviews* 2001, Issue 2, Art. No.: CD002246.
- Dobbins, M., S. Jack, H. Thomas, and A. Kothari (2007), 'Public health decision-makers' informational needs and preferences for receiving research evidence', *Evidence Based Nursing*, 4(3): 156–163.
- Egan, M., M. Petticrew, D. Ogilvie, and V. Hamilton (2003), 'New roads and human health: a systematic review', *American Journal of Public Health*, 93(9): 1463–1471.
- Egger, M., P. Juni, C. Bartlett, F. Holenstein, and J. Sterne (2003), 'How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study', *Health Technology Assessment*, 7(1): 1–76.

- Glanville, J.M., C. Lefebvre, J.N. Miles, and J. Camosso-Stefinovic (2006), 'How to identify randomized controlled trials in MEDLINE: ten years on', *Journal of the Medical Library Association*, 94(2): 130–136.
- Greenhalgh, T., E. Kristjansson, and V. Robinson (2007), 'Realist review to understand the efficacy of school feeding programmes', *British Medical Journal*, 335: 858–861.
- Kristjansson, E.A., V. Robinson, M. Petticrew, B. MacDonald, J. Krasevec, L. Janzen, T. Greenhalgh, G. Wells, J. MacGowan, A. Farmer, B.J. Shea, A. Mayhew, and P. Tugwell (2007), 'School feeding for improving the physical and psychosocial health of disadvantaged elementary school children', *Cochrane Database of Systematic Reviews*, Issue 1.
- LaPelle, N.R., R. Luckmann, E. Hatheway Simpson, and E.R. Martin (2006), 'Identifying strategies to improve access to credible and relevant information for public health professionals: a qualitative study', *BMC Public Health*, 6: 89.
- Lavis, J., H. Davies, A. Oxman, J.L. Denis, K. Golden-Biddle, and E. Ferlie (2005), 'Towards systematic reviews that inform health care management and policy-making', *Journal of Health Services Research and Policy*, 10(1): 35–48.
- Liabo, K. (2004), 'Traffic calming and childhood injury on the road', National Children's Bureau, London, 2 pages.
- Lucas, P.J., J. Baird, L. Arai, C. Law, and H.M. Roberts (2007), 'Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews', *BMC Medical Research Methodology*, 7: 4 (available at: <http://www.biomedcentral.com/1471-2288/7/4>).
- Macintyre, S. (2003), 'Evidence based policy making', *British Medical Journal*, 4: 326(7379): 5–6.
- McCarthy, M. (1999), 'Transport and health', in M. Marmot and R. Wilkinson (eds), *The Social Determinants of Health*, Oxford: Oxford University Press.
- Millward, L., M.P. Kelly, and D. Nutbeam (2003), *Public Health Interventions Research: The Evidence*, London: Health Development Agency.
- Munro, S., S. Lewin, T. Swart, and J. Volmink (2007), 'A review of health behaviour theories: how useful are these for developing interventions to promote long-term medication adherence for TB and HIV/AIDS?', *BMC Public Health*, 7: 104.
- Nichols, H. (1891), 'The psychology of time', *American Journal of Psychology*, 3: 453–529.
- Ogilvie, D., M. Egan, V. Hamilton, and M. Petticrew (2005), 'Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go?', *Journal of Epidemiology and Community Health*, 59: 886–892.
- Ogilvie, D., R. Mitchell, N. Mutrie, M. Petticrew, and S. Platt (2006), 'Evaluating health effects of transport interventions methodologic case study', *American Journal of Preventive Medicine*, 31(2): 118–126.
- Ogilvie, D., C.E. Foster, H. Rothnie, N. Cavill, V. Hamilton, C.F. Fitzsimons, and N. Mutrie (2007), *British Medical Journal*, 334: 1204.
- Petrosino, A., C. Turpin-Petrosino, and J. Buehler (2002) '“Scared Straight” and other juvenile awareness programs for preventing juvenile delinquency', *Cochrane Database of Systematic Reviews*, Issue 2. Art. No. CD002796.
- Petticrew, M. and H. Roberts (2006), *Systematic Reviews in the Social Sciences*, Oxford: Blackwell.

- Pope, C., N. Mays, and J. Popay (2007), *Synthesising Qualitative and Quantitative Evidence: A Guide to Methods*, Maidenhead: Oxford University Press/McGraw Hill.
- Roberts, H., K. Liabo, P. Lucas, D. DuBois, and T.A. Sheldon (2004), 'Mentoring to reduce antisocial behaviour in childhood', *British Medical Journal*, **328**: 512–514.
- Sackett, D.L., W.M.C. Rosenberg, J.A. Muir Gray, R.B. Haynes, and W.S. Richardson (1996), 'Evidence based medicine: what it is and what it isn't', *British Medical Journal*, **312**: 71–72.
- Shemilt, I., M. Mugford, S. Byford, M. Drummond, E. Eisenstein, M. Knapp, J. Mallender, K. Marsh, M. McDaid, L. Vale, and D. Walker (in press), 'Campbell Collaboration Economics Methods Policy Brief', C2 Methods Groups (Online).
- Thomson, H., M. Petticrew, and D. Morrison (2001), 'Health effects of housing improvement: systematic review of intervention studies', *British Medical Journal*, **323**(7306): 187–190.
- Tugwell, P., M. Petticrew, V. Robinson, E. Kristjanansson, and L. Maxwell (2006), 'Cochrane and Campbell Collaborations, and health equity', *Lancet*, **367**: 1128–1130.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.