Meeting the challenge: Developing systematic reviewing in social policy

ARTICLE in POLICY & POLITICS · OCTOBER 2004
Impact Factor: 0.7 · DOI: 10.1332/0305573042009444 · Source: DAI

4 AUTHORS, INCLUDING:

Alison Wallace
The University of York
15 PUBLICATIONS 66 CITATIONS

Karen Croucher
The University of York
17 PUBLICATIONS 137 CITATIONS

Deborah Quilgars
The University of York
36 PUBLICATIONS 211 CITATIONS

Available from: Karen Croucher
Retrieved on: 15 September 2015
Meeting the Challenge: Developing Systematic Reviewing in Social Policy.

By Alison Wallace¹, Karen Croucher², Deborah Quilgars² and Sally Baldwin¹

Paper presented to:
Social Policy Association Conference
University of Teesside
16 July 2003 9am

Further information from Alison Wallace aw152@york.ac.uk

¹ Dept of Social Policy and Social Work
² Centre for Housing Policy
Meeting the challenge: Developing systematic reviewing in social policy.

Wallace A, Croucher K, Quilgars D, and Baldwin S

Abstract
This paper uses the experience of reviewing the evidence on the financial support available for (potentially) defaulting homeowners to consider the opportunities and challenges systematic review methods present to social policy. It addresses concerns regarding the examination of the strength of given evidence using quality appraisal and ‘hierarchies of evidence’, and what are perceived to be purely technical methods to pass judgement on large bodies of literature. It argues that there is merit in utilising the method to provide policymakers and practitioners, and indeed the research community, with transparent summaries of the most robust evidence with minimum bias. The paper outlines the challenges presented in synthesising evidence, the determination of ‘good evidence’ and its interpretation, to answer questions relating to highly contextualised complex interventions. Systematic reviews have a role to play in the accumulation of sound knowledge, and social policy researchers have a valuable contribution to make to the developing methods.

Introduction
‘If research per se is considered a useful endeavour, then bringing together the results of individual pieces of research addressing similar questions should be seen as even more useful and largely unproblematic.’ Gough and Elbourne (2002:234 original emphasis). On the contrary, systematically reviewing existing research literature, where authors are explicit about how original studies were identified and selected for inclusion, elicits strong reactions in many academics, often based on concerns and misconceptions about the method. Petticrew (2001) identifies common perceptions of systematic reviews being characterised by the use of statistics, randomised controlled trials, and concerned only with the bio-medical view of health. Whilst emphasising the use of systematic review methods in diverse fields, the inclusion of a variety of research methods and the utility of systematic review methods to wider audiences, he suggests many of the criticisms are fallacious and ill founded. This paper suggests these misgivings are based on misrepresentations of a methodology that is rapidly evolving.

The aims of this paper are to discuss some of the concerns held by some in the research community regarding the methods of systematic reviewing, from a position of having empirically tested the methodology. This paper speaks to the criticisms levelled at the methodology, and at the same time presents challenges to the development of systematic review methods in social policy as we see it. Some of the misgivings regarding the review methods are outlined initially, as it has been suggested that systematic reviewing adheres too closely with one research paradigm, is mechanical
and excludes too much of the evidence available. The second section addresses our experience of reviewing the safety nets for mortgagees’ literature and answers some of the criticisms. Next, the merits of reviewing existing research evidence in this way are examined, their ability to critique the evidence base as well as provide useful summaries of concepts and research evidence is considered. The fourth section examines the methodological challenges that are apparent in developing the methodology in social policy. The paper concludes that systematic reviewing is a desirable development in social policy, but does require further work to ensure high standards of reviewing are used.

Misgivings and misconceptions

Positivist

One of the main criticisms of systematic reviews is their conformity to the positivist paradigm and all that many perceive that entails. Hammersley (2002:6) argues that the concept of science built into discussions of systematic reviews is largely discredited, positivist and amounts to no more than following an explicit method. Marks (2002) suggests that evidence based policy, is premised on seven myths: rationality, intentionality, objectivity, ‘methodolatory’ (or the adulation of RCTs), trainability and opinion eliminating. Evidence-based policy, from which the interest in systematic reviewing is derived, stands accused of being wrongly premised on rational policy making (Packwood, 2002), and lacks understanding of the best evidence from cognitive, social and organisational psychology about influences on decision making (Marks, 2002:23). Some of these points are explored below.

Clearly, policy makers draw on several streams of knowledge when making decisions about policy, such as personal experience, research evidence, practice wisdom and political beliefs (Newman, 2002). Nutley et al (2002) adapt a typology from Weiss (1998) to suggest that even when research evidence is considered, it may be used in four very different ways: 1) instrumentally, where it feeds directly into decision-making; 2) conceptually, where it may not be used but produces new insights and understandings of situations or events; 3) it may mobilise support and provide legitimacy for an already chosen course of action; or 4) it may have wider influence, where the research is synthesised by practitioners or researchers altering policy paradigms or belief communities in the longer term. A linear or rational relationship does not exist therefore between research and its use in practice, and Packwood (2002) is perhaps correct in asserting that evidence ‘informed’ policy and practice is the best that one can hope for. It is wrong however to infer, from an understanding of the messy and contested nature of policymaking and implementation, that researchers somehow refrain from having any sort of dialogue with policymakers.

Systematic review methods are intended to reduce bias that may unwittingly enter into ‘traditional’ literature reviews, by searching extensively for literature and being explicit about how studies are appraised and selected for inclusion. The explicit reporting of these methods enables readers to assess the integrity and validity of the review (Peersman et al, 2001). This transparency is not apparent in ‘traditional’ reviews, but somehow attracts derision. For Marks (2002), the methods
represent additional filters on the social construction of an evidence base. Extensive searches for published or unpublished research from a variety of stakeholders, and the testing of a priori decisions regarding the focus of the review and the review findings against expert views from the field or service users (Forbes et al, 2001; Arksey and O'Malley, forthcoming; Croucher et al, 2003; Boaz, in progress) aims to reduce bias from processes that mediate research production, discussed by Marks (2002), such as publication bias, funding and politics, and values the tacit and experiential knowledge held by these participants.

In terms of the ‘methodolatory’ idolisation of RCTs, there are many examples of reviews that include other study designs including qualitative research (see Harden et al, 1999; Harden et al, 2001; Godfrey et al, 2000; Campbell et al, 2003). Some of the points about the strength of evidence are made below, but the pursuit of methods that can synthesise quantitative and qualitative methods and different study designs are vital for reviews to provide meaningful accounts of the evidence available. Indeed, a review of the influences on the take-up of childhood immunisations found that several key determinants were missed if quantitative methods alone had been included in the review (Roberts et al, 2002).

The arguments put forward in opposition to systematic reviewing are partial and do not reflect the many developments made outside of clinical medicine, some of which Marks (2002) identifies and are outlined in Forbes and Griffiths (2002), such as meta ethnography and critical realist synthesis. Badger (2000:229) suggests that whilst systematic reviewing has it roots in effectiveness research and quantitative approaches, the parameters of the methodology are being continually expanded, increasing its potential application. Boaz et al (2002) notes the contribution social science can make to the development of the methodology in its openness to theoretical analysis and diverse methodologies. In social policy, the methods are far from being embedded and researchers therefore have the benefit of adapting the methods using current debates about the methodology, instead of repeating outmoded arguments.

Losing the evidence
Much of the anxiety some researchers feel over the systematic review methodology centres around the exclusion of literature from the review, either because it is deemed of poorer quality, or the study designs employed are thought to hold insufficient explanatory power compared to other study designs available (see White, 2001; Marks, 2002). This concern about loss of evidence is framed in three main ways. Apprehension exists about the lack of consistent use of consensually agreed quality appraisal criteria (White, 2001; Hopayian, 2001). Ways of valuing different types of research for different types of questions is an ongoing concern (see Gowman and Coote, 2000). However, often it seems researchers and policymakers feel uncomfortable about the scrutiny to which the systematic review methodology will potentially subject their own work or ideas.

There are lively debates about what constitutes a ‘good study’ on which to base policy and practice (see Black, 1996; Oakley, 2000; Marks and Godfrey, 2000; Gowman and Coote, 2000; Rychetnik,
2002; Stafford, 2002), which arguably, could be rehearsed within social policy to our advantage in order to increase the rigor and persuasiveness of research inquiry. Hierarchies of evidence are used in reviews of *effectiveness* and are the cornerstone of evidence-based medicine, but for different types of review questions, and possibly in different fields, experimental research may not constitute the best evidence. Forbes and Griffiths (2002) note a number of areas where evidence based medicine techniques as widely practiced may be inappropriate. These may include where questions relate to values or beliefs (Mays, Roberts and Popay 2001); where the intervention is multi-faceted or involves multiple transactions (Gomm and Davies 2000); or where the outcomes of interest are measures of subjective experience, such as quality of life.

Valuing the contribution different types of research and study designs can make to the evidence base is important. Although critical of some aspects of systematic reviewing, Marks (2002) suggests ‘horses for courses’, a more pluralist and inclusive approach to evidence to ensure research studies match the field of inquiry. Pawson *et al* (2003) consider the variety of knowledge sources that may inform policy and practice in social care, ranging from organisational knowledge, practitioner knowledge, policy community knowledge, research knowledge and user knowledge, some of which will be a product of experience, ideology, politics, empirical study, inspection data, or theory, or indeed a mixture. Pawson *et al’s* study was a precursor for the Social Care Institute for Excellence conducting reviews and disseminating good practice in the social care domain. Gorard (2002) implores the research community to value the best research tools for a particular line of enquiry, a middle way, and avoid paradigmatic wars. There is great merit in valuing evidence from a range of sources and research traditions, yet when considering the review findings and recommendations it is fundamental that researchers conducting systematic reviews engage with the question of what evidence is the most convincing by virtue of source, study designs or quality. Pawson (2001) argues however that the most important issue regarding strength of evidence is that what is claimed for the research should not be overstated.

‘Losing’ evidence in a review due to a judgement of quality is perhaps an uncomfortable proposition. Oliver and Peersman (2001:p83) make the point that ‘...critical appraisal is not about faultfinding and should not result in destructive nihilism, which concludes nothing can be learned from ‘imperfect’ research.’ It is not about only reporting ideal type research, but about ensuring recommendations, perhaps for policy, practice and public investment, are based upon research in which you can be reasonably confident about the findings. In addition, the explicit criteria mean that all studies are appraised using the same guidelines, acting as another tool to minimise reviewer bias. The critical appraisal of research does not automatically result in exclusion from the review but may also help explain different findings of the studies that have entered the review (NHS CRD, 2001).

How these judgements regarding quality are made is an important concern that continues to exercise many, with numerous schema put forward for use in reviews (see Oakley, 2000; Popay, Rogers and Williams, 1998; Mays and Pope, 2000; Mulrow and Oxman, 1997). It is clear however there is little consensus over what criteria to use for certain types of research studies (Dixon-Woods
et al, 2001), few criteria have been developed for some study designs such as secondary analysis of large-scale datasets or social surveys, and there is little advice to reviewers on their practical application (CRD 2001). Hopayian (2001) compares three reviews of effectiveness in health and found they had discordant findings. Two reviews had used the same quality criteria but had produced different answers and the third review used different criteria, and again came up with different conclusions. Hopayian (2001) argues the clinician or expert should have a greater role in the review process, but perhaps another point is also the need for the development of widely accepted methods of quality appraisal.

When appraising the strength of evidence through study design or quality, Edwards et al (1998) suggest examining the ‘signal’ versus the ‘noise’ of studies, as interesting and valuable insights (the signal) can be gained from some methodologically weaker studies (the noise). This approach is appealing, but offers solutions that rest on greater subjectivity in the review decision-making process. It must be acknowledged that subjectivity is inherent in the interpretation of the inclusion and quality criteria, but can be minimised by explicit methods.

A mechanical pursuit
Implicit in Hamersley’s accusation of blind adherence to a method, is that systematic reviews are unscholarly. Gough and Elbourne (p233) discuss different approaches to the synthesis of qualitative research. They argue that interpretative research requires different forms of synthesis than quantitative research, which has hitherto been characterised by the statistical pooling of the data from several studies, or meta-analysis. They cite several authors who suggest that qualitative meta-analysis should be more than an uninspiring catalogue of research studies, should produce new conceptualisations of original results and provide interpretative constructions rather than generalisations (Schwandt 1998,p410; Schreiber, Crooks and Tern 1997, p314; Noblitt and Hare 1988, p23). There are arguments for using theory in evaluation and research synthesis more frequently, thinking about what the research studies are telling us about the context in which different programme mechanisms produce certain outcomes (Pawson 1997, 2001). There is a case therefore for researchers to be more analytical of any review findings. The use of meta-ethnography (Noblitt and Hare, 1988) as a developing method of synthesising research in reviews is an example of such an approach, developing higher-level concepts from bringing research knowledge together (Campbell et al, 2003; Britten et al, 2003). There are calls for systematic reviews to be given the same status as other research activities by the Research Assessment Exercise (RAE), or its successor, for these reasons (Evans and Benfield 2001, Badger 2001).

Mortgage Safety Nets Review
Researchers at the York node of the ESRC Centre for Evidence Based Policy and Practice ‘Evidence Network’ conducted the mortgage safety nets review (www.evidencenetwork.org.uk). The network consists of research units examining different aspects of producing, synthesising and utilising research evidence for policy and practice. Participation in the network originated from a concern that there was little accumulation of research knowledge over the years, and what did exist
was considered quite selective. The York node is interested in developing the methodologies for systematically reviewing in the field of social policy and social care. Questions that we wish to address include: What would happen if a systematic review were conducted in a social policy field? Is social policy different, and if so in what way? Can the methodology be usefully applied to complex interventions for policy making? Beyond developing an evidence base in a number of topic areas, a key aim of our work is to investigate the methodologies required for the identification, appraisal and synthesis of evidence.

Our review of the literature relating to the effectiveness of mortgage safety nets for homeowners (Croucher et al, 2003) sought to collate and synthesise current evidence relating to interventions intended to support homeowners in paying their mortgage if experiencing financial difficulties, and to assess how effective these interventions are at reducing mortgage arrears and repossessions. Mortgage Payment Protection Insurance (MPPI) and Income Support Mortgage Interest (ISMI), flexible mortgages and other private insurances that provide an income or lump sum payment were the focus of the review. Sustainable home ownership is a key policy objective of the government (DTLR, 2000). The review topic reflected concerns within the research community over the operation of the existing safety nets provided by the state and private financial organisations, and whether they adequately address the needs and circumstances of homeowners at risk of default.

Government policy is based upon a range of publicly and privately provided protection for homeowners. Figure 1 depicts the area that was the focus of the review, situated against a backdrop of various macro, meso and micro level variables. This meant it was quite a complex policy arena in which to interpret and synthesise research from different domains.

**Figure 1: Contextual Map of Sustainable Homeownership**

![Contextual Map of Sustainable Homeownership](image)

**Macro level factors:**
- Economic cycles
- Unemployment rates
- Interest rates
- House Prices

**Meso level factors:**
- Welfare regime
- Private insurance market
- Regulation
- Lending and recovery procedures

**Micro level factors:**
- Individual financial awareness and planning
- Causes of arrears and repossessions

**Sustainable Homeownership**

*Source: Croucher et al (2003)*
A small expert panel advised the review, made up of representatives of lenders, government regulatory bodies, consumer groups, and academics in the substantive areas, an academic experienced in review methods, and a government department. The review sought to answer the question *How effective are public and private safety nets in assisting mortgagors in unforeseen financial difficulties to avoid arrears and repossessions?* A review protocol outlining *a priori* decisions at each stage of the review (the question, the search strategy, the inclusion criteria, the quality criteria, and synthesis framework) was drawn up and discussed with the expert panel before the review commenced. This protocol was influenced, but not constrained, by the work of the NHS Centre for Reviews and Dissemination (NHS CRD) at the University of York, and the Evidence for Policy and Practice Information Centre, (EPPI) based at the University of London. The protocol, which was written for our Expert Panel unfamiliar with the methodology, remains available on the web ([http://www.york.ac.uk/inst/chp/ESRC1881.pdf](http://www.york.ac.uk/inst/chp/ESRC1881.pdf)), although some decisions were subsequently changed and documented in the final report (Croucher *et al*, 2003).

**Inclusive Searching**

The review team acknowledged from the start that competing positions existed regarding the value of the state benefit and private insurance schemes, both to homeowners and as a mainstay of public policy supporting sustainable homeownership. With this in mind, the expert panel was set up aiming to reflect these views. The final decisions relating to the boundaries of the review were, however, the responsibility of the review team. It was felt to be beyond the scope of a single review to address all aspects affecting the support of mortgagors, and a decision was made to focus the review on the main policy tools. Full explanations of these decisions were provided however, and the potential value of advice and information services, or responsible lending and recovery practices, for example, was acknowledged. It may not have been possible to reconcile the wishes of all parties, but it is hoped the decisions we made were fair and justified.

The extensive search strategy also aimed to identify studies from many stakeholders and disciplines and fields. Electronic databases reflecting the social sciences, economics and psychology academic fields were searched. ‘Grey’ literature, published or unpublished material not subject to peer-review, was also comprehensively sought from consumer, mortgage and insurance organisations by using specialist electronic databases, a trade library, the internet and exhaustive contact with individual organisations and prominent experts.

A pluralist line was adopted in that the review team was quite prepared to include any credible empirical research that helped answer the question posed. *Credible*, because ethically, research to inform public policy and public investment should be trustworthy; and *any empirical research* as a range of views and experiences regarding the interventions was important to examine our broad definition of effectiveness. The review focussed on the structure and process of the interventions, in addition to the raw outcomes of reducing arrears and repossessions (Wagner 1989). It was important to capture wider impacts of, and attitudes to, the public and private protection schemes. This wide interpretation of effectiveness offered evidence on what mediated any outcomes identified
and enhanced the potential value to policymakers and future users. However, process evidence to explain outcomes may be important, but the question of ‘what works?’, or ‘Does something work?’ remains crucial. The review questions’ focus on outcomes highlighted the paucity of data on whether ISMI and MPPI worked in reducing arrears and repossessions, with only two studies examining the effects of the main policy tools on reducing arrears and repossessions.

Government policy towards sustainable homeownership involves a range of interventions, most of which are operated by private financial institutions which individuals have to opt in to in advance of any problems arising. It is not a field that lends itself to experimental study being highly contextualised and the interventions outwith the control of public agencies and subject to consumer relationships within a private market. This approach therefore, did not take randomised control trials and meta-analysis to be the only credible evidence and method of reviewing (Packwood, 2002:267) but provided a platform for a variety of research methods to be examined to capture all dimensions of our question. Qualitative studies, secondary analysis of large-scale data sets and mixed methods, combining social surveys with qualitative interviews, dominate the evidence base.

*Losing the evidence?*

There are concerns about the conclusions of reviews being drawn from too little evidence as much of the initially retrieved literature is discarded (White, 2001, Marks, 2002). Table 1 shows how the initial set of references was reduced to the 22 final studies that entered the review. The soft nature of many social policy terms means that many studies were retrieved by the electronic searches that had no bearing on the review. Of the 1832 references found, 766 were related to mortgages or homeownership in some way and were considered generally relevant. Of these however, only 49 addressed how our target safety net interventions were working for mortgagors and therefore met our inclusion criteria. Twenty-two of these were considered sufficient quality to enter the final review.

**Table 1: Number of studies at key stages of review**

<table>
<thead>
<tr>
<th>Whole reference set</th>
<th>Generally relevant</th>
<th>Met Inclusion Criteria</th>
<th>Met quality threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1832</td>
<td>766</td>
<td>49</td>
<td>22</td>
</tr>
</tbody>
</table>

*Source: Croucher et al (2003)*

This breakdown of the initial reference set may cause alarm to some, but the reasons for discarding much of the apparent ‘evidence’ is quite mundane and primarily did not relate to issues of poor quality or weak study designs. Imprecise searching, indexing and abstracting on social sciences databases accounts for the large number of references that were not considered ‘generally relevant’. These may have been about homeownership or mortgages, but addressed concerns such as negative equity or the mortgage securitisation of lending portfolios by large financial institutions. The most common reasons for excluding studies were that they were simply unhelpful in answering
the question. Our inclusion criteria ensured the selected studies addressed mortgagors, the four interventions of interest, included aspects of effectiveness and were empirical research. The search strategy threw up a large number of trade articles from mortgage industry periodicals, and a plethora of discursive pieces based on academic, policy and trade expert opinion and therefore the most common component criteria missing was that the studies were not considered empirical research. The use of this literature informed the background to the review, but it offered insufficient evidence of effectiveness, despite the wide interpretation of effectiveness employed.

Credible evidence
The review did not seek evidence only from ‘ideal type’ studies; nevertheless, determining whether studies that met the inclusion criteria were good enough to enter the review was the most problematic stage of the review project. Existing quality criteria addressing secondary analysis of large scale datasets or social surveys were not evident and an initial set of questions were derived from the study of other criteria and advice available for distinguishing good research (Popay, Rogers & Williams, 1998; Mays and Pope, 2000; Pope and Rogers, 1998; Denscombe, 1998; Gomm et al, 2000; Dale, Arber and Procter, 1998 and Godfrey et al, 2000). However, there was a lack of clarity amongst the reviewers regarding the initial set of questions presented in the protocol, which hampered readers’ understanding of how decisions were made using the answers to the questions. The questions were therefore revised to determine which ones were essential to meet a quality threshold, which were desirable to explain variation or aid interpretation of the study and which could be dispensed with, resulting in the solution in Table 2, which provides the set of questions asked of each study. The quality threshold was set, based upon those criteria that were felt to secure the basic validity of the research, generalisability being more difficult to interpret.

It was practical, for the review team’s resources and for the reader, to use the one set of criteria for all quantitative or qualitative studies, but we are aware of the epistemological objections some may hold to this approach (see discussion in Mays and Pope, 1996). All decisions made by the first reviewer were checked by a second person and disagreements resolved consensually within the wider review team, which was valuable in further reducing or minimising reviewer bias. We are also aware that there is the potential for the collection of studies to have been different had the boundaries of the quality threshold been drawn differently. However, after extensive discussion and amendments, the review team feel in the absence of any broader agreement, these criteria can now be operationalised in a way meaningful to users of this and future reviews, albeit contributing to the ‘cottage industry’ of quality criteria and review methods by numerous reviewing teams (Dixon-Woods, 2003).

It is not be necessary to exclude studies on the grounds of quality as the critical appraisal may be used to explain variation between study findings (CRD, 2001), however it was important to maintain standards of evidence if policy recommendations were to flow from the review given the variety of types of research that was found. The EPPI review method excludes poor intervention studies on
the basis of a set of criteria developed by their team but interestingly this does not cover qualitative studies (EPPI, 2001) although poor qualitative studies were found in Croucher et al (2003).

Table 2: Quality criteria for mortgage safety nets review

<table>
<thead>
<tr>
<th>1</th>
<th>Question</th>
<th>Is the research question clear?</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Theoretical perspective</td>
<td>Is the theoretical or ideological perspective of the author (or funder) explicit, and has this influenced the study design, methods or research findings?</td>
<td>D</td>
</tr>
<tr>
<td>3</td>
<td>Study design</td>
<td>Is the study design appropriate to answer the question?</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>Context</td>
<td>Is the context or setting adequately described?</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sampling (Qualitative)</td>
<td>Is the sample adequate to explore the range of subjects and settings, and has it been drawn from an appropriate population?</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Quantitative) Is the sample size adequate for the analysis used and has it been drawn from an appropriate population?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Data collection</td>
<td>Was the data collection adequately described and rigorously conducted to ensure confidence in the findings?</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>Data analysis</td>
<td>Was there evidence that the data analysis was rigorously conducted to ensure confidence in the findings?</td>
<td>E</td>
</tr>
<tr>
<td>8</td>
<td>Reflexivity</td>
<td>Are the findings substantiated by the data and has consideration been given to any limitations of the methods or data that may have affected the results?</td>
<td>D</td>
</tr>
<tr>
<td>9</td>
<td>Generalisability</td>
<td>Do any claims to generalisability follow logically, theoretically and statistically from the data?</td>
<td>D</td>
</tr>
<tr>
<td>10</td>
<td>Ethics</td>
<td>Have ethical issues been addressed and confidentiality respected?</td>
<td>D*</td>
</tr>
</tbody>
</table>

E = essential, D = desirable, * Ethics may be essential in other sensitive fields


Whilst some studies were clearly not conducted rigorously and their findings undoubtedly less trustworthy according to the criteria set, other studies however, were much harder to discern. For example, often evidence relating to safety nets interventions were embedded within wider studies,
which overall were conducted very well, but were less helpful and weaker when informing our particular question. This was because the relevant part of the sample was very small, or there were problems with the data used in this part of the analysis. Care was taken in making quality decisions explicit, ensuring the studies were not considered poor quality for other purposes. Having said this, it was striking how some oft-cited references did not stand up well to closer examination. Studies that did not meet the quality threshold for the review were from all sources, but were concentrated in the voluntary sector or trade organisations, perhaps reflecting a lack of research resources or different objectives in the commissioning of research, for campaigning or marketing purposes, for example. Moreover, the safety nets review found the level of methods reporting was generally poor, perhaps due to the dominance of the idea that policymakers and other users have little interest in this area (Best, 2003). This does beg the question as to why research should be considered more robust or persuasive, if the processes that generated the findings is not differentiated from other competing sources of knowledge and opinion.

Unscholarly Pursuit?
Implicit in Hammersley’s (2002) criticism of systematic reviewing being reduced to just following a method, is that it is essentially an unscholarly or unanalytical pursuit and that it can be conducted by anyone. From the experience of the mortgage safety nets review, we would contend otherwise. The mortgage safety nets review represents our initial approach to utilising the systematic review methodology for a complex social policy issue, and it is not claimed to be definitive in its approach. Policymakers are adept at identifying reports that lack a nuanced understanding of the field, rendering reviews less useful. Therefore, the involvement of those who have a comprehensive understanding of the broader research field is clearly a prerequisite for conducting any such review. It was important for the review team to include those with substantive content knowledge and experience to ensure critical reflection on the applicability of the review questions, meaningful data extraction, interpretation and synthesis of the research studies.

The review may have benefited from the development of concerns regarding issues such as the ability of private insurance to cover labour market risk or theories of individual’s attitudes to risk further in relation to the empirical evidence found. In this way it may have become ‘greater than the sum of [its] parts’ achieving greater understandings of the review topic (Campbell, 2002).

Merits in systematic reviewing

Undoubtedly there are advantages in reviewing research literature systematically, if only for providing the capability to make sense of the information explosion that ICT developments have offered users and policymakers, who may not have either the time or critical appraisal skills to conduct such a task themselves. Maynard (2003) however, suggests other short run rewards being to: sort the ‘wheat from the chaff’, identify gaps and weaknesses in the evidence base and increase access to credible knowledge; and in the long run, increase the ratio of ‘wheat’, encouraging consistent standards, better access to research, and potentially better policy and practice. Campbell
et al (2002) review papers that do not cite the work of papers published before them, missing opportunities to bring together knowledge and compare findings, which in and of itself may be a justification for using research synthesis. They also suggest research synthesis can be used to confirm, refute or modify relevant bodies of theory, after the synthesis revealed important concepts not evident in the individual studies alone.

Our experience confirmed the short run benefits and enforced the view that there are major advantages to critically appraising literature for review in a uniform manner. In competing arenas it affords an opportunity to move beyond the anecdote and look at robust evidence, to present summaries of critically appraised studies and to add to the accumulation of research knowledge. The experience of embracing the systematic review method in medicine and health services suggests a positive contribution to the driving up of research standards and their methodological reporting (Evans and Benfield 2001), which may be a welcome development in social policy. Other long run benefits are harder to determine from our review. Work is to be conducted to get a sense of how well policymakers received the safety nets for mortgagors review, both in terms of their reception of the new methodology, and whether the impact of the findings were fed directly into the policymaking process, or whether they have been incorporated into the current understanding of those active in the field.

Comparison with an unsystematic review
Several literature reviews regarding mortgage safety nets were found in the literature search. Most of these provided scant methodological details of the reports or research studies that were cited, and often included the same small pool of studies. One review of note however, provided an evaluation of the data and evidence on mortgage safety nets and proved useful in considering the merits of systematic reviewing. Ford and England’s (2000) Data and Literature on Mortgage Interest: State Provision and Private Insurance. An Evaluation Report and Source Book is used below to illustrate the difference between the two approaches to reviewing existing research. Ford and England’s review provides a discussion of the literature and a synopsis of the forty-three included studies, including their methodology and stated limitations and perhaps represents one of the best examples of an unsystematic approach to synthesising existing research.

Ford and England aimed to review ‘the available literature and research relevant to safety net provision’. Their review had a wider remit than the systematic review and placed the discussion of the evidence on safety nets within the context of the owner occupied market, and mapped the literature against the wider trends. The systematic review (Croucher et al, 2003) used a specific review question, searched extensively for literature that could answer that question, specified what type of studies were to be sought and appraised the studies to ensure rigor and confidence in the findings. Ford and England’s review differed markedly on all these points. Table 3 presents the breakdown of the studies included in Ford and England’s review by their classification in this review.
Table 3: Classification of Ford and England's studies in this systematic review.

<table>
<thead>
<tr>
<th>Included in systematic review</th>
<th>Passed inclusion, but not the quality threshold</th>
<th>Did not pass Inclusion Criteria: reason</th>
<th>Systematic review not able to retrieve studies in time allowed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>8</td>
<td>6</td>
<td>13</td>
<td>2</td>
</tr>
</tbody>
</table>


Ford and England's review was of the available literature. It may have been beyond the scope of that study to search extensively for additional literature, but it is unclear whether other studies were considered for inclusion, and if so, on what basis other studies were excluded. The initial reference set for our systematic review included several reports based upon stakeholders’ expert opinions, studies addressing the contextual issues of perceptions of risk, or causes of mortgage arrears, for example, that could have been considered for Ford and England's review but were not included. In addition there were six studies included in the systematic review, which, for reasons other than date, were not included in Ford and England. There were also studies included in their review that did not meet the quality threshold. It is recognised that they addressed issues that were pertinent and valuable, but because of problems with the methods or data used in the analysis, the findings were not wholly persuasive. Ford and England's synopsis of each study included a classification of the methods and any stated limitations, but robust empirical studies and other studies reporting expert opinion or theoretical work enjoyed parity. It is left up to the reader to ascertain their worth.

The findings from the two reviews are similar, but for a different question or in a different field, this may not be the case. A policy field where there are many competing voices has the potential to produce very different review conclusions than a review from one ‘camp’ where the recommendations may be more partisan. The distance between the Expert Panel or advisory group and the reviewers may also provide added value for some end users of a systematic review.

The cost of Ford and England’s review would have been significantly below that of the systematic review and was completed more quickly, but the costs and benefits of a resource intensive systematic review and more traditional synthesis of existing research cannot yet be determined until its long run impact and reception by policymakers is assessed. There may be occasions when a quicker but less rigorous review may suffice, but clarity is required over what may be claimed for the findings.
Methodological challenges

The use of systematic reviewing in social policy may be welcome but is not unproblematic. The challenges presented to us in conducting our review were based upon: finding the literature, appraising the quality of the work, and synthesising that evidence from different studies produced against a background of complex contextual factors. In addition, matters of ironing out semantic debates about new reviewing methods need appropriate attention.

Searching

An initial challenge exists in the searching and retrieval of literature in the field. There are problems with the soft nature or multiple uses of many terms, and poor subject indexing of studies, resulting in electronic databases producing a high level of unrelated material. The quality of abstracts is also problematic which led to the unnecessary purchase of many documents that later proved to be discursive pieces or irrelevant. Time resources or inter-lending library costs could have been less if abstracts were structured in such a way to clearly indicate the nature of the evidence or research project reported.

Methodological reporting

Poor methodological reporting was not confined to the abstracts however, many studies paid scant attention to explanations and justifications of the methods employed. Campbell et al (2002) identified the poor reporting of qualitative research methods. It is a matter of concern for a number of reasons. Firstly, research evidence is not immune from the increased scrutiny placed on professions (Boaz et al, 2002), and it cannot assume people’s trust and belief without demonstrating the rigor of its means. Secondly, there is an increase in the evidence based policy and practice movement with many social care organisations conducting critical appraisal training for front line qualified workers such as social workers (see Centre for Evidence Based Social Services (CEBBS); Spittlehouse et al, 2000; Critical Appraisal Skills Programme (CASP)). This training is designed to support the utilisation of research for front line practice, and so researchers wishing wide dissemination of their findings will find it necessary to persuade others of the trustworthiness of their research. Thirdly, whilst little consensus exists of what constitutes good systematic reviews in social policy the methodology is attracting increasing interest from research commissioners. Methodological clarity is important for the positive inclusion of studies in future reviews. Despite extensively searching for related papers that may explain the methodology of a pertinent part of a large research project, the mortgage safety net review was unable to include one study because of the paucity of information relating to the research methods. Journal editors, research commissioners and researchers themselves must take these issues on board.
**Synthesis**

The mortgage safety nets review adopted a narrative account of the evidence, examined using a framework of effectiveness based on the structure, process and outcomes of an intervention. Evidence was presented in much the same way as the authors, and little further analysis was conducted to draw the emerging themes together. It must be said that a gross underestimate was made of the time needed for this stage of the review. It is apparent that considerations of how heterogeneous evidence is presented are problematic, especially if some evidence is stronger than others. On which do you base any policy recommendations? Many researchers are considering how qualitative and quantitative evidence can be synthesised with many current research projects underway. The statistical pooling of quantitative data by meta-analysis is no longer helpful for all reviews, when evidence from disparate study designs are brought together, and new methods of data synthesis are developing. Meta-ethnography looks promising (Noblitt and Hare, 1988; Campbell et al, 2002) but is inductive and driven by the literature. How are gaps in the literature to be identified, comparison made to stakeholders experiences or different strengths of evidence indicated? Forbes and Griffiths (2003) offer a solution combining inductive and deductive review methods with a large-scale consultation exercise.

One difficulty our review encountered that cannot be underestimated was the shifting context to the existing research evidence, and the subsequent interpretation of that evidence. Shifting economic cycles, interest rates, unemployment rates, policy changes, changing intervention standards, lenders changing forbearance practices, all mediated the outcomes of the safety net interventions. The interventions were set against these wider contextual factors that also shifted temporally. This was indicated to readers by the provision of a contextual map (Fig.1) highlighting the fact that it was beyond the resources of the review to capture the wider context, and that the review was tightly targeted on the operation of the interventions and only the micro processes that may mediate outcomes. A timeline was also supplied; plotting the included studies against a set of key milestones relating to how other macro contextual factors had changed. It was intended that these measures would indicate to readers of the review that these factors should be considered when interpreting the evidence. Although medical understanding may change over time, the pace of change in social policy and the impact of social, economic, political and cultural events are arguably greater. Thoughtful consideration of these issues is needed for some subjects under review.

**Consensual standards for reviewing**

Methodological standards for systematic reviewing need to be developed in social policy fields. The Campbell Collaboration oversees systematic reviews in the fields of social, behavioural and educational interventions (www.campbellcollaboration.org). It parallels the Cochrane Collaboration, who conducts strict reviews of effectiveness in health care and medicine. The Campbell Collaboration supports many methods groups looking at developing standards, and there are many working on methods of synthesis or types of acceptable evidence in other arenas (ESRC Research Methods Programme). It is important for the methodology to maintain its position as one that
provides a survey of the best quality evidence that is available and relevant to answer policy or practice questions with the minimum of bias.

Judging by the number of calls for review work, it is apparent that many research commissioners, attracted by the results of systematic reviews, expect full reviews in a time span of only two or three months, which is simply unfeasible. Conducting the extensive search and document retrieval for the mortgage safety nets review took this long alone, and checking the relevance, inclusion criteria, data extraction and synthesis took considerably longer. The EPPI centre suggest two reviewers at 50% fte for 9 months as the basic resources required to conduct a review, but those considering systematically reviewing should also consider the support of research assistance in managing the database and document retrieval, and the factoring in of information scientist support for the conducting the electronic searching. Systematic reviews are therefore resource intensive.

Conclusions

Forbes and Griffiths (2002) make points about evidence-based health care (EBHC), the sentiment of which can easily be applied to social policy or social care. Although working in the field of nursing research, they suggest ‘the principles of EBHC have much to offer all professional groups in resolving the common questions: what is in the best interest of the people whom I seek to help? What is required is not therefore a rejection or demonisation of EBHC but the development of rigorous methods complementary to the principles of EBHC, which can deal with complexity and different forms of empirical expression and are acceptable to practitioners and policymakers who will use their output to inform their clinical decisions’ (p143).

The systematic review method provides a discipline in which sense can emerge from large volumes of literature, when in other circumstances researchers or policymakers may find themselves simply overwhelmed. The focus on what is confidently known about a subject, rather than what is perceived as known, is illuminating. The methods are based upon transparency and explicitness in the reporting of how the review and conclusions have been constructed, and need not be bound by certain study designs, types of questions or topics, as long as reviewers are as reflexive about their review as is expected in primary research. Challenges are present that hopefully co-ordinated work in the methodological field will address in the coming period. Those in social policy have a valuable contribution to make to the developing methodology to ensure the products of future reviews are nuanced presentations of robust concepts and evidence, to inform service users, policymakers and the research community alike.

Acknowledgement

We would like to thank Dr Hilary Arksey for her ever helpful feedback on the original draft of this paper, and the ESRC who supported this work Systematic Reviewing in Social Policy and Social Care.
References


Beishon M (undated) Proving a Point [http://www.hda-online.org.uk/hdt/0301/evidence.htm](http://www.hda-online.org.uk/hdt/0301/evidence.htm) accessed 16/6/03


Centre for Evidence Based Social Services (CEBBS) [http://www.ex.ac.uk/cebss/](http://www.ex.ac.uk/cebss/)

Critical Appraisal Skills Programme (CASP) [http://www.phru.org.uk/~casp/casp.htm](http://www.phru.org.uk/~casp/casp.htm)


Gorard S (2002) *How do we overcome the methodological schism (or can there be a compleat researcher)?* Occasional Paper 47, Cardiff University School of Social Sciences.


Lucas P and Liabo K (2003) *One to One, non-directive mentoring programmes have not been shown to improve behaviour in young people involved in offending or other anti-social activities* Evidence Nugget, London, What Works For Children. ESRC Evidence Network Node. [www.whatworksforchildren.org.uk](http://www.whatworksforchildren.org.uk)


Popay, Rogers and Williams (1998) ‘Rationale and standards for the systematic review of qualitative literature in health services research’, *Qualitative Health Research*, 8, 3, 341-51.


