



Health Development Agency

Integrative approaches to qualitative and quantitative evidence

*Mary Dixon-Woods¹, Shona Agarwal¹, Bridget Young²,
David Jones¹ and Alex Sutton¹*

1. Department of Health Sciences, University of Leicester

2. Department of Psychology, University of Hull

Corresponding author:

Dr Mary Dixon-Woods, Senior Lecturer in Social Science and Health, Department of Health Sciences, University of Leicester
Tel: 0116 2523156, Fax: 0116 2523272, Email: md11@le.ac.uk

Copies of this publication are available to download from the HDA website (www.hda.nhs.uk).

Health Development Agency

Holborn Gate

330 High Holborn

London

WC1V 7BA

Email: communications@hda-online.org.uk

ISBN 1-84279-255-5

© Health Development Agency 2004

About the Health Development Agency

The Health Development Agency (www.hda.nhs.uk) is the national authority and information resource on what works to improve people's health and reduce health inequalities in England. It gathers evidence and produces advice for policy makers, professionals and practitioners, working alongside them to get evidence into practice.

Contents

Foreword	iv
Summary	1
What is the role of qualitative approaches in traditional trials and experimental studies?	2
Background	2
Multi-method research	2
Qualitative research in trials	3
Why synthesise evidence from quantitative and qualitative studies?	4
At what point in the development of a field of knowledge is it appropriate to pull qualitative and quantitative learning together?	6
Generating hypotheses and questions	6
Informing the selection of outcomes for review	6
Extending or guiding sampling	6
Providing explanations and informing conclusions	6
What constitutes good evidence from qualitative studies?	7
Are there hierarchies of evidence within the different types of qualitative investigation?	10
How may the findings of qualitative and quantitative research be synthesised?	11
Advice from the NHS Centre for Reviews and Dissemination	11
The interpretive/integrative distinction	11
Narrative summary	12
Thematic analysis	15
Grounded theory	15
Meta-ethnography	17
Estabrooks, Field and Morse's aggregation of findings approach	19
Qualitative meta-analysis	19
Qualitative meta-synthesis	19
Meta-study	21
Miles and Huberman's cross-case data analysis techniques	22
Content analysis	23
Case survey	24
Qualitative comparative analysis	24
Bayesian meta-analysis	25
Meta-needs assessment	27
Discussion	28
Quantitising and qualitisising	28
Issues in qualitative synthesis	28
Conclusions	31
References	32

Foreword

In 2000 the Health Development Agency (HDA) was established to, among other things, build the evidence base in public health, with particular reference to reducing inequalities in health (Department of Health, 2001). Since then the HDA has been engaged in a programme of developing methodologies and protocols to do precisely that (Swann *et al.*, 2002).

There are two important ideas behind the HDA's remit, political and scientific. The political imperative is a clear commitment to tackling the long-term problem at the heart of public health – that, as the health of the population as a whole continues to improve, at the same time the gradient in inequalities in health across the population, from the most to the least advantaged becomes worse (Acheson, 1998) There is, in other words, a long-standing problem of inequalities in health. The scientific principle is that the best available evidence should be used in order to tackle the problem of inequalities. This is, then, an evidence-based solution to a question of social justice.

The HDA has therefore taken on the task of synthesising and building an evidence base about effective interventions for health improvement, especially for the most disadvantaged. The original questions that Archie Cochrane (1972) posed about clinical medicine provide the Agency's focus. Do we know whether an intervention works? Do we know whether it does any harm? What is the best way to determine whether it works or does any harm? And, if possible, do we know what it costs?

So the HDA was faced with the job of drawing together the known evidence and putting it in a form that could be usable for practitioners and policy-makers, and in such a way that the evidence could inform questions about effectiveness and inequalities. This was a huge undertaking. There is a wealth of potential evidence with everything from randomised controlled trials, through observational studies, to a vast

array of qualitative and action research studies, as well as umpteen local project evaluations. This material is highly diverse, methodologically very mixed, and based on all sorts of different philosophical and scientific assumptions.

Our chosen first step was to tackle the evidence that was most easily dealt with – quantitative studies, especially the existing literature on public health improvement which uses trials and related methods. The beginning of our programme of building the evidence base involved synthesising the world's English language systematic review literature on a number of discrete public health topics. These include social support in pregnancy (Bull *et al.*, 2004), sexually transmitted infections (Ellis and Grey, 2004), obesity (Mulvihill and Quigley, 2003) and many others. A full set of documents can be found on the HDA's website at www.hda.nhs.uk/evidence.

However, this was only the first step, and presently the HDA has embarked on a programme of work to develop ways of introducing other types of evidence into our reviews. This paper deals with just one of those sorts of potential evidence – that drawn for qualitative approaches. There is potentially an extremely rich source of evidence and data, generated using qualitative techniques, which needs to be embraced in the evidence base.

The best way to proceed in order to draw different types of evidence together is not immediately self-evident – whether we are interested to synthesise findings from qualitative studies themselves, or to combine qualitative and quantitative evidence. The HDA has therefore been conducting a number of pieces of work to assist in finding practical solutions to the problem of evidence synthesis across different methodological traditions. This commissioned report is one of those pieces of work.

The issue of combining data from different research traditions is of paramount importance. Trial data reported

in systematic reviews, while very useful, covers only part of the picture and deals with some of the issues. Thus the HDA has found recurrent biases as we have synthesised quantitative evidence from systematic reviews. There are a number of issues. There is a simple absence of information. We have found that only a very small amount of research actually deals with interventions (Millward *et al.*, 2003a). Aside from the missing data, we have found content bias – certain research questions are not investigated because they are not amenable to being answered by a randomised controlled trial. One of the most striking content biases is an absence of studies dealing with inequalities as a research question in its own right (Millward *et al.*, 2003b). We have found process bias, the exclusion of process data from trial reports. We have found artefact bias, by which we mean that the method used obscures the meaning of the data because the research questions have been asked in particular ways. So the kinds of things that are of central importance to patients – their families, for example – are simply not tapped into. So methods such as those produced in qualitative research, which can allow us to draw on these things, must be a part of the lexicon of the evidence base. The question is, how best to do it?

In this paper Mary Dixon Woods and her colleagues have reviewed the literature. This very comprehensive and detailed survey of the field identifies a number of possible avenues which could, relatively easily, be used as a means of developing synthesis across different methodological traditions. The report also shows how contested the field is. At the heart of the controversy is a view held by some pundits that the methodological and philosophical differences between the ‘positivistic’ and the ‘interpretive’ traditions are so great, and the epistemologies and ontologies that underpin them so profoundly separate, that the exercise of evidence synthesis is philosophically impossible and scientifically meaningless. Dixon-Woods acknowledges the argument, but in a highly measured and scholarly way demonstrates some of the potential ways forward.

The position of the HDA, though, is very straightforward and robust. First, and most importantly, inequalities in health and the human pain and misery that flow from them are too great to be ignored on the grounds of philosophical or methodological problems. Certainly there are epistemological and practical differences between research traditions, but this in itself is not a reason for inaction. The HDA cannot detain itself with what, in our view, is a misrepresentation of the philosophical issues. The practical problem to be solved is that of the premature mortality and excess morbidity that

disproportionately affect the poor to a far greater degree than the well-to-do. Second, the characterisation of the debate as an irresolvable one between positivism and interpretivism is disingenuous in our view. It is a device that obscures more than it reveals. This is because, as this paper shows, the interpretivist side is heterogeneous and very subtle and parts of it, at least, are far from incompatible with quantitative reasoning. The same is true of the so-called positivist side of the fence. We contend that the label ‘positivist’ is usually misused in a stereotypical and stigmatising way, which does an injustice to the original ideas of positivism and, more importantly, labels some very serious science as if it were in some sense second-rate. The HDA’s position is that much of the methodological debate is completely unhelpful from a point of view of trying to bring about reductions in inequalities in health, does nothing to help develop policy and practice, is a gross oversimplification of vast amounts of important scientific work in a range of methodological traditions, and – as a final shot – is a misrepresentation of the philosophical principles which supposedly are the origins of the so-called divide. Anyone with a familiarity with Plato (1974) or the other Greek philosophers and mathematicians in whose work these divides are said to originate will know that the old chestnut of positivism versus constructivism or interpretivism is simply neither a fair nor reasonable account of the philosophical principles involved.

This paper deals with some of the technical complexities, and opens up a number of possibilities which the HDA will be taking forward over the coming several years.

Professor Mike Kelly, Director of Evidence and Guidance
Dr Catherine Swann, Head of Research and Evidence

References

- Acheson, D. (1998). *Independent inquiry into inequalities in health*. London: The Stationery Office.
- Bull, J., McCormick, G., Swann, C. and Mulvihill, C. (2004). *Ante- and post-natal home-visiting programmes: a review of reviews: Evidence briefing*. London: Health Development Agency.
- Cochrane, A. L. (1972). Effectiveness and efficiency: random reflections on health services. London: *British Medical Journal*/Nuffield Provincial Hospitals Trust. (1989 Reprint with postscript and obituary.)
- Department of Health (2001). *A research and development strategy for public health*. London: Department of Health.
- Ellis, S. and Grey, A. (2004). *Prevention of sexually transmitted infections (STIs): a review of reviews of non-clinical interventions: Evidence briefing*. London: Health Development Agency.
- Millward, L. M., Kelly, M. P. and Nutbeam, D. (2003a). *Public health interventions research: The evidence*. London: Health Development Agency.

Millward, L. M., Morgan, A. and Kelly, M. P. (2003b). *Prevention and reduction of accidental injury in children and older people: Evidence briefing*. London: Health Development Agency.

Mulvihill, C. and Quigley, R. (2003). *The management of obesity and overweight: an analysis of diet, physical activity and behavioural approaches: Evidence briefing*. London: Health Development Agency.

Plato (1974). *The Republic*, trans. D. Lee, 2nd edn. Harmondsworth: Penguin.

Swann, C., Falce, C., Morgan, A. and Kelly, M. (2002). *HDA evidence base: process and quality. Standards manual for evidence briefings*. London: Health Development Agency.

Summary

This report is an informal review of the literature on integrating qualitative and quantitative forms of evidence. It explores five key themes.

First, it discusses the role of qualitative approaches in traditional trials and experimental studies, beginning with a general discussion of the rationale for multi-method research. It then describes some current attempts to integrate qualitative approaches with trial designs. These approaches may have particular value in sensitising trial design to the needs and priorities of patients.

Second, the report discusses at what point in the development of a field of knowledge it is appropriate to pull qualitative and quantitative learning together. Most of the precedents in this area derive from primary research, but there has been little discussion or research.

Third, the report addresses the complex question of how to determine what constitutes good evidence from qualitative studies. The appraisal of qualitative studies is a contested area, reflecting diverse disciplinary traditions and allegiances, and many of the limitations of appraisal processes reflect methodological anarchy at a primary research level.

Fourth, we offer a brief discussion of whether there are hierarchies of evidence within the different types of qualitative investigation, and conclude that it is unlikely that consensus can be achieved.

Finally, the report considers in some detail how the findings of qualitative and quantitative evidence may be synthesised. It offers an overview and critique of the following strategies, illustrated where possible and appropriate with examples: informal narrative review; thematic analysis; grounded theory; meta-ethnography; aggregation of findings; qualitative meta-synthesis; meta-study; Miles and Huberman's cross-case data analysis techniques; content analysis; case survey methods; qualitative comparative

analysis; and Bayesian meta-analysis. A discussion of some of the theoretical and methodological issues that remain to be resolved is presented, and the report concludes with some directions for research and development.

The report seeks to address five related questions:

- 1** What is the role of qualitative approaches in traditional trials and experimental studies?
- 2** At what point in the development of a field of knowledge is it appropriate to pull qualitative and quantitative learning together?
- 3** What constitutes good evidence from qualitative studies?
- 4** Are there hierarchies of evidence within the different types of qualitative investigation?
- 5** How may the findings of qualitative and quantitative research be synthesised?

What is the role of qualitative approaches in traditional trials and experimental studies?

Background

Any discussion of the role of qualitative approaches in trials and experimental studies should be seen in the broader context of the rationale for multi-method research. Until recently the apparent paradigm schism between qualitative and quantitative research has inhibited the development of studies combining both approaches. This rift is based on apparently deep-seated epistemological and ontological concerns and, in particular, the divide between positivism and interpretivism which has been increasingly emphasised in debates from the 1960s onwards. Barbour (1998) goes further, stressing not only the paradigmatic distinctions between qualitative and quantitative research, but also the distinctions between different types of qualitative research. It is increasingly being argued that these paradigmatic divisions are overstated, and that they can be ignored (Bryman, 1988). Multi-method research is seen as offering a way forward, particularly in addressing especially complex questions.

Multi-method research

Multi-method research involves using more than one type of research technique or data source within a study. The notion of triangulation – strategically comparing a range of perspectives on a particular issue – forms part of the rationale for multi-method research. Triangulation within psychological research was initially conceived by Campbell and Fiske (1959). It was argued that using multiple methods could enhance confidence in the findings, in particular by increasing the ability to evaluate convergent and discriminant validity. More recently, as Bryman (2001) outlines, the rationale of multi-method research has been seen to lie in its perceived ability to compensate for the weaknesses of any single research design, although Bryman suggests that multi-strategy research may be a better way to conceptualise research designs involving qualitative and quantitative elements.

Greene *et al.* (1989) suggest the following rationale for combining qualitative and quantitative methods:

- To achieve convergence of results
- To identify overlapping facets that emerge on closer inspection using multiple methods
- To augment the information gained from an initial approach
- To identify and examine contradictions obtained from multiple sources
- To add scope and breadth to a study
- To guide the use of additional sampling, data collection and analysis techniques.

Duffy (1987) proposes a number of other benefits of triangulation, including external validation of empirically generated constructs, use of case studies to illustrate statistically derived models, and using quantitative assessment to correct for the 'holistic fallacy' (the perception that all views are congruent, when only those interviewed by the researcher may have held that particular view).

More recently much of the debate has focused on how far one or another method should be most dominant. Hammersley (1996) helpfully outlines three approaches to multi-strategy research:

- Triangulation, in which the aim is corroboration (one method is used to verify the findings of the other)
- Facilitation, in which one strategy facilitates or 'assists' the other
- Complementarity, in which two strategies are employed to investigate different aspects of a problem.

Popay and Williams (1997) distinguish between a 'difference' and an 'enhancement' model of the role of qualitative research in healthcare. The enhancement model may:

- Assist understanding of why interventions work
- Improve the accuracy and relevance of quantitative evidence
- Identify appropriate variables to be measured
- Explain unexpected results
- Generate hypotheses to be tested through quantitative research.

Popay and Williams appear to criticise the tendency for qualitative research to be used mainly in the enhancement role, rather than having something distinctively different to say that is sufficient in its own right. They imply that an enhancement model sees qualitative research as having a role that is complementary but subsidiary to quantitative research. A difference model, by contrast, emphasises how qualitative research can contribute to evidence-based practice independently of the research generated by other methodologies.

Qualitative approaches are increasingly penetrating traditional trials and experimental studies. Qualitative research has played both enhancement and difference roles in a number of recent studies, with increasing evidence that qualitative research is being used to challenge taken-for-granted assumptions in trial design and practices.

Qualitative research in trials

Qualitative researchers have often aligned themselves politically with more vulnerable groups, and have argued that qualitative research can empower usually disenfranchised groups in research processes by giving them a voice (Guba and Lincoln, 1989). In healthcare, in particular, it is argued that qualitative researchers are in some sense the patient's advocates. Qualitative research has been seen to be well suited to identifying the concerns and priorities of patients, and is often argued to be well placed to promote their interests. The legitimacy and desirability of this position is open to question, and it could be argued that some forms of quantitative research could also operate as a form of advocacy for patients.

Qualitative research to ensure acceptability of trial objectives and design and to select the outcomes of interest to patients

The use of qualitative research to inform the design of studies and the information used in recruiting patients is increasingly being advocated. The arguments for this rest on the particular ability of qualitative research to elicit what participants in a process see as being important and

significant. Thornton and Dixon-Woods (2002) have argued that well conducted qualitative research is necessary to establish the feasibility and acceptability of trial designs for participants. They suggest that patients' perspectives are too often ignored in developing and designing studies, resulting in two key problems. First, the study design may be unacceptable to patients or questions may appear irrelevant or illegitimate. Second, the processes of recruitment (explanation, information, support) may be poorly managed. These elements are the most under-researched and least well understood of trial practices.

Qualitative research could help identify trial questions that are relevant to patients' concerns and ensure the trial design and recruitment processes are optimally fitted to patients' own priorities. For example, interventions that focus exclusively on 'clinical' outcomes such as tumour regression may not adequately address patients' concerns about the impact of the intervention on other outcomes that may be important to them, such as physical appearance or the effort and time required to comply with the intervention. The importance of this is underlined by Tallon *et al.* (2000) who used qualitative research, among other methods, to demonstrate a mismatch between the priorities of consumers and the interventions studied in trials. Koops and Lindley (2002) showed that qualitative research encouraging the involvement of consumers in designing trial leaflets significantly changed the content and form of the leaflets, while Donovan *et al.* (2002) demonstrated improvements in recruitment rates when qualitative research informed the explanations of the different arms of the trial. These uses of qualitative research may not be necessary in all trials, but they represent important opportunities for those conducting trials.

Qualitative research to evaluate trials from the participants' perspective

Qualitative research has played an important role in illuminating participants' experiences of trials. Cox (2000) interviewed 55 patients who participated in early phase cancer trials, and found that patients did not fully understand what they were taking on when they entered a trial, and felt participation to be burdensome and demanding.

Qualitative research has also been successful in investigating some aspects of professionals' experiences of involvement with trials. Langley *et al.* (2000) used semi-structured interviews to explore clinicians' views on recruiting patients into cancer trials. This was important in identifying some of the problems and concerns that clinicians experience.

Qualitative research to investigate aspects of trial practices

Qualitative research has been used to investigate aspects of trial practices, including the commonly advocated notion that trial results should be fed back to participants. Snowdon *et al.*'s (1998) qualitative study indicated that a cautious approach to feedback is warranted, even in seemingly harmless circumstances, and that it is very important that only those who opt to receive information about the outcomes of studies are sent it.

Participants' understanding of key concepts in trials, including randomisation, has been explored using qualitative research. Featherstone and Donovan (1998) reported important differences between lay and professional concepts of randomisation and trial, and identified the consequences for informed consent. Snowdon *et al.* (1999) noted that Zelen randomisation, in which randomisation to trials precedes consent, had been the subject of academic and professional discussion, but the views of those who had experience of participating in such trials were not known. They conducted in-depth interviews to explore the views on Zelen randomisation of parents whose children who had been recruited to a neonatal trial, and found that parents were divided in their acceptance or rejection of this form of randomisation. Those who would find the method most difficult to accept were those who were allocated to the control arm. Snowdon and colleagues concluded that Zelen randomisation might widen rather than narrow perceived inequities in trials, although Braunholtz (1999) in a later commentary emphasised that this effect was likely to have been particularly prominent in the trial Snowden and colleagues investigated, and might vary depending on the trial.

Qualitative research to help understand the links between evidence and practice

One of the key roles of qualitative research is in addressing difficult questions of how evidence can inform practice (Green and Britten, 1998) and, in particular, to understand professional and patient influences on the implementation of evidence from trials. For example, a qualitative study of how general practitioners accessed and used evidence about statin drugs in their management of hypercholesterolaemia showed that for doctors, information from trials acquires its status as practical knowledge not from scientific rigour, but from clear local consensus that takes account of the context in which the trial data are to be used (Fairhurst and Huby, 1998). Qualitative research can also help inform the

development of interventions to encourage the uptake of evidence into practice. A study by Butler *et al.* (1998) set out to identify interventions to reduce unnecessary prescribing of antibiotics for sore throats. Qualitative interviewing of a sample of GPs established that doctors were well informed of the scientific evidence, but often prescribed in order not to jeopardise the doctor–patient relationship. Educational interventions targeted at changing GPs' behaviour were therefore unlikely to be effective, and the authors concluded that future trials should focus on interventions to influence doctors' consultation techniques.

Why synthesise evidence from quantitative and qualitative studies?

Rationale for synthesising qualitative research

Estabrooks *et al.* (1994) make a case for undertaking what they call 'aggregation' of qualitative studies, arguing that analysis and synthesis of multiple studies can result in a contribution to theory-building more powerful than any single study. Synthesis of evidence can allow the construction of larger narratives and more general (and generalisable) theories (Hammersley, 1997; Sandelowski *et al.*, 1997). It can therefore overcome the problem of isolation associated with qualitative research, allowing cross-study themes or higher-order analytical categories to be developed. Stiles (1993) gives the example from primary research of a study by Wiersma (1988) in which multiple case studies of women who had made midlife transitions from a domestic career to one outside the home revealed more general insights into how society maintains gender segregation in the world of work.

A further argument for pooling qualitative studies concerns the optimal use of primary data. Some questions can be answered only by examining a range of data sources, particularly when events are too rare to be identified frequently within a single study (Thorne, 1994). For example, ethnographic studies may identify aspects of a phenomenon left uncovered by interview-based studies (eg 'errors' in the delivery of interventions), while many studies may be required to identify rarely occurring aspects of that phenomenon (eg errors relating to gross incompetence). Synthesis, by examining the topic with access to a broader, larger sample, can provide more powerful explanations than those available from single studies (Sherwood, 1999). It may also have a particularly valuable role in inquiry into sensitive topics or hard-to-reach populations, because such analysis enables maximum value to be gained from studies that have been able to overcome problems with accessing the settings.

More pragmatically, failing to optimise the outcomes of the time-consuming investment in qualitative research is clearly wasteful (Thorne, 1994). This was a concern empirically confirmed by Campbell *et al.* (2003), who found that of seven qualitative studies on patients' experiences of diabetes, not one referred to another. This problem can undermine qualitative inquiry as a collective and incremental enterprise, and lead to reinvention of the wheel.

There are now strong precedents from primary research for combining qualitative and quantitative evidence in secondary analysis.

Other arguments for qualitative synthesis include:

- Helping to explore the diversity of effects across studies, settings and groups (NHS CRD, 2001)
- Identifying differences between interventions in apparently similar studies or between the contexts within which interventions are delivered (NHS CRD, 2001)
- Illuminating the impact of contextual factors (NHS CRD, 2001)
- Integrating, synthesising and organising research results into coherent patterns that can be more easily evaluated for application to clinical practice (Sherwood, 1999)
- Generating operational models that can be tested or applied to practice (Sherwood, 1999)
- Developing grounded theory for future hypothesis testing (Sherwood, 1999)
- Specifying concepts to clarify the phenomenon of interest in operational terms (Paterson *et al.*, 2001)
- Consolidating widely scattered literature on a defined topic
- Generating bibliographies of research in particular areas
- Identification of gaps where further research is required (Paterson *et al.*, 2001)
- Generating models and theories (Estabrooks *et al.*, 1994; Thorne and Paterson, 1998; Finfgeld, 1999)
- Validating and assessing the empirical support for theories (Forte, 1998)
- In methodological reviews, determining the existence of certain bodies or schools of thought
- Charting the development of concepts longitudinally in the literature over a period of time (Thorne and Paterson, 1998)
- Complementing the findings and interpretation of quantitative systematic reviews
- Identifying significant domains or attributes to enable the development of prototype instruments or scales.

The historical antagonism between quantitative and qualitative methods in primary research is now often eschewed in favour of multi-method approaches integrating a range of different research perspectives and techniques.

At what point in the development of a field of knowledge is it appropriate to pull qualitative and quantitative learning together?

Little has been written on when it might be appropriate to begin to draw qualitative and quantitative evidence together, but some inferences can be drawn from the literature on primary multi-method research.

Generating hypotheses and questions

Either qualitative or quantitative research can be used to generate hypotheses or research questions that can be addressed through subsequently synthesising both types of evidence.

Informing the selection of outcomes for review

An important task for reviewers is selecting the variables to be studied. As described above, qualitative research is increasingly being used to help sensitise research designs to the concerns and priorities of patients. In primary research, it is increasingly rare in clinical trials for outcome measures such as patient satisfaction and quality of life to be selected without qualitative research first being used to identify patients' concerns and then to develop and test appropriate measures (Fitzpatrick *et al.*, 1998). However, despite recent moves towards more inclusive reviews, there is a continuing tendency for some systematic reviews, particularly those conducted within the Cochrane movement, to focus exclusively on clinically defined outcomes rather than on those that might be most important to patients (Dixon-Woods *et al.*, 2001). There appear to be good arguments for using qualitative research to help identify or refine the questions to be addressed in a systematic review in ways that would give more recognition to patients' concerns. Such an approach could help challenge the tendency of systematic reviews to reflect clinicians' rather than patients' priorities for care, and restore a more broadly conceived critical perspective to reviews (Berkwits, 1998). Moreover,

as demonstrated recently, failure to identify the outcomes that have been studied in both qualitative and quantitative research means that reviews are likely to be incomplete (Roberts *et al.*, 2002).

Extending or guiding sampling

Multi-method approaches in primary research, particularly those emphasising the qualitative paradigm, have promoted the ability of combined approaches to inform sampling. Much of the rationale for this derives from the concept of theoretical sampling in primary qualitative research, where the selection of samples is guided by ongoing analysis. For purposes of synthesis, it might similarly be possible to select further papers for review based on the emergent analysis, although any selection would be limited by the availability of papers or different types of paper. However, so far there has been relatively little discussion of these issues in the published literature.

Providing explanations and informing conclusions

There are excellent precedents in primary research of using one research strategy to explain the findings of another. For example, Sparks *et al.* (1994) used qualitative data to explain differential childhood accident rates. It is much rarer to find examples of the use of a synthesis of studies done using one type of strategy to explain the findings of a synthesis of studies done using a different type of strategy.

What constitutes good evidence from qualitative studies?

It is difficult to answer with any precision the question of what constitutes good evidence from qualitative studies. Good evidence clearly depends on high quality research, but the characteristics that define good quality research remain contested. Arguments abound about whether criteria for quality in qualitative research could, or even should, exist. Henwood and Pidgeon (1992) argue that evaluative criteria are helpful to keep in mind during analysis and the reading of analyses as they help provide guidance as to 'good practices'. Others have called for an end to 'criteriology' (Schwandt, 1996), arguing that this leads to privileging of method as a 'sacred prescription' that is rooted in positivist philosophical traditions. Criteria are sometimes seen as ideals to which researchers might aspire, but which are practically impossible to achieve, at least simultaneously in the same study. It could also be argued that criteria might overly discipline what are the essentially creative interpretive processes that characterise qualitative research, and could stifle methodological innovation.

Nonetheless, it is clear that some means is needed to determine the quality of studies to be included in a review, for several reasons. First, without a means of quality assessment it will be difficult for reviews involving qualitative research to make the sorts of claims to credibility usually enjoyed by systematic reviews of quantitative studies. Second, the absence of quality assurance in determining which studies are included in a review can cause difficulties in producing reviews. Campbell *et al.* (2003) note that failure to exclude a poor quality study from their meta-ethnography caused difficulties in interpretation and synthesis. This study had been evaluated using the CASP checklist (Critical Appraisal Skills Programme, www.phru.nhs.uk/~casp/casp.htm), but subsequently was found to be unsatisfactory. Third, if studies of differing quality are included in a review they may cause variable but unknown distorting effects in any synthesis.

Various attempts have been made to resolve the current impasse in finding a way to incorporate quality assessments for qualitative research in systematic reviews. This has involved a large number of proposals for quality criteria, which have been useful in identifying criteria that might be used, but which have also led to an unhelpful proliferation and diversity, with often little evidence of common ground (NHS CRD, 2001). An alternative approach is to produce a consensus or core set of criteria, one example of which is the recent version by the National Centre for Social Research (Spencer *et al.*, 2003).

There are, however, several problems with the consensus approach. Engel and Kuzel (1992) argue that consensus on what constitutes 'good' qualitative research is impossible, because there are no 'in principle' arguments that can address this issue across the varieties of qualitative inquiry. A second problem is that the consensus set of criteria being produced by the National Centre for Social Research (Spencer *et al.*, 2003) is based on 28 sets of proposed criteria for good qualitative research. We have identified more than 100 sets of proposals for quality in qualitative research, suggesting that there are possible additional criteria and that the consensus may be only partial. These 100+ sets of proposals require systematic overview which is beyond the remit of the current report.

A third and extremely important problem is that most approaches to the assessment of quality have treated qualitative research as a unified field. Given the diversity of approaches within qualitative research, this is a flawed assumption doomed to produce criteria that suit only particular approaches. Under an Economic and Social Research Council (ESRC)-funded Research Methods project led by Dixon-Woods and colleagues (www.ccsr.ac.uk/methods), research is being undertaken to determine in what ways different guidelines would produce different decisions about inclusion and exclusion of papers, compared with

intuitive expert judgement of quality. This project will provide empirical evidence for the impact of using different types of guidelines and insights into how far different guidelines tend to favour particular approaches. Such a project is necessary because many of the differences between the 100 or so sets of proposals represent divides along methodological and disciplinary allegiances, with little apparent consensus. For example, the criteria developed by Seale and Silverman (1997) and those devised by Popay *et al.* (1998) appear to have several non-overlapping features (Chapple and Rogers, 1998). The Seale and Silverman criteria emphasise the need to support generalisations by counts of events and the use of computer software to assist in data analysis, while the Popay *et al.* criteria do not emphasise the use of quasi-statistics and argue that computer software is neither necessary nor sufficient for rigorous qualitative analysis. Some sets of criteria recommend respondent validation (eg Lincoln and Guba, 1985), a process where the participants in research are asked to check the researchers' interpretations of the accounts they have given or the observations made of the participants. Respondent validation is also suggested (with caution) by Mays and Pope (2000), who observe that a social-scientific interpretation of accounts or observations may be very different from a 'lay' one, but do commend it as a means of 'error reduction' which can also generate further original data. Respondent validation, however, may be an entirely unsuitable form of validity checking for some forms of qualitative research, including for example, discourse analysis, where its often anti-realist emphasis is on the multiple accounts that can be produced of any phenomenon, rather than seeking a 'single' verifiable account.

The situation in qualitative research contrasts with the development of methods for appraising quantitative research, where it is recognised (eg in the work of CASP and the NHS CRD guidance) that different study types (randomised controlled trials, case control studies, studies to evaluate screening programmes, etc) may demand different criteria. This allows the very precise formulation of flaws that would be fatal or very damaging to the rigour of a particular study type. For example, failure to randomise properly in a randomised controlled trial, or to select appropriate controls in a case control study, could be deemed very serious problems. In qualitative research there is a need to recognise that focus groups, interviews, participant observation etc also constitute different study types, and therefore may also need different criteria to allow their appraisal. However, it is vitally important to recognise that, to a much greater extent than in quantitative research, the execution of a qualitative research study type is crucially related to the theoretical perspective

in which the researchers have chosen to locate the study. An interview-based study conducted within a grounded theory framework may therefore have very different characteristics from an interview-based study conducted within a discourse analysis framework, and this needs to be reflected in the framework for appraising a study. The key task then lies in defining what the expectations should be for a particular study design within a particular theoretical field.

This task is by no means straightforward. It is often difficult even to be sure about which study design, and within which disciplinary field, a published study has been undertaken. This problem arises because researchers may describe something as qualitative research when, in the judgement of others, it is not (Campbell *et al.*, 2003); a study may also be described as using a specific approach such as grounded theory when, in the judgement of others, this approach has not been used (Paterson *et al.*, 2001).

The problem is at its worst in the case of grounded theory. Here, one of the most frequently cited social science texts is that of Glaser and Strauss (1967). However, the original vision outlined by Glaser and Strauss is more frequently honoured in the breach than in the observance (Bryman and Burgess, 1994). Inconsistencies, misappropriations and mislabelling of studies purporting to use grounded theory are common problems (Stern, 1994). Some of the reasons for this 'democratization of the scholarship process' (Thorne, 1997) lie in the practical difficulties of implementing grounded theory in its original form (Barbour, 2001; Bryman, 2001). The problem is further exacerbated by the fissure that arose between Glaser and Strauss themselves, who went on to develop two quite distinctive approaches to grounded theory. Strauss's later work focused on the development of technique and procedure, and could be seen as prescriptive in character. Glaser's work, by contrast, can appear to lack a grounding in practicalities and has avoided the development of techniques. For example, his exhortation (Glaser, 1978) that researchers enter the field with no preconceived theories or expectations is hardly realistic, particularly in today's academic and practice environments. It is, however, clear that some elements of grounded theory – in particular the constant comparative method of analysis – are especially attractive and useful, but the tendency to select some of these techniques to create ad hoc and a la carte approaches to qualitative analysis, and still retain the label 'grounded theory', is very unhelpful. The result is that, despite the tendency for researchers to assume that describing a study as 'grounded theory' is sufficient, it is far from self-evident what such a study might involve.

The diversity – near anarchy – in qualitative methodology means that it is very difficult to identify, or at least gain agreement on, what might constitute a fatal flaw in, for example, an interview-based study using an opportunistic sample and the constant comparative method for analysis. There is no ‘gold standard’ set of standard methodological properties of such a study, and therefore little to guide an objective assessment of whether such a study has complied with good practice.

A further problem concerns the failure of many suggestions for assessing the quality of qualitative research (also true of many criteria for quantitative research) to distinguish between issues relating to the quality of reporting and those relating to design and execution. For example, the frequently cited CASP checklist contains a number of detailed questions relating to issues such as how the research was explained to the participants, why some participants chose not to take part, etc. Arguably medical journals rarely publish papers of a length sufficient to allow descriptions of such features of studies, and these issues relate mainly to transparency.

Finally, the failure of most sets of proposals on appraising qualitative research to distinguish between fatal and trivial flaws further hampers efforts to determine which studies should be selected to enter a review. Even Cesario *et al.* (2002), one of the groups who have gone furthest towards producing a system of ranking qualitative research, appear to rely on aggregated scores in which each criterion is treated as being of equal weight. The problem of identifying what might constitute fatal flaws is inherent in the absence of clear and consistent standards to govern study design and execution in qualitative research. The absence of such standards may have many advantages, allowing researchers to devise studies that are flexible and well suited to particular settings and resources. It also has many disadvantages, as is evident in the poor quality of many published qualitative studies. Common problems include the failure to identify which analytical approach has been used, failure to explain how the analysis was undertaken, or claiming to use a particular analytical approach when the analysis itself demonstrates that most of the requirements of this approach have been violated. Most damagingly from the point of view of systematic reviews, however, the problem of non-agreement on quality standards means that it is extremely difficult to select the best, and only the best, qualitative evidence for inclusion, and even more difficult to demonstrate that only the best has been selected.

Are there hierarchies of evidence within the different types of qualitative investigation?

The short answer to whether there are hierarchies of evidence within the different types of qualitative investigation is 'probably not – or at least, not any that everyone would agree on'. Mays *et al.* (2001) argue that the epistemological diversity of the field renders any single hierarchy of evidence inappropriate for qualitative research. Barbour (1998) also cautions against the development of hierarchies of evidence, referring for example to the dangers of using interview data as the yardstick against which to measure the validity of focus group data. Few other commentators have even approached the question of a hierarchy of evidence within qualitative inquiry.

How may the findings of qualitative and quantitative research be synthesised?

Methods for the synthesis of quantitative data are becoming increasingly sophisticated, with techniques such as meta-analysis expanding to include forms of cross-design synthesis that deal with study designs other than randomised controlled trials. Methods for synthesis of qualitative forms of evidence have developed much more slowly, and many are currently in the early stages of proposal and evaluation. In this section we present a selection of approaches to synthesising qualitative and quantitative forms of evidence. For each approach we provide a description and offer a critique, illustrated where possible by examples.

It is necessary first, for the sake of clarity, to distinguish the synthesis of qualitative primary studies from the secondary analysis of qualitative data. The secondary analysis of qualitative research has become a focus of interest in recent years, partly stimulated by the establishment of the ESRC Qualidata Archive Resource Centre, which stores qualitative data for use by other researchers. It is important to note that by 'secondary analysis of qualitative research' (Heaton, 1988; Thorne, 1994) some qualitative researchers appear to mean the re-analysis of data sets originally collected for a different purpose. West and Oldfather (1995), for example, describe a process of 'pooled case comparison' for secondary analysis of data from separate qualitative studies, which involved a re-analysis of the original data sets. Here we discuss techniques for synthesising evidence from studies, and do not discuss in any detail methods for secondary analysis except where these might be useful in informing the discussion.

Advice from the NHS Centre for Reviews and Dissemination

The NHS CRD recently updated its guidance on conducting systematic reviews to include advice on the use of qualitative evidence (NHS CRD, 2001). This was a welcome step forward (Dixon-Woods and Fitzpatrick, 2001), but the guidance

acknowledges the need for research and development in this area, and the problems of approaching such contested issues as how to appraise qualitative research. As a result much of the guidance remains rather general, with little in the way of specific advice on techniques: no example of a data extraction form is provided for qualitative research, for example. However, the guidance does provide a framework within which reviewers can begin to think about including qualitative evidence in systematic reviews, and identifies many of the problems that need to be tackled.

The interpretive/integrative distinction

Noblit and Hare (1988) introduce a useful distinction between integrative and interpretive reviews. They suggest that integrative reviews are rooted in the positivist paradigm. They involve techniques, such as meta-analysis, concerned with assembling and pooling data, and require a basic comparability between phenomena so that data can be aggregated for analysis. Noblit and Hare regard integrative reviews as primarily suitable for synthesising quantitative studies. They argue that interpretive reviews, on the other hand, see the essential tasks of synthesis as involving both induction and interpretation, and are concerned not to predict but to 'anticipate' what might be involved in analogous situations and to understand how things connect and interact. Again, they suggest that interpretive reviews are suitable for synthesising interpretive studies.

While the distinction between integrative and interpretive review is very important and useful, we wish to elaborate on Noblit and Hare's original conceptualisation and to propose a new way of thinking about these different forms of synthesis, rather than different forms of review. First, we want to be careful not to identify integrative with positivist (and all the negative connotations of that term), or with quantitative. Instead, we suggest that integrative syntheses

are those where the focus is on summarising data, and where the concepts (or variables) under which those data are to be summarised are assumed to be largely secure and well specified. For example, in an integrative synthesis of the impact of educational interventions on uptake of influenza immunisation in older people, the key concepts (educational intervention, uptake, older people) would be defined at an early stage in the synthesis and would effectively form the categories under which the data extracted from any empirical studies are to be summarised. This summary may be achieved through pooling the data, using techniques such as meta-analysis, or less formally through techniques such as providing a descriptive account of the data (often referred to as a narrative summary). It is important not to exaggerate how secure such categories might be – there may be debates within and external to a review team about the properties of categories, for example whether ‘older people’ should be defined as those within specific age bands, or by some other measure of chronology. The primary focus of an integrative synthesis is not, however, on the development of concepts, or their specification. This does not prevent integrative syntheses from fulfilling theoretical or interpretive functions. Indeed, Noblit and Hare argue that all types of synthesis involve interpretation. The kinds of theory that integrative syntheses may be especially likely to produce will often be theories of causality, and may also include claims about generalisability.

Second, we want to be careful not to argue that interpretive syntheses can be carried out only on qualitative primary studies. We suggest that in an interpretive synthesis, the primary concern is with the development of concepts and theories that integrate those concepts. An interpretive synthesis will therefore avoid specifying concepts in advance of the synthesis; in contrast with an integrative synthesis, it will not be concerned to fix the meaning of those concepts at an early stage so as to facilitate the summary of empirical data relating to those concepts. The analysis that yields the synthesis is conceptual in process and output, and the product is not aggregations of data, but theory. Again it is important not to caricature an interpretive synthesis as floating free of any empirical anchor – an interpretive synthesis of primary studies must be grounded in the data reported in those studies.

We do not see interpretive and integrative forms of synthesis as completely distinct. They share considerable areas of overlap, and there may be considerable blurring of the boundaries of these types of syntheses. Nonetheless, most forms of synthesis can be characterised as being primarily interpretive or primarily integrative in form and process,

although every integrative synthesis will include elements of interpretation, and every interpretive synthesis will include elements of aggregation of data. In the discussion below we present a selection of methods for synthesis, together with an informal commentary on issues involved in using these methods. The different methods of synthesis (summarised briefly in Table 1) can be grouped in terms of their epistemological and ontological foundations, and whether the aim of synthesis is primarily interpretive or integrative. Clustering towards the interpretive end of the spectrum are the methods of narrative summary, grounded theory, meta-ethnography, meta-synthesis, meta-study, and Miles and Huberman’s (1994) data analysis techniques, while lying at the more integrative end of the spectrum are content analysis, case survey, qualitative comparative analysis and Bayesian meta-analysis. Within these clusters or groups, elements of the methods show some degree of overlap. Grounded theory method as used by Kearney (2001), meta-ethnography and meta-synthesis all involve some form of creative process whereby new constructs are fashioned by identifying related concepts in the original studies, which are then reworked and reformulated to further extend theorising, while still encompassing key features of the original concepts. Content analysis, case survey and Bayesian meta-analysis, on the other hand, each involve the quantification of qualitative data, and here the emphasis is on achieving a systematic integration of the data.

Narrative summary

Narrative review is a term used to describe the informal selection, assembly and summary of studies for review. Before the advent of systematic review methodology, which by contrast is explicit, transparent and systematic about the selection, appraisal and synthesis of studies, narrative review was probably the most common form of review. It is possible to distinguish narrative reviews – which can take a variety of forms – from narrative summary. Narrative summary typically involves the selection, chronicling and ordering of evidence to produce an account of the evidence, often including some kind of commentary or interpretation. Narrative techniques of summary can be combined with systematic techniques for searching and appraisal, and can account for complex dynamic processes, allowing the voice of the synthesiser to be heard most clearly. They can offer complex explanations that emphasise the sequential and contingent character of phenomena (Abbott, 1990).

Mays *et al.* (2001) identify three narrative approaches to the synthesis of findings across diverse study designs: theory-led, analytical, and triangulation.

Table 1 Summary of main approaches to joint synthesis of qualitative and quantitative evidence

Method	Outline of approach	Problems	Strengths
Narrative summary	Narrative description and ordering of primary evidence (perhaps selected) with commentary and interpretation	Lack of transparency re selection and later stages of process	Flexible procedures; can cope with large evidence base comprising diverse evidence types
Thematic analysis	Identification of major/recurrent themes in literature; summary of findings of primary studies under these thematic headings	Lack of transparency re process decisions Largely descriptive basis to groupings, not necessarily developing theory, accounting for contradictions	Can cope with diverse evidence types
Grounded theory	Constant comparative method identifies patterns and inter-relations in primary data; sampling responds to analysis; principles of theoretical sampling may be used	Lack of transparency	Seeks generalised explanations/theories Sampling to theoretical saturation limits number of papers to review Could include quantitative as well as qualitative evidence (by qualitisng)
Meta-ethnography	Reciprocal translational analysis identifies key themes in each study, then seeks to translate these into the context of each other study; themes with best overall fit/explanatory power adopted Attempt made to explain contradictions Seeks general interpretive (lines of argument) synthesis	Does not guide sampling May be order-of-synthesis effects	Seeks higher-order (generalised) theories Could include quantitative as well as qualitative evidence (by qualitisng)
Meta-study	General framework for and guidance on technique, including question formulation and selection of primary studies; division into meta-theory, meta-data, meta-method	Laborious and demanding process	Provides helpful guidance on a range of issues in addition to synthesis Could include quantitative as well as qualitative evidence (by qualitisng) although not explicitly designed to do so

Continued

Table 1 *Continued*

Method	Outline of approach	Problems	Strengths
Miles and Huberman	Evidence from each primary study first summarised and coded under broad thematic headings; evidence then aggregated and summarised within theme across studies, with brief citation of primary evidence Commonalities and differences between studies noted	No guidance on sampling of primary studies	'Systematic' method Audit trail of process possible May allow inclusion of quantitative evidence
Content analysis	Evidence from each primary study coded under broad thematic headings using extraction tool designed to aid reproducibility; occurrences of each theme counted and tabulated	Emphasis on frequency may limit interpretive process	Software available Can incorporate quantitative evidence
Case survey	Each primary study treated as a 'case'; study findings and attributes extracted using closed-form questions, for reliability Survey analysis methods used on extracted data	Applicable to outcomes, but less adequate for process Lacks sensitivity to interpretive aspects of evidence	Can incorporate quantitative evidence
Qualitative comparative analysis	Boolean analysis of necessary and sufficient conditions for particular outcomes to be observed, based on presence/absence of variables and outcomes in each primary study	Focused on causality determination, not interpretive aspects of qualitative data	Transparent Systematic; can incorporate both qualitative and quantitative primary study evidence Allows competing explanations to be explored and retained
Bayesian meta-analysis	Quantified beliefs about effects of variables from qualitative studies formally combined (through Bayes's paradigm) with evidence from quantitative studies	Conceptually simple but may be technically complex to implement (and thus lose transparency)	Impact of analysts' prior beliefs can be explicitly explored

- Theory-led approaches involve exploring the relationship between findings from studies in the context of theories. Mays and colleagues give the example of Harden *et al.* (1999) who addressed evidence from process evaluation studies to consider theories relating to aspects of control and autonomy experienced by young people in relation to health promotion.
- Analytical approaches refer to syntheses involving techniques such as grounded theory, meta-ethnography,

and Miles and Huberman's data analysis techniques, all described later in this report. While Mays *et al.* (2001) describe these as narrative techniques, we would disagree, and feel that narrative techniques in general should refer to more informal approaches.

- Triangulation approaches involve comparing qualitative and quantitative data from different studies, and here Mays *et al.* suggest that reviewers draw on the strategies used by primary researchers.

Narrative, informal and reflexive techniques may be well suited to synthesising diverse literature, particularly that including large bodies of qualitative evidence, and most closely mirror processes at a primary level, but are likely always to be subject to criticism for their lack of transparency.

Thematic analysis

Thematic analysis involves identifying prominent or recurrent themes in the literature and summarising the findings of the different studies under thematic headings. Summary tables, providing descriptions of the key points, can then be produced (Mays *et al.*, 2001). However, in many reports the underlying processes in achieving a thematic analysis are not made explicit, and there is considerable ambiguity about, for example, whether the thematic analysis should be confined to the themes reported in the primary studies; whether some degree of interpretation or adjustment of thematic labels is permissible; and whether themes external to the primary studies, but with some explanatory value, can be imported.

Several recent attempts at providing structured or systematic overviews of diverse areas of evidence have adopted thematic analysis. Garcia *et al.* (2002), for example, report a thematic analysis of women's views of ultrasound in pregnancy, involving initial tabulation of papers and counts of numbers of papers that contributed data on each theme. They did not exclude studies on the grounds of poor quality, and noted the need for quality criteria to determine the studies to be included in a review. They also noted the difficulties of trying to formulate in advance a precise and highly delimited review question (as would be required for quantitative systematic reviews), because the review question was itself moderated by the literature, in a process that mirrors primary qualitative research. Garcia *et al.* (2002) illustrate their thematic analysis by reporting data excerpts (eg quotations) from the primary studies. Other papers that appear to have adopted thematic approaches to synthesis include Harden *et al.* (1999).

There is a tendency in some papers to label (inappropriately) thematic analysis as 'content analysis' – eg Neill's (2000) review of the literature on acute childhood illness at home suggests that content analysis techniques were used, when the reporting indicates that, in fact, thematic analysis was employed.

Thematic analysis offers several advantages: it allows clear identification of prominent themes, and organised and structured ways of dealing with the literature under these themes. It is a flexible approach, allowing considerable latitude to reviewers.

However, thematic analysis also suffers from several important problems.

- There is lack of clarity about exactly what thematic analysis involves and the processes by which it can be achieved, and a lack of explicitness about procedures and aims, including how far thematic analyses should be descriptive or interpretive. For example, it is unclear how themes can be identified, whether the structure of the analysis should reflect the frequency with which particular themes are reported, or whether the analysis should be weighted towards themes that have a high level of explanatory value. Thematic analysis can be either data-driven (driven by the themes identified in the literature itself) or theory-driven (oriented to evaluation of particular themes through interrogation of the literature). The failure of much writing on thematic analysis to distinguish adequately between these two approaches has resulted in a lack of transparency.
- It is limited in its ability to deal with contradictions, other than by describing them.
- If thematic analysis is limited to summarising the themes reported in primary studies, it offers little in the way of theoretical structure within which to develop higher-order thematic categories that go beyond the categories identified from the literature. This limits the ability of thematic analysis to take full advantage of the potential offered by synthesis.

Grounded theory

Grounded theory is a primary research approach that has been hugely influential in the development of qualitative methods in health for more than 35 years. Originally formulated by Glaser and Strauss (1967), the approach describes methods for sampling, data collection and data analysis, broadly within a symbolic interactionist paradigm. It sees the overriding concern of qualitative research as the generation of theory, meaning generalisable explanations for social phenomena. The key features of grounded theory are:

- Data collection and analysis proceed iteratively
- Analysis (the constant comparative method) involves identification of patterns in the data and relationships between these patterns
- Analysis guides sampling
- Theoretical saturation determines the point at which sampling is adequate.

It is clear that, in principle, grounded theory offers a potentially suitable approach to the synthesis of primary

studies. As Estabrooks *et al.* (1994) point out, although Glaser and Strauss themselves appear to have discussed only single studies, there is considerable scope within the original formulation of grounded theory (Glaser and Strauss, 1967) to use multiple studies to move towards a higher, more abstract level of interpretation and theory-building. The constant comparative method, perhaps the most widely used element of grounded theory, is perhaps the most obviously useful element of grounded theory, in part (especially in later formulations by Strauss) because of its appeal in offering a set of procedures by which qualitative data may be analysed.

Examples of grounded theory for synthesis

Despite the arguments in favour of the use of grounded theory to synthesise studies, there are few empirical examples. Dixon (1996) collected 16 qualitative studies of parents' experiences with chronically ill children and, by examining and categorising the terms used by the authors, identified four unifying concepts. Kearney (2001) offers a full-scale grounded theory analysis of qualitative studies of women's experience of domestic violence. Her study explicitly aimed to develop formal theory. She selected studies for inclusion in the study based on her judgement of their methodological rigour and their theoretical contribution, eventually identifying 15 reports of 13 studies for inclusion. From these, she extracted data and assembled them onto a grid to facilitate cross-case comparison.

Kearney proceeded to use the constant comparative method as described by Strauss and Corbin (1998). This involved identifying and clustering concepts identified across studies into new categories, followed by axial coding in which the nature of the categories was more fully specified, and the relationships between categories within and across studies were tested. Selective coding was undertaken to refine and substantiate the core categories that linked the other categories. Theoretical sampling was undertaken by returning repeatedly to the study reports, although it is not very clear what this entailed in practice. Kearney describes using memos with specific links to the source texts to document the analysis. During the course of her analysis, Kearney rejected concepts for which she felt there was no convincing evidence in the primary studies, providing a further level of quality checking, but suggests that it is very unlikely to be possible to exclude qualitative studies in advance on the basis of specified criteria, as is often possible for trials or some other quantitative study designs.

Kearney assessed the quality of her final formal theory using Glaser and Strauss's notions of fitness (accurate

representation of what is going on, to the extent that a person familiar with the phenomenon would be able to recognise her experience); understanding (the richness and sensitivity of the final theory); generality (applicability or adaptability to a range of situations); and control (an indication of how people might be able to influence process or outcomes). Kearney comments that fitness and generality were judged to be stronger than understanding or control. It is unclear how exactly Kearney conducted this assessment, for example, whether in assessing 'fitness' she asked people who have experienced domestic violence whether they would recognise the concepts she derived, although she comments that the strong agreement across the studies appeared to indicate fitness.

Moving forward with grounded theory as a means of synthesis

Kearney's analysis represents one of the most theoretically grounded and robust applications of grounded theory to qualitative study synthesis, and it is clear that grounded theory offers a number of important advantages in attempting syntheses. It is a widely used and accepted approach in primary research, so that there is a cadre of trained researchers who would be able to utilise the approach. Grounded theory very helpfully deals with sampling as well as analysis issues and, particularly in the notion of theoretical saturation and theoretical sampling, offers a welcome means of limiting the number of papers that must be reviewed, rather than the exhaustive approach traditionally associated with quantitative systematic reviews. Grounded theory offers a means of generating higher-order categories and encourages reflexivity on the part of the reviewer as well as preserving the interpretive properties of the underlying data. The approach could also potentially deal with quantitative data by 'qualitising' them – converting quantitative data to qualitative form, for example through a narrative descriptive process. Glaser and Strauss (1967) themselves propose that quantitative data can be used to generate theory, although this is one of the most rarely used elements of their book.

Grounded theory does, however, have several disadvantages as a method for review. The limitations that Kearney herself identifies as characterising this approach to synthesis include the problem that the full details of the original participants are not fully available, because of the necessarily selective nature of research reports. The influences of the original investigators were often difficult to identify, and there was a danger of mixing experiences that were culturally or historically incomparable in ways that were not detectable.

These limitations are, however, likely to affect any attempt to synthesise qualitative studies.

For the reviewer seeking to synthesise both qualitative and quantitative evidence, grounded theory has a number of limitations, particularly if the aim is to incorporate the evidence within traditional systematic review paradigms. Although it is probably technically feasible to include quantitative data in a grounded theory review (though as yet there are few empirical demonstrations of this), quantitative forms of evidence are likely to be the 'poor relation' in any review of this kind. It is also not easy to see how grounded theory syntheses could be converted into quantitative form (eg as variables) or to derive estimates for use in a quantitative synthesis.

The concern of grounded theory with theory-building will not always suit the requirements of reviewers (or their funders). In some cases the synthesis of qualitative studies will be needed to perform lower-level tasks (such as identifying variables to be studied quantitatively), and here the arguments of the methodological purists might become rather tiresome. Moreover, even in the more highly proceduralised forms of grounded theory typified by the later writings of Strauss, grounded theory as an interpretive method, is inherently lacking in transparency and may therefore lack appeal to the systematic review community. Grounded theory also offers no advice on how to appraise studies for inclusion in a review.

Perhaps the most important problem with grounded theory is the methodological anarchy that characterises the area. This should not be underestimated as a barrier to the development of this approach as a means of synthesising primary studies. Those who claim that grounded theory is violated when researchers use concepts imported from the existing literature or disciplines to label categories identified in their own data (eg Wilson and Hutchinson, 1996) are unlikely to recognise the legitimacy of this approach when generating a synthesis of studies.

Meta-ethnography

Meta-ethnography is a set of techniques for synthesising qualitative studies. It was first proposed by Noblit and Hare (1988) and involves three major strategies.

- Reciprocal translational analysis. The key metaphors, themes or concepts in each study are identified. An attempt is then made to translate the metaphors, themes or concepts of other studies into each other. For example, one would assess if a concept of 'embarrassment' from

one study translates into the concept of 'stigma' from another. Judgements about the ability of a concept from one study to capture concepts from others are based on attributes of the themes themselves, including cogency, economy and scope. The concept that is 'most adequate' is chosen.

- Refutational synthesis. The key metaphors, themes or concepts in each study are identified and contradictions between the reports are characterised. The 'refutations' are examined and an attempt is made to explain them.
- Lines of argument synthesis. This involves building a general interpretation grounded in the findings of the separate studies.

Examples of meta-ethnography

Paterson *et al.* (1998) report what appears to be the largest meta-ethnography to date, involving the synthesis of 43 qualitative studies of the lived experience of diabetes. They describe using various criteria borrowed from primary research to assure the quality of their analysis, including aiming for trustworthiness by using multiple researchers, identifying deviant and negative cases, and evaluating rival hypotheses. Nelson (2002) describes using meta-ethnography to synthesise 12 studies on mothering other-than-normal children. Using reciprocal translational analysis, 13 common themes were identified, and further analysis yielded four more abstract themes. Beck (2002a) describes using meta-synthesis based on meta-ethnography techniques to synthesise the findings of six qualitative studies on mothering of multiples. Beck (2002b) also used meta-ethnography to synthesise 18 qualitative studies on postnatal depression. There is also a small body of papers that have used Noblit and Hare's proposals informally, without necessarily adopting the specific techniques or processes (eg Cook *et al.*, 2001).

Some authors have used meta-ethnography alongside other techniques within a review. Booth (2001) identifies the following systematic reviews, which he sees as being underpinned to varying degrees by meta-ethnography:

- A systematic review for the NHS Health Technology Assessment Programme on the use of modelling in the planning and prioritisation of clinical trials (Chilcott *et al.*, 2004)
- A systematic review for the NHS Health Technology Assessment Programme on the use of health-related quality-of-life measures in economic evaluation (Brazier *et al.*, 1999)
- A systematic review for the English National Board for Nursing, Midwifery and Health Visiting on the nursing

contribution to rehabilitation (Nolan *et al.*, 1997)

- A systematic review for the American Association for Retired Persons on coping with multiple chronic health conditions in older persons (Dowzer *et al.*, 2000).

Campbell *et al.* (2003) offer a detailed and well documented evaluation of meta-ethnography through a synthesis of papers on lay experiences of diabetes and diabetes care. They purposively sampled 10 papers, which varied in terms of country of study, ethnic groups studied, and qualitative approaches employed. The papers were appraised using an adapted form of the CASP checklist, and seven papers were judged suitable for inclusion in the review after this process. Lists of the papers' findings were generated using the authors' terms and concepts, and the 'key concepts' identified. A systematic search for the presence or absence of these concepts in all seven papers was then undertaken, in keeping with the requirements of reciprocal translational analysis. A higher-order analysis was then achieved through lines-of-argument synthesis, and allowed the identification and characterisation of what the authors call 'strategic non-compliance'. Third-order constructs and in particular the notion of achieving a balance were found to be very important in explaining people's experiences of diabetes, but were not evident in the primary studies.

Britten *et al.* (2002) offer a similarly well documented meta-ethnography of four papers on the meanings of medicines. In order to be explicit about how the concepts compared with one another, the authors created a grid into which they placed the concepts from each paper. The first four rows of the grid included relevant details of the study setting and research design, which the authors saw as essential contextual information for the synthesis. From the fifth row onwards each row of the grid represented a key concept. In labelling the rows, they aimed to use terminology that encompassed all the relevant concepts from each paper. The last row of the grid represented the main explanation or theory arising from each paper. In developing the grid, Britten and colleagues used Schutz's (1962) notion of first- and second-order constructs. Schutz used the notion of first-order construct to refer to the everyday understandings of ordinary people, and the term second-order construct to refer to the constructs of the social sciences. The explanations and theories in the last row of the grid were the second-order interpretations. Britten and colleagues developed what they call third-order interpretations by building on the explanations and interpretations of the constituent studies, which were at the same time consistent with the original results and also extended beyond them. They argue that these third-order interpretations justify the claim that

meta-ethnography achieves more than a traditional literature review, and also note that the process of qualitative synthesis cannot be reduced to a set of mechanistic tasks.

Moving forward with meta-ethnography

Meta-ethnography represents one of the few areas in which there is an active programme of funded methodological research for qualitative synthesis – Campbell and colleagues currently hold a Health Technology Assessment Programme grant to develop this area. It is thus a field of considerable potential and interest. It offers several advantages including, as Britten *et al.* (2002) note, its systematic approach. Despite the arguments of Estabrooks *et al.* (1994), meta-ethnography does appear to offer considerable potential for preserving the interpretive properties of primary data. Like grounded theory, it can potentially deal with quantitative data by treating it as a theme, although again there are no empirical examples, and no examples of meta-ethnography being used to derive quantitative information. However, a project funded by the NHS Service Delivery and Organisation on access to healthcare by vulnerable groups, led by Dixon-Woods and colleagues (www.prw.le.ac.uk/research/qualquan), is currently attempting to integrate both quantitative and qualitative evidence using meta-ethnography.

Estabrooks *et al.* (1994) have four major criticisms of meta-ethnography. They claim that meta-ethnography is a 'context-stripping' activity, not an interpretive one. Second, they suggest that it does not have an explicit goal of theory-building. Third, they feel that Noblit and Hare (1988) limited their discussion to ethnography. And fourth, they argue that Noblit and Hare reject the notion of knowledge as cumulative. These arguments are not necessarily sustainable, and appear to be undermined by the empirical demonstrations of meta-ethnography in practice. Jensen and Allen (1996) further argue that meta-ethnography should be used only to synthesise studies within a single paradigm. This could clearly present difficulties, especially when there are only a few qualitative studies in a particular area. However, there is no consensus on the epistemological and ontological debates underpinning Jensen and Allen's arguments, and pragmatically, as demonstrated by Campbell *et al.*'s (2003) work, many reviews will seek to integrate studies across different qualitative traditions.

Notwithstanding these theoretical debates, it is clear that there are a number of issues to be resolved if meta-ethnography is to develop in ways that are helpful and useful to reviewers. Campbell *et al.* (2003) point to the problem of determining the order in which papers should be synthesised,

as well as issues about searching for qualitative research and use of an appraisal tool. A further important problem with meta-ethnography is that, unlike grounded theory, it offers no guidance on sampling and is solely a means of synthesis. Meta-ethnography is a demanding and laborious process, and would benefit from the development of suitable software.

Estabrooks, Field and Morse's aggregation of findings approach

Estabrooks *et al.* (1994) describe what they call 'aggregation' of qualitative study findings. They suggest that aggregation should be limited to studies focused on similar populations or themes, using similar methods and similar means of categorising their data. The task of the synthesis, they suggest, is to compare and interpret across studies by undertaking critical activities such as the constant comparing and contrasting of themes from one study to the next. They suggest that such aggregation could generate mid-range theories – those that are moderately abstract and inclusive, identifying key variables and concepts, and helpful in informing practice.

Estabrooks and colleagues do not describe any set of techniques that might be used to undertake this form of analysis. They do, however, identify a number of problems with implementing their approach. First, the original researchers may have 'misinterpreted' their data or used a labelling convention that obscured the data. Second, there is the risk of weak studies being included in the aggregation. Third, there is a need for the analysis to be undertaken by an experienced researcher.

The aggregation of findings approach thus far represents a minor contribution to this area which appears to have enjoyed little uptake. It is likely to be overtaken by some of the approaches that are developing more rapidly.

Qualitative meta-analysis

Schreiber *et al.* (1997), drawing on Noblit and Hare's (1988) conceptualisation of meta-ethnography, define qualitative meta-analysis as aggregating a group of studies for the purpose of discovering the essential elements and translating the results into an end product that transforms the original results into a new conceptualisation. Schreiber and colleagues distinguish three different purposes of qualitative meta-analysis:

- Theory building, in which the aim of qualitative meta-analysis is the development of formal theory through the exploitation of data from diverse sources
- Theory explication, which would further elaborate abstract concepts
- Theoretical development, in which the synthesis of findings is thickly descriptive and comprehensive.

Schreiber *et al.* (1997) outline five principles by which a qualitative meta-analysis might be conducted:

- The reviewer should identify the purpose and likely outcome of the meta-analysis early in the process, although these decisions should not be binding
- The reviewer should adhere to the philosophical tenets of the naturalistic paradigm and make decisions congruent with it – in particular, context should not be sacrificed for commonality
- The reviewer should make an audit and decision trail, recording decisions about sampling, scope, and strengths and limitations
- The reviewer should maintain perspective about the purpose of the study and the intended relationship of the analysis to theory development
- The issues relating to context should be clarified for each study sampled.

Schreiber *et al.* offer little more than a general framework, and do not describe original procedures or techniques by which the analysis might be conducted. Perhaps the key problem with their approach is their appropriation of the term 'meta-analysis' to mean synthesis, when meta-analysis in fact refers to a very specific form of quantitative synthesis involving a family of quantitative synthesis methods. Again there appear to be few examples of the empirical application of this approach.

Qualitative meta-synthesis

'Meta-synthesis' appears to be a term used generically to describe the synthesis of primary qualitative studies, and has been used to refer to such techniques as meta-ethnography (Nelson, 2002) as well as less well defined approaches. Higginson *et al.* (2002) describe using qualitative meta-synthesis to synthesise studies on hospital-based palliative teams. By this they mean a synthesis of quantitative studies based on the themes, interventions and results of the primary studies. Standardised tables are used, into which data from studies are extracted. In effect, what they appear to mean by 'qualitative meta-synthesis' is a non-quantitative overview.

Jensen and Allen (1996), however, appear to mean something more specific in their discussion of qualitative meta-synthesis. Drawing on Noblit and Hare's (1988) meta-ethnography approach, they argue that interpretive synthesis involves the translation of the key metaphors of studies into each other, and suggest that the goal of meta-synthesis is interpretive, not integrative. Like Estabrooks *et al.* (1994), Jensen and Allen propose that in identifying studies for inclusion in a review, studies that use differing qualitative methods should not be combined. They also argue that studies should not be excluded on the basis of (lack of) scientific merit. Perhaps the most important and useful element of their work is their proposal that interpretive synthesis consists of hermeneutic and dialectic aspects. The hermeneutic aspects refer to the accurate portrayal of individual constructions, whereas the dialectic aspect refers to comparing and contrasting these individual constructions with the aim of generating a new construction on which there is consensus. The importance of this distinction is illustrated in Britten *et al.*'s (2002) meta-ethnography, where there was a need to preserve the second-order constructions as separate from third-order constructions.

In terms of implementing or proceduralising qualitative meta-synthesis, Jensen and Allen (1996) suggest that in the initial phase of reading the texts the findings of studies should be standardised via common codes, outlines and reporting formats, with attention given to the details of the individual accounts. The various studies within each group are then clustered together and initial assumptions made about the relationships between the studies. A list of the key metaphors, themes, etc of each study is created and juxtaposed with those of the other studies. Reciprocal translational analysis then follows, involving translation of the studies into one another, as outlined by Noblit and Hare (1988). The final phase involves synthesising these translations and making the clustered meta-matrices more refined.

Jensen and Allen (1996) propose that the rigour of meta-syntheses should be assessed using the criteria of credibility, auditability and fittingness (Guba and Lincoln, 1981). They argue that a meta-synthesis is credible when it represents descriptions of experience that people having that experience would recognise – a process similar to the respondent validation approach found in primary studies. 'Fittingness' is achieved when the findings can fit into contexts outside the studies. Perhaps most controversially, Jensen and Allen propose that an interpretive synthesis is auditable when the same or comparable conclusions can be reached by others,

given the data. The implicit notion of inter-rater reliability here would be unacceptable to qualitative researchers from some traditions.

Examples of studies using meta-synthesis

Sherwood (1997, 1999) uses 'meta-synthesis' to describe a process based on a combination of methods derived from meta-ethnography as well as other methods for qualitative synthesis (eg Estabrooks *et al.*, 1994), apparently ignoring the antagonism of Estabrooks and colleagues to meta-ethnography. This process she describes as involving six phases:

- Delineate focus of study. Sherwood suggests that the parameters for inclusion are set in relation to the study question and the underlying theoretical framework, while the inclusion criteria are based on the areas of commonality binding the studies together
- Locate studies meeting inclusion criteria
- Read and re-read all studies, give ratings for inclusion criteria and evaluate for scientific adequacy. Sherwood commends the use of criteria for appraising studies (somewhat in opposition to Jensen and Allen), and suggests that maintaining an audit trail of accepted and rejected studies will answer later questions about the decision-making process
- Determine how studies are related, examine for homogeneity across studies, and sort according to characteristics. A table summarising bibliographic information and other aspects of studies should be constructed, and studies grouped according to method to ensure groupings are compatible for analysis
- Translate studies into one another, merge data from across studies into a common data pool and study for common elements, and collapse data into essential patterns with explanatory themes through a process of analysis and synthesis. Here Noblit and Hare's (1988) methods for ethnography are deployed to translate studies into one another, and when the clusters of metaphors have become sufficiently refined to arrive at a description of the phenomenon common themes, specific descriptions, and quotations from the primary studies are used to elaborate on the meaning of each pattern and to provide clarity
- Express synthesis with a composite description narrating patterns and/or themes to explain what is taking place. Here a narrative similar to that found in primary qualitative research is presented, to report on findings and explicate themes.

Moving forward with meta-synthesis

Meta-synthesis, at least as outlined by Sherwood (1997, 1999), appears to offer some useful guidance on conducting syntheses, and emphasises the need for good practice (eg preserving an audit trail of decision-making about inclusion). Many of the processes appear to be soundly based on primary research methods. However, partly because of the varying uses to which the term 'meta-synthesis' has been put, qualitative meta-synthesis does not represent a coherent and consensual set of techniques. For example, questions remain about whether it is acceptable to apply quality criteria to select studies for inclusion in a review, and issues about how results from different disciplinary backgrounds should be combined remain largely unresolved. Again, meta-synthesis at this stage appears to represent a fairly minor and little-used contribution to the field.

Meta-study

Paterson *et al.* (2001), following on from Zhao (1991), use 'meta-study' as a term to encompass the overview of fields including theory, method and data. They argue that meta-study must involve analysis of the theoretical, methodological and contextual foundations and features of studies. A meta-study should include a temporal or historical context, and may consist of a series of synthesised studies. They propose a number of techniques, including the use of specific sampling techniques. These mirror those used in primary research, for example including the notion of searching for negative or disconfirming cases to challenge emerging theory.

Similarly to those who have described qualitative meta-synthesis, Paterson and colleagues try to provide a general framework and a clear set of techniques that can address the many varied concerns of reviewers, rather than focusing on one element (such as synthesis). They deal, for example, with issues relating to formulating the research question, study selection and appraisal. They suggest that if the number of potentially relevant studies identified is too large, the review may be limited by obtaining sufficient reports to permit comparisons among the selected dimensions and constructs to answer the question fully. This approach appears to draw on the notion of theoretical sampling in primary qualitative studies. Unlike some others writing in this area, they do commend excluding poor quality reports from the review, and also suggest that specific data from included studies can also be excluded – for example if the themes reported are not supported by the data presented.

In terms of synthesis, Paterson and colleagues propose (unlike some others) that studies representing the entire range of methodological approaches may be included in a synthesis, but they distinguish between meta-data synthesis, meta-method synthesis, and meta-theory synthesis. Meta-data synthesis refers to the synthesis of evidence presented in reports, and here they suggest that the choice of analytical approach is up to the reviewers. They suggest a number of techniques available to reviewers, including grounded theory, meta-ethnography, thematic analysis and interpretive descriptive analysis (by which they seem to mean a narrative critical review). Whatever approach is used, they recommend that studies be grouped according to characteristics such as disease, nature of the sample, and so on, and that each grouping should be treated as a case. They emphasise the contextual nature of primary research data, for example pointing to the excess of white, well educated participants in studies of diabetes, which they discovered in their own meta-ethnography of patients' experiences of this condition (Paterson *et al.*, 1998). They offer some useful guidance on how to manage the process of synthesis, suggesting that a combination of strategies is required including the use of qualitative analysis software supplemented by, for example, thesaurus, word-processed texts, diagrams and tables.

In approaching meta-method, Paterson and colleagues propose that the key concern is in identifying how the methods applied to an area of study shape our understanding of it. They suggest that a meta-method synthesis would involve specifying the methodological characteristics of reports and searching for patterns and contradictions that explain the findings of different methods (eg interview-based studies compared with ethnographies).

Meta-theory, on the other hand, involves a critical exploration of the theoretical frameworks that have provided direction to research, and synthesis would aim at understanding contrasts and similarities between different theoretical constructs (eg adaptation and coping as a way of understanding people's experiences of chronic illness).

Paterson and colleagues use the term 'meta-synthesis' to describe bringing together the ideas that have been deconstructed in the three meta-study processes. The primary goal of such an analysis is to develop mid-range theory grounded in qualitative research. They acknowledge that the processes they describe are laborious and time-consuming, and suggest that they can be done effectively only as a team effort.

Moving forward with meta-study

Meta-study, as described by Paterson and colleagues, provides a comprehensive framework with guidance on sampling, appraisal and synthesis. It helpfully distinguishes between different types of synthesis (meta-data, meta-method, meta-theory) and is largely systematic in its approach. Some well documented examples of meta-study have been published by Paterson and colleagues, so there are exemplars available. The approach is not as prescriptive as some others, for example in allowing considerable flexibility in the choice of method for synthesis.

Although it provides a very useful framework, little of meta-study is original in its conceptualisation, and it relies heavily on the rigour of the underlying methods. As discussed above, many of these methods, including meta-ethnography, are still in the early stages of development and evaluation, and there is some way to go before many of the methodological issues will be resolved. Meta-study also suffers from the limitation that it is not currently designed explicitly to cope with quantitative evidence.

Miles and Huberman's cross-case data analysis techniques

Miles and Huberman (1994), in their textbook on qualitative methods, offer a number of strategies for conducting cross-case analysis. These include meta-matrices for partitioning and clustering data in various ways, using summary tables based on content analysis, using case-ordered displays, and using time-ordered displays. Though Miles and Huberman discuss the secondary analysis of primary data using these techniques, they are readily transferable to the synthesis of study reports.

McNaughton (2000) describes using Miles and Huberman's data analysis techniques to conduct a synthesis of qualitative home-visiting research. McNaughton began by developing a start list of codes (Miles and Huberman, 1994) to produce broad categories for organising the data and to facilitate analysis. These initial codes were considered broad enough to organise the data into meaningful clusters for analysis, while avoiding the imposition of a conceptual or theoretical framework on the data. McNaughton describes analysing the 14 research reports she included in her study in chronological order, beginning with the oldest published report. Within-case analysis, referring to the task of examining each case in isolation from other cases (Miles and Huberman, 1994), was completed first. This process involved coding individual reports and writing a case summary. As coding progressed,

additional categories were added to accommodate and differentiate subthemes, although no new codes were needed after three or four reports had been analysed. The author kept an analytical journal to record memos on the thought processes involved in the analysis.

McNaughton then transcribed the findings sections of each research report and organised them, using a word-processing package, according to codes. The purpose of this step was to aggregate all data relating to the codes in one location. The analysis proceeded by examining the transcripts and summarising the content into major themes and concepts, noting them in the margins. A table was then constructed for each report in which small portions of text (phrases or paraphrases) were entered into columns labelled according to the start list of codes, with the aim of summarising and reducing the data in preparation for cross-case analysis. The data were then further reduced into a single table. The final reduction consisted of three or four conceptual descriptors for each category along with the 14 studies, yielding approximately 80 cells of data. Finally, McNaughton undertook cross-case analysis, consisting of noting commonalities and differences between the studies.

Moving forward with Miles and Huberman's data analysis techniques

Miles and Huberman (1994) offer a range of techniques for synthesising evidence from study reports. Their approach is highly systematic and has the advantage that it may be easier to teach and learn than some other approaches. The emphasis on data display assists in ensuring transparency, and the results of the synthesis are likely to be capable of being readily converted to variables that can be used quantitatively. Software is available that can fairly easily cope with this approach.

However, Miles and Huberman's emphasis on highly disciplined procedures has already made the approach unpopular among some of the qualitative research community, who see it as unnecessarily and inappropriately stifling interpretive processes and encouraging an emphasis on highly stylised themes, rather than theory. Miles and Huberman, because they did not have the notion of synthesis of study reports explicitly in mind, offer no advice on sampling or appraisal of the primary papers and, despite the affinity of the approach with quantitative study methods, it is not immediately clear how their techniques would deal with quantitative evidence because of the lack of 'richness' of quantitative data.

Content analysis

Content analysis is a technique for categorising data and determining the frequencies of these categories (Bryman, 2001). Content analysis differs from more qualitative methods in several ways. First, it is a requirement of content analysis that category specifications are sufficiently precise to allow multiple coders to achieve the same results when they code a collection of data. Second, it is firmly rooted in the quantitative traditions, relying as it does on the systematic and objective application of neutral rules (Bryman, 2001). Third, content analysis tends to draw on the same concepts of validity and reliability more usually found in the positivist sciences (Silverman, 2001).

Like Miles and Huberman's data analysis techniques, content analysis has developed as a way of conducting primary research. However, it is clear that it offers a means of synthesising study reports by allowing a systematic way of categorising and counting themes.

Example of content analysis for synthesis

Evans and Fitzgerald (2002) describe using content analysis in a systematic review of reasons for physically restraining patients and residents. The papers were not assessed for quality prior to inclusion. Content analysis was used to categorise the reasons cited for using physical restraint. First, the reports were read and re-read to develop initial impressions of the body of the literature. All reports were then grouped according to setting (acute or residential care) based on the description provided in the study methods. An abstraction tool was developed and pilot tested. This tool was used to extract the reasons for using restraints. It also allowed comparison of the abstracted data by the two reviewers. The two reviewers used the abstraction tool to create a list of reasons for each study. Initial categories for coding these lists of reasons were identified. The purpose of this coding system was to enable the reasons for restraint to be condensed into a smaller number of similar categories. The initial coding categories were pilot tested and revised as necessary by the two reviewers, using a random sample of the study reports. Two reviewers then coded the lists of reasons for each study. Reliability of the coding process was not formally assessed, but all differences in coding between reviewers were discussed until agreement was reached. The coded reasons were then listed by category (for example, treatment-oriented reasons for restraint). Coded reasons were tabulated according to

their setting: acute care or residential care. These categories were then used to present the results of the analysis. For example, the authors identify 'safety' as one of the categories, and show that this is cited in 92% of reports as a reason for using physical restraint.

Moving forward with content analysis

Content analysis is a well developed and widely used method in the social sciences, with an established pedigree. It is transparent in its processes and easily auditable. Software packages for undertaking the analysis are available, making the task of reviewers easier to manage. The synthesised evidence resulting from content analysis is relatively easy to present, generally in tabular form. Content analysis converts qualitative data into quantitative form, making it easier to manipulate within quantitative frameworks should this be appropriate.

There are, however, a number of disadvantages of content analysis that limit its usefulness as a means of synthesis. One is the problem mentioned above, that the term 'content analysis' may be a misused term, often used to refer to thematic analysis. More generally, content analysis has a number of weaknesses derived from its limitations as a primary research method. Content analysis is inherently reductive and tends to diminish complexity and context. It may be unlikely to preserve the interpretive properties of underlying qualitative evidence. The frequency-counting aspects of content analysis may mean the results do not reflect the structure or importance of the underlying phenomenon – in the example above 'safety' may have been the most cited reason, but that does not mean it was the most important reason. A related problem is the danger that absence of evidence (non-reporting) could be treated as evidence of absence (not important) and that, because content analysis deals better with manifest content than abstracted theory, the results may be oversimplified and may record what is easy to classify and count, rather than what is truly important.

A further problem with content analysis is that there is a need for caution over analyst perspective, in particular the micro-political orientation of the reviewer. Content analysis has traditionally had a role in media analysis, where it has been involved in, for example, exposing a range of biases in reporting (Henderson *et al.*, 2000). Clearly a different orientation might be required where the intention is to use content analysis to produce an integrative synthesis.

Case survey

The case survey method proposed by Yin and Heald (1975) is a formal process for systematically coding relevant data from a large number of qualitative cases for quantitative analysis, and could be compared with a primary research questionnaire survey method. The reader-analyst constructs a set of highly structured closed questions to enable the extraction of data from individual case studies. These data are converted to quantitative form subsequently used for statistical analysis. Yin and Heald see the assessment of study quality as important, using *a priori* methodological criteria to assess papers and removing those which fail to meet them from the set to be analysed.

Larsson (1993) further elaborated case survey as a means of quantitative analysis of multiple case studies. Individual cases are scored by multiple coders, with the results subjected to inter-coder reliability analysis. Jensen and Rodgers (2001) also describe an extension to the case survey method, proposing a method for 'cumulating' case studies. They advocate the use of meta-analytical schedules, or cross-case summary tables, in which the rows are case studies and the columns are variable-related findings or other study attributes (eg timeframe, research entity, case study design type, number and selection method for interviewees, threats to validity). The cell entries may be simple check marks indicating that a particular study supported a given variable relationship, or they may be brief summaries of findings on a given relationship, or a brief description of study attributes.

It is extremely difficult to find examples of the application of this method in the health field: to date, most published uses of the techniques have been in the area of organisational studies.

Moving forward with the case survey method

The case survey method has a number of strengths. Chief among these, perhaps, is its ability to synthesise both qualitative and quantitative evidence. It converts qualitative evidence into a quantitative form which makes it easy to manipulate within quantitative frameworks. The approach may be very useful in addressing questions relating to organisation or policy, where it may be possible to synthesise case studies from areas outside healthcare.

The approach also has a number of important limitations, however. It relies on having a sufficient number of cases to make quantitative analysis worthwhile. In much healthcare research there are unlikely to be sufficient numbers of

studies to make quantitative analysis meaningful. Statistical techniques used in case survey analysis typically lack power due to the limited number of observations, as each case study is treated as an observation. Many qualitative researchers would probably reject the description of their studies as 'cases'. In particular, they might be unhappy about their study being treated as one case. The case survey method would have difficulty in coping with the interpretive properties of qualitative data – contextual factors that might be important in explaining the features of particular cases will be stripped out as the data are reduced to quantitative form.

Yin and Heald (1975) themselves acknowledge that the case survey method is more suited to studies of outcomes than processes, and that it should be used in carefully selected situations.

Qualitative comparative analysis

Qualitative comparative analysis (QCA) was originally proposed by Ragin (1987), based on his view that the same outcome may be achieved by different combinations of conditions, and that causation must be understood in terms of necessary and sufficient conditions. The approach was suggested as a means of analysing complex causal connections using Boolean logic to explain pathways to a particular outcome. Ragin suggests that causal explanations:

'... must focus on how conditions combine in different settings to produce the same or different outcomes.

The identification of patterns of multiple conjunctural causation provides a basis for specifying, at a more abstract level, the underlying similarities responsible for similar outcomes and the underlying differences responsible for different outcomes.'

(Ragin, 1987, p. 49)

Ragin points out that qualitative studies tend to examine only a few cases at a time, but their analyses address many aspects of cases as well as examining how the different parts of a case fit together. QCA attempts to synthesise all cases in a data set so that cases with the same outcome are explained by a parsimonious model. The model contains one or more terms, each of which explains a number of cases with this outcome, while no cases with a different outcome are explained by any of the terms. All variables and outcomes are treated as categorical. Boolean methods of logical comparison represent each case as a combination of causal and outcome conditions.

These combinations can be compared with each other and subsequently logically simplified through a process of paired comparison.

The method uses a matrix known as a 'truth table' to show all logically possible combinations of the presence and absence of independent variables and the corresponding outcome variable. Data are coded into a binary format (1 or 0) according to whether or not they meet certain criteria. Searches are then undertaken to find actual cases in the data that match each possible combination of independent variables and the outcome. Where cases differ only in the presence of one variable, yet produce the same outcome, this variable is considered logically irrelevant and is therefore deleted. This Boolean minimisation process is continued until all logically inconsistent variables have been omitted. The Boolean minimisation procedure allows the elimination of many potential explanations so that the analyst is left with only a handful of combinations of variables to account for particular outcomes. The remaining set of variables represents a parsimonious and logically consistent model of the combination of variables associated with the outcome under study (Haggerty, 1992). Recent refinements of the method have included the introduction of 'fuzzy' logic, so that it is not necessary to dichotomise variables so precisely (Ragin, 2000). Cases are not viewed as collections of distinct variables, but as configurations of set memberships.

Examples of application of QCA

Melinder and Andersson (2001) report one of the few health-related studies using QCA. Their study explores possible structural determinants underlying injuries into two categories: those with a predominantly social genesis (eg suicide), and those largely related to technical and environmental hazards (eg traffic fatalities). QCA is complemented in Melinder and Andersson's study by consideration of correlations between variables. Although Melinder and Andersson's study is interesting, in fact it uses exclusively quantitative data.

Cress and Snow (2000), by contrast, do use QCA to analyse a range of types of data, and complement it with a number of other types of synthesis. Their review, looking at outcomes of mobilisation by the homeless, examined 15 case studies. Using QCA they produced a truth table that allowed them to show under what circumstances social movements could lead to a range of outcomes, including rights. Effectively this analysis showed which conditions must be satisfied in order to achieve specific outcomes.

Moving forward with qualitative comparative analysis

Qualitative comparative analysis has the advantage that it does not require as many cases as the case survey method. It can be used with previously conducted studies as well as with new studies, and thus encourages an evolutionary and integrative approach to knowledge creation. Ragin argues that his approach allows complex and multiple patterns of causation to be explored, and allows evaluation of competing explanations. It allows easy integration of both qualitative and quantitative forms of evidence, and is transparent and systematic.

However, like any approach that relies on converting qualitative evidence into quantitative form, qualitative researchers are likely to argue about the ontological and epistemological assumptions of QCA. It is mainly appropriate for when a causal pathway is sought, and may be ill suited to the more usual concerns of qualitative research, including the meanings that people give to their experiences. QCA appears to be designed primarily to deal with case studies, and may not cope well with the more usual form of qualitative study reports. It also suffers from a number of disadvantages, like those of the case survey method: it does not deal well with issues on which there may be no evidence, and the data may not reflect the underlying structures of the phenomenon.

Bayesian meta-analysis

Meta-analysis is the formal quantitative synthesis of data from multiple studies, in which evidence is pooled using statistical techniques. Bayesian meta-analysis has become the focus of increased interest and development in recent years because of its flexibility in handling data from diverse forms of study. From a Bayesian perspective, synthesis of research evidence is a decision-making process, and individuals who make decisions do not approach evidence in isolation: pre-existing beliefs, subjective judgements, and access to external sources of evidence will shape their interpretations. Bayesian statisticians express their belief about the size of an effect by specifying some prior probability distribution before seeing the data, and then they update that belief by deriving a new probability distribution, taking the data into account (Egger *et al.*, 1997). In the conventional terminology, Bayesian analysis begins with prior beliefs that are updated by data (summarised in the likelihood function) to form posterior conclusions. Bayes's theorem defines the process by which the prior beliefs, formally expressed as a probability distribution, is modified when combined with the current

evidence in the likelihood function. The resultant posterior distribution is thus a synthesis of the prior distribution and the data.

Example of Bayesian meta-analysis

The method proposed by Roberts *et al.* (2002) for synthesising data from both qualitative and quantitative primary studies used qualitative evidence as a source of external evidence to inform the development of the prior distribution. Qualitative evidence was therefore used in the role of a precursor, to help identify the relevant variables to include in the synthesis, but importantly it also contributes to the evidence itself by informing initial judgements about the likely effects of these variables. The prior distribution is then combined with evidence from quantitative studies to produce an overall synthesis, the results of which are expressed as a 'posterior distribution'.

Roberts *et al.* (2002) attempted to evaluate the feasibility of a Bayesian meta-analysis approach in identifying factors potentially affecting uptake of childhood immunisation in developed countries and in assessing the likely impact of each factor on levels of uptake. They noted that issues relating to searching for and evaluating qualitative papers remain to be resolved, and that the elicitation of prior beliefs is an underdeveloped area. They therefore made assumptions or took pragmatic steps at several stages in order to produce a complete analysis.

The final meta-analysis was based on 32 primarily quantitative studies and 11 primarily qualitative studies, with no exclusions on the grounds of quality. In order to generate the prior distribution for the meta-analysis, Roberts *et al.* combined subjective opinion with a synthesis of the qualitative studies across five different individuals. The aim of using five individuals was to represent a broad range of prior knowledge and expertise in the prior distribution. A content analysis approach was used to generate categories to represent the factors identified by this synthesis, producing a list of 11 factors which were ranked by each reviewer. Rankings for each factor were then combined across reviewers using a linear opinion-pooling method.

Data were extracted from quantitative studies using a predefined and prepiloted data extraction form developed for the review. The data were coded into the categories that had already been generated by content analysis of the qualitative data. However, where factors influencing uptake identified in the quantitative studies did not fit under this schema, new categories were generated. Odds of being immunised were

calculated for each factor that influenced uptake from the combined data from the quantitative studies.

It became clear during the data extraction that the qualitative and quantitative studies identified some factors in common, but other factors had been reported only in the qualitative or only in the quantitative studies. Only six of the 11 factors identified in the prior distribution had been investigated in the quantitative studies. Conversely two variables commonly studied in the quantitative literature had not been identified in the prior distribution.

Meta-regression methods were used to combine the prior probability (based on the qualitative evidence and subjective beliefs) with the quantitative data, to form a 'posterior probability' that each of the factors identified was important in determining uptake of immunisation.

The addition of the quantitative data modified some of the prior probabilities. For two of the factors, the posterior probabilities (which reflected both the quantitative and qualitative evidence) were much increased from the prior probabilities (which were based solely on the qualitative evidence). For two other factors, addition of the quantitative evidence resulted in a reduction in importance compared with that based on the qualitative evidence alone.

Moving forward with Bayesian meta-analysis

Bayesian meta-analysis allows the integration of qualitative and quantitative forms of evidence, and explicitly allows qualitative evidence to contribute to meta-analysis by identifying variables to be included and providing evidence about effect sizes. It therefore reflects important precedents from primary research, where qualitative research is often used to identify the variables of interest before conducting a qualitative study. It demonstrates the importance of synthesising qualitative studies in order to identify variables for investigation. It may therefore help to ensure meta-analyses properly reflect the diversity of evidence at primary level, and to show where quantitative data relevant to patients' concerns might be absent, by identifying where qualitative research has determined the importance of a theme which has been neglected in quantitative studies.

Although Bayesian meta-analysis is conceptually straightforward its implementation may prove difficult, causing problems with transparency. Because of its emphasis on converting qualitative data into quantitative form, it may

lack appeal to some sections of the 'qualitative community'. The techniques for achieving this form of analysis are still under development and many methodological issues remain to be resolved, including those relating to elicitation of prior beliefs and the impact of different methods of qualitative synthesis. Other approaches to Bayesian analysis, including the confidence profile method (Eddy, 1989; Eddy *et al.*, 1992), may also offer methods for synthesis which will need to be evaluated to assess their usefulness for combining qualitative and quantitative evidence.

Meta-needs assessment

Gaber (2000) proposes that quantitative and qualitative evidence can be combined for purposes of conducting meta-needs assessments. He proposes that the sequential steps in conducting a meta-needs assessment study are the same as conducting a meta-analysis. Problem formulation is identified as the first step of meta-analysis. This is followed by identification of the relevant literature and supporting documentation, very broadly defined. The second major step involves retrieval of the documents. Gaber suggests that if there are too many documents it may be necessary to refine the problem to narrow down the number of reports required. This contrasts with the notion of determining the point of theoretical saturation as a means of limiting the number of papers for inclusion.

The third step in meta-needs assessment, according to Gaber (2000), is the evaluation of collected data. He suggests undertaking a content analysis of each document to determine the primary needs by systematically identifying elements of a message. Finally, he suggests comparing and contrasting the varying research findings from the different research reports using a combination of two established systematic review techniques: narrative procedure and vote counting. The narrative procedure requires careful reading of each document after which the researcher constructs an overarching narrative based on the different accounts presented in the literature. Vote counting involves tabulating the distribution of different observations among the reviewed papers and establishing frequencies for them. These procedures provide a qualitative descriptive analysis of each reviewed document through the narrative account, while concurrently providing a quantitative ranking of observations through the vote-counting procedure.

Moving forward with meta-needs assessment

Meta-needs assessment provides a general framework combining elements of other approaches rather than a distinctive new approach in its own right. It is a recently published approach, and as yet there are few empirical examples of its application.

Discussion

Quantitising and qualitising

It is evident from the discussion above that synthesis of diverse forms of evidence will generally involve quantitising – conversion of qualitative data into quantitative form – or qualitising – conversion of quantitative data into qualitative form (Tashakkori and Tedlie, 1998). Once (aspects of) qualitative evidence have been converted into quantitative form, a range of more-or-less standard quantitative data analysis methods is available for application to the quantitative summaries. In some cases no further analysis is required, and the simplest of descriptive accounts suffices. For example content analysis may be reported in terms of percentages of studies (and/or subsets of participants therein) which cite particular terms. In principle, regression modelling of the variation of these percentages could be undertaken, but this is rarely helpful in this context unless the summary data set is very extensive.

Yin and Heald's (1975) case survey method yields cross-classified tables summarising the findings of each study, which may be presented descriptively and/or analysed by usual survey analysis methods, typically describing the frequency with which particular findings arose, and their inter-relationships and dependencies on characteristics of the studies and their samples. Some of Miles and Huberman's (1994) data analysis techniques might produce similar types of data capable of similar forms of analysis.

Where a quantitative summary of the results of each qualitative study is obtainable, and if an overall summary (pooled) estimate over all the studies is sought, standard meta-analysis techniques can be applied. These allow variations within the data set to be explored, for example variations by quality or other characteristics of each study. If results of quantitative studies are available in a compatible form, formal combination of the results from quantitative and qualitative studies will be possible through this approach.

However, such combinations may often be of doubtful validity, and qualitative results will usually be swamped by quantitative results. Other methods use quantitative methods more specific to their distinctive approach. For example, QCA yields a summary of presence or absence of combinations of explanatory and outcome variables which is subjected to Boolean analysis intended to identify minimum necessary sets of combinations.

While it is clear in principle that many of the approaches that involve quantitising qualitative data lend themselves easily to subsequent analysis, few approaches that have dealt with qualitative synthesis have dealt directly or explicitly with how to incorporate quantitative evidence. Issues of how to qualitise quantitative data, and how to deal with it within a qualitative synthesis, require further discussion and development.

Issues in qualitative synthesis

Is synthesis of qualitative evidence legitimate?

There is a body of argument against the synthesis of qualitative research. The argument rests primarily on the epistemological and ontological commitments that are assumed to underlie qualitative research, and in particular the inviolable relationship between researcher, data and interpretation. Qualitative researchers tend to emphasise the importance of their relationships with research participants and take account of the context in which data were collected (Fielding and Fielding, 2000), and these are usually inaccessible to the secondary analyst. Qualitative studies, it is therefore suggested, are unsuitable for 'summing up' (Sandelowski *et al.*, 1997). However, Barbour (1998) points out that qualitative researchers use contradictory evidence collected within a single method to explore the boundaries of emergent typologies or theoretical explanations, and that it is therefore acceptable to analyse data from different

sources together to focus on contradictions and exceptions in developing, expanding or reformulating theories.

Answering well defined questions

Estabrooks *et al.* (1994) argue that studies should be selected to focus on similar populations or themes. However, Jensen and Allen (1996) argue that the identification of the phenomenon of interest is not straightforward, and point out that in primary research the definitions of the phenomenon emerge from the data. Whether one should start with an *a priori* definition of the phenomenon for purposes of a secondary synthesis is therefore an important topic. A related issue is whether, where there is a need to limit the number of papers included in a review that includes qualitative evidence, the question should be very narrowly focused (eg in terms of participant characteristics), or whether some alternative strategy (such as the notion of theoretical sampling) should be used to limit the number of papers required.

Identifying papers for inclusion

The problems of searching electronic databases for qualitative research have been noted (Dixon-Woods and Fitzpatrick, 2001; Hawker *et al.*, 2002; Barroso *et al.*, 2003). Under an ESRC-funded project led by Dixon-Woods *et al.* (www.ccsr.ac.uk/methods) some of these problems are now being evaluated.

Much of the debate has focused on how to determine the number of papers that need to be included in a review. Schreiber *et al.* (1997) suggest drawing on the sampling techniques of primary qualitative research, arguing that the principles guiding the theoretical sampling technique depend on the research questions, the desired end product, and the ontological and epistemological framework of the analysis. In primary research, theoretical sampling is conducted with a view towards the evolving theoretical development of the concepts. A researcher continues sampling until theoretical saturation is reached – after each new interview or observation, no new relevant data seem to emerge regarding a category, either to extend or contradict it (Strauss and Corbin, 1990). Using this approach in reviewing evidence would suggest that an exhaustive sample of papers fitting the review criteria, as required under systematic review procedures, would potentially be redundant: the reviewer could stop reading and synthesising papers once satisfied that sufficient data existed for each category generated by the review. Thorne and Paterson (1998) similarly suggest that there must be a sufficient number of studies to answer the question and allow comparisons among selected dimensions

and constructs. Booth (2001) argues that this approach would be entirely consistent with the forms of sampling used in primary research, suggesting that 'If we think of the papers in a qualitative review as being "informants" then we seek to identify specific groups of papers that possess characteristics that are relevant to the phenomenon being studied. We aim to include a wide range of types of papers (ie reflecting as many of the themes or schools of thought as possible) and preferably to select "key informants" (ie papers that lead us to important sources of knowledge in the form of additional citations) ... Once a particular theme has been identified further occurrences of this theme are only of interest in strictly quantitative terms unless they expand on or modify an already-identified theme.' Booth further points out that quantitative researchers are also currently seeking to establish a law of diminishing returns beyond which there is little benefit in further searching. However, the application of this form of sampling has rarely been tested empirically, although Sandelowski *et al.* (1997) suggest that where there are more than 10 topically similar papers, it will be necessary to use a 'clearly defined purposeful sampling strategy in order to set tighter boundaries for the synthesis'. Jensen and Allen (1996) and Sherwood (1999), however, express anxiety that selective sampling may result in the omission of relevant data, thus limiting the understanding of the phenomenon and the context in which it occurs.

Does it matter if the studies used different methods or different theoretical approaches?

Barbour (1998) offers a vivid discussion of the distinctions, tensions and conflicts between the different approaches to qualitative inquiry. Some of the most intense debates in the area of qualitative synthesis have focused on how far it may be acceptable to synthesise qualitative studies derived from different traditions.

Estabrooks *et al.* (1994) argue that studies for synthesis should share a similar research approach. They give the example of a synthesis of studies that had addressed issues around women and pregnancy. In all but one study, which used a phenomenological approach, a grounded theory approach had been employed. Estabrooks and colleagues suggest that although similar themes could be identified across all studies, the mixing of methods led to difficulties in developing theory because of the major differences in the epistemological foundations of the methods. Jensen and Allen (1996) concur with this view, arguing that if studies from different methods are mixed, it is unclear what has been obtained and how it has been verified. Sherwood (1999)

similarly argues that studies with different methods should be grouped together to allow for comparable analysis.

Others have taken a more pragmatic approach, including Paterson *et al.* (2001) and Campbell *et al.* (2003). Paterson and colleagues deal with the problem of different approaches partly through their techniques of meta-method and meta-theory. They note in some detail the problem that qualitative study reports frequently mislabel or misreport the approach used: as noted above, grounded theory and content analysis are frequently applied to studies that are in fact simple thematic analyses or some other form of analysis.

Appraising papers

The issue of how, or whether, to appraise qualitative papers for inclusion in a review has received a great deal of attention. The NHS Centre for Reviews and Dissemination guidance (NHS CRD, 2001) emphasises the need for a structured approach to quality assessment for qualitative studies to be included in reviews, but also recognises the difficulties of achieving consensus on the criteria that might constitute quality standards. Estabrooks *et al.* (1994) argue that weak papers should be excluded 'just as a researcher may exclude a suspect interview in the process of secondary selection of informants'. How a weak or suspect study should be identified is not entirely clear, although they suggest that one criterion for selecting studies for inclusion in an aggregation concerns the reporting of the data, arguing that if the category labels in the studies are not close to the data then the aggregated analysis is more prone to error.

Sandelowski *et al.* (1997), on the other hand, argue that papers should not be excluded for reasons of quality, particularly where this might result in synthesisers discounting important studies for the sake of 'surface mistakes', and argue that synthesisers have to be 'connoisseurs' to be able to distinguish between these surface mistakes and mistakes so fatal that they invalidate the findings. Jensen and Allen (1996) appear to agree, suggesting that data 'germane to the purpose of the investigation' are likely to be excluded if studies are eliminated based on lack of scientific merit.

Several published qualitative syntheses have used quality criteria to judge the quality of papers for inclusion – although, perhaps unsurprisingly, not the same criteria. Thorne and Paterson (1998) used five criteria: explicit purpose; sampling and interpretation of findings guided by theoretical framework; clear inclusion and exclusion criteria; findings given alternative explanations; and sufficient detail to allow replication. Barroso and Powell-Cope (2000), in

their meta-synthesis of studies of living with HIV, use a grid to guide assessments along the following criteria: descriptive vividness; methodological congruence; analytical precision; theoretical connectedness; and heuristic relevance.

Other examples are given above of where reviewers have chosen either not to appraise the papers (eg Garcia *et al.*, 2002, page 15), or to appraise the papers using a formalised approach (eg Campbell *et al.*, 2003, page 18). Until the debate on quality appraisal moves closer to resolution, this is likely to be an area of continuing diversity.

Other issues

There has been surprisingly little discussion so far of the ethics of synthesis of qualitative evidence. Some issues that could be addressed include whether permission should be sought from the original authors; whether it is acceptable to use quotations from primary papers to illustrate third-order constructs; and whether respondent verification should be sought from the primary researchers or even the participants in primary studies.

Possibilities are beginning to develop for using software such as SPSS (www.spss.com) and NVivo (www.qsr.com.au) to handle data arising from multi-method research (eg Research Support Pty. Limited, www.researchsupport.com.au). However, as yet most software has limitations in dealing with both qualitative and quantitative forms of evidence.

Conclusions

This report offers an overview of a range of issues in using qualitative and quantitative evidence for informing policy and practice. We suggest that there is a strong rationale for multi-strategy research, with a growing number of examples of qualitative research being incorporated into traditional trial designs, often improving their sensitivity to the priorities of patients. There are persuasive arguments for integrating data generated by a range of strategies within reviews, although work is needed on the points at which it might be appropriate to pull diverse types of knowledge together. An important barrier to the deployment of qualitative evidence remains the continuing intense, and sometimes bitter, debates about how and whether the quality of qualitative research should be appraised. More constructive ways forward, informed by well grounded research, are now required. These may include assessments of the effects of using different quality criteria on decisions as to what is to be included in a review, and a move towards criteria that recognise the underlying diversity in qualitative research, rather than trying to impose a single unified framework on all qualitative research.

Finally, we offer a review of a range of strategies for synthesising qualitative and quantitative forms of evidence. Questions can be asked about how far apparent differences between the different strategies reflect relatively superficial differences in terminology, or the degree to which the method has been specified. At least some of these differences may reflect how the models have been developed in isolation, rather than more substantive or fundamental points of divergence. This raises the possibility that, certainly within the strategies, some of the methods themselves may be amenable to synthesis or at least some form of cross-fertilisation to develop more encompassing methods that incorporate the most useful element of each and allow for greater transparency. Synthesis of the methods is undoubtedly a more distant prospect, but even here there

may be points of possible convergence, for example the axial coding of grounded theory emphasises the importance of the relational properties of the data, as does qualitative comparative analysis with its emphasis on analysing the nature of the connections within the data. Work that compares the results of applying the different methods of synthesis will be useful in distinguishing the trivial and non-trivial points of divergence between the methods.

Decision-makers at all levels deserve to benefit from the evidence that can best help them to address the questions that confront them, whether this evidence be qualitative, quantitative, or both. For this reason it is important that the methodological and theoretical issues identified in this report should be addressed.

References

- Abbott, A. (1990) Conceptions of time and events in social science methods: causal and narrative approaches. *Historical Methods* 23: 140-50.
- Barbour, R. (1998) Mixing methods: quality assurance or qualitative quagmire? *Qualitative Health Research* 8: 352-61.
- Barbour, R. (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal* 322: 1115-7.
- Barroso, J. and Powell-Cope, G.M. (2000) Meta-synthesis of qualitative research on living with HIV infection. *Qualitative Health Research* 10: 340-53.
- Barroso, J., Gollop, C.J., Sandelowski, M., Meynell, J., Pearce, P.F. and Collins, L.J. (2003) The challenges of searching for and retrieving qualitative studies. *Western Journal of Nursing Research* 25: 153-78.
- Beck, C.T. (2002a) Mothering multiples: a meta-synthesis of qualitative research. *MCN: American Journal of Maternal/Child Nursing* 27: 214-21.
- Beck, C.T. (2002b) Post-partum depression: a meta-synthesis. *Qualitative Health Research* 12: 453-72.
- Berkwits, M. (1998) From practice to research: the case for criticism in an age of evidence. *Social Science and Medicine* 47: 1539-45.
- Booth, A. (2001) Cochrane or cock-eyed? How should we conduct systematic reviews of qualitative research? Presented at the Qualitative Evidence-based Practice Conference, *Taking a Critical Stance*, Coventry University, UK, May 14–16 2001. www.leeds.ac.uk/educol/documents/00001724.doc
- Braunholtz, D.A. (1999) A note on Zelen randomization: attitudes of parents participating in a neonatal clinical trial. *Controlled Clinical Trials* 20: 569-72.
- Brazier, J., Deverill, M., Green, C., Harper, R. and Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment* 3 (9): 1-164.
- Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M. and Pill, R. (2002) Using meta-ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research and Policy* 7: 209-15.
- Bryman, A. (1988) *Quantity and quality in social research*. London: Routledge.
- Bryman, A. (2001) *Social research methods*. Oxford: Oxford University Press.
- Bryman, A. and Burgess, R.G. (eds) (1994) *Analysing qualitative data*. London: Routledge.
- Butler, C., Rollnick, S., Pill, R., Maggs-Rapport, F. and Stott, N. (1998) Understanding the culture of prescribing: qualitative study of general practitioners' and patients' perceptions of antibiotics for sore throats. *British Medical Journal* 317: 637-42.
- Campbell, D.T. and Fiske, D. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81-105.
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. and Donovan, J. (2003) Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science and Medicine* 56: 671-84.
- Cesario, S., Morin, K. and Santa-Donato, A. (2002) Evaluating the level of evidence of qualitative research. *Journal of Obstetric Gynecological and Neonatal Nursing* 31: 708-14.
- Chapple, A. and Rogers, A. (1998) Explicit guidelines for quality research: a step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism? *Family Practice* 15: 556-61.
- Chilcott, J., Brennan, A. and Booth, A. (2003) The use of modelling in the planning and prioritisation of clinical Trials. *Health Technology Assessment* 7: 1-125.
- Cook, D.J., Meade, M.O. and Perry, A.G. (2001) Qualitative studies on the patient's experience of weaning from mechanical ventilation. *Chest* 120, Suppl. S: 469-73.
- Cox, K. (2000) Enhancing cancer clinical trial management: recommendations from a qualitative study of trial participants' experiences. *Psycho-Oncology* 9: 314-22.
- Cress, D. and Snow, D. (2000) The outcomes of homeless mobilization: the influence of organization, disruption, political mediation, and framing. *American Journal of Sociology* 105: 1063-104.
- Dixon, D. (1996). Unifying concepts in parents' experiences with health care providers. *Journal of Family Nursing* 2: 111-32.
- Dixon-Woods, M. and Fitzpatrick, R. (2001) Qualitative research in systematic reviews has established a place for itself. *British Medical Journal* 323: 765-6.

- Dixon-Woods, M., Fitzpatrick, R. and Roberts, K. (2001) Including qualitative research in systematic reviews: problems and opportunities. *Journal of Evaluation in Clinical Practice* 7: 125-33.
- Donovan, J., Mills, N., Smith, M., Brindle, L., Jacoby, A., Peters, T., Frankel, S., Neal, D. and Hamdy, F. (2002) Improving the design and conduct of trials by embedding them in qualitative research: ProctecT study. *British Medical Journal* 325: 766-70.
- Dowzer, C., Basford, L., Booth, A. and Poon, L.W. (2000) *Living with multiple health conditions: a systematic search and literature review*. Report for the American Association of Retired Persons. Sheffield: School of Health and Related Research (SchHARR), University of Sheffield.
- Duffy, M.E. (1987) Methodological triangulation: a vehicle for merging quantitative and qualitative research methods. *IMAGE, Journal of Nursing Scholarship* 19: 130-3.
- Eddy, D.M. (1989) The confidence profile method: a Bayesian method for assessing health technologies. *Operations Research* 37, 210-28.
- Eddy, D.M., Hasselblad, V. and Shachter, R. (1992) *Meta-analysis by the confidence profile method*. San Diego, CA: Academic Press.
- Egger, M., Davey Smith, G. and Phillips, A.N. (1997) Meta-analysis principles and procedures. *British Medical Journal* 315: 1533-7.
- Engel, J.D. and Kuzel, A.J. (1992) On the idea of what constitutes good qualitative inquiry. *Qualitative Health Research* 2: 504-10.
- Estabrooks, C.A., Field, P.A. and Morse, J.M. (1994) Aggregating qualitative findings: an approach to theory development. *Qualitative Health Research* 4: 503-11.
- Evans, D. and Fitzgerald, M. (2002) Reasons for physically restraining patients and residents: a systematic review and content analysis. *International Journal of Nursing Studies* 39: 735-43.
- Fairhurst, K. and Huby, G. (1998) From trial data to practical knowledge: qualitative study of how general practitioners have accessed and used evidence about statin drugs in their management of hypercholesterolaemia. *British Medical Journal* 317: 1130-4.
- Featherstone, K. and Donovan, J.L. (1998) Random allocation or allocation at random? Patients' perspectives of participation in a randomised controlled trial. *British Medical Journal* 317: 1177-80.
- Fingfeld, D.L. (1999) Courage as a process of pushing beyond the struggle. *Qualitative Health Research* 9: 803-14.
- Fielding, N.G. and Fielding, J.L. (2000) Resistance and adaptation to criminal identity: using secondary analysis to evaluate classic studies of crime and deviance. *Sociology* 34: 671-89.
- Fitzpatrick, R., Davey, C., Buxton, M. and Jones, D. (1998) Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 2:14.
- Forte, J.A. (1998) Power and role-taking: a review of theory, research, and practice. *Journal of Human Behaviour in the Social Environment* 1: 27-56.
- Gaber, J. (2000) Meta-needs assessment. *Evaluation and Program Planning* 23: 139-47.
- Garcia, J., Bricker, L., Henderson, J., Martin, M., Mugford, M., Nielson, J. and Roberts, T. (2002) Women's views of pregnancy ultrasound: a systematic review. *Birth* 29: 225-50.
- Glaser, B. (1978) *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.
- Glaser, B.G. and Strauss, A.L. (1967) *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter.
- Green, J. and Britten, N. (1998) Qualitative research and evidence based medicine. *British Medical Journal* 316: 1230-2.
- Greene, J.C., Caracelli, V.J. and Graham, W.F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis* 11, 255-74.
- Guba, E.G. and Lincoln, Y.A. (1981) *Effective evaluation*. San Francisco, CA: Jossey Bass.
- Guba, E.G. and Lincoln, Y.A. (1989) *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hammersley, M. (1996) The relationship between qualitative and quantitative research: paradigm loyalty versus methodological eclecticism. In: Richardson, J.T.E. (ed.) *Handbook of research methods for psychology and the social sciences*. Leicester: BPS Books.
- Hammersley, M. (1997) Qualitative data archiving: some reflections on its prospects and problems. *Sociology* 31: 131-42.
- Haggerty, T.R. (1992) Unravelling patterns of multiple conjunctural causation in comparative research: Ragin's qualitative comparative method. *Journal of Comparative Physical Education and Sport* 14:19-27.
- Harden, A., Weston, R. and Oakley, A. (1999) *A review of the effectiveness and appropriateness of peer-delivered health promotion for young people*. London: Social Science Research Unit, Institute of Education, University of London.
- Hawker, S., Payne, S., Kerr, C., Hardey, M. and Powell, J. (2002) Appraising the evidence: reviewing disparate evidence systematically. *Qualitative Health Research* 12: 1284-99.
- Heaton, J. (1998) Secondary analysis of qualitative data. *Social Research Update*, 22. www.soc.surrey.ac.uk/sru/SRU22.html
- Henderson, L., Kitzinger, J. and Green, J. (2000) Representing infant feeding: content analysis of British media portrayals of bottle feeding and breast feeding. *British Medical Journal* 321: 1196-8.
- Henwood, K. and Pidgeon, N. (1992) Qualitative research and psychological theorising. *British Journal of Psychology* 83: 97-111.
- Higginson, I.J., Finlay, I., Goodwin, D.M., Cook, A.M., Hood, K., Edwards, A.G., Douglas, H.R. and Norman, C.E. (2002) Do hospital-based palliative teams improve care for patients or families at the end of life? *Journal of Pain and Symptom Management* 23: 96-106.
- Jensen, L.A. and Allen, M.N. (1996). Meta-synthesis of qualitative findings. *Qualitative Health Research* 6: 553-60.

- Jensen, J.L. and Rodgers, R. (2001) Cumulating the intellectual gold of case study research. *Public Administration Review* 61: 236-46.
- Kearney, M. (2001) Enduring love: a grounded formal theory of women's experience of domestic violence. *Research in Nursing and Health* 24: 270-82.
- Koops, L. and Lindley, R.I. (2002) Thrombolysis for acute ischaemic stroke: consumer involvement in design of new randomized controlled trial. *British Medical Journal* 325: 415.
- Langley, C., Gray, S., Selley, S., Bowie, C. and Price, C. (2000) Clinicians' attitudes to recruitment to randomised trials in cancer care: a qualitative study. *Journal of Health Services Research and Policy* 5: 164-9.
- Larsson, R. (1993) Case survey methodology: quantitative analysis of patterns across case studies. *Academy of Management Journal* 36: 1515-46.
- Lincoln, E.G. and Guba, Y.S. (1985) *Naturalistic inquiry*. Beverley Hills, CA: Sage.
- Mays, N. and Pope, C. (2000) Qualitative research in healthcare: assessing quality in qualitative research. *British Medical Journal* 320: 50-2.
- Mays, N., Roberts, E. and Popay, J. (2001) Synthesising research evidence. In: Fulop, N., Allen, P., Clarke, A. and Black, N. (eds). *Studying the organization and delivery of health services*. London: Routledge.
- McNaughton, D.B. (2000) A synthesis of qualitative home visiting research. *Public Health Nursing* 17: 405-14.
- Melinder, K.A. and Andersson, R. (2001) The impact of structural factors on the injury rate in different European countries. *European Journal of Public Health* 11: 301-8.
- Miles, M.B. and Huberman, M. (1994) *Qualitative data analysis: an expanded sourcebook*. London: Sage.
- Neill, S.J. (2000) Acute childhood illness at home: the parents' perspective. *Journal of Advanced Nursing* 31: 821-32.
- Nelson, A.M. (2002) A meta-synthesis: mothering other than normal children. *Qualitative Health Research* 12: 515-30.
- NHS CRD (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*. CRD Report No. 4, 2nd edn. York: NHS Centre for Reviews and Dissemination.
- Noblit, G.W. and Hare, R.D. (1988) *Meta-ethnography: synthesising qualitative studies*. Newbury Park, CA: Sage.
- Nolan, M., Booth, A. and Nolan, J. (1997) *New directions in rehabilitation: exploring the nursing contribution*. Research Reports Series No. 6. London: English National Board for Nursing, Midwifery and Health Visiting.
- Paterson, B.L., Thorne, S. and Dewis, M. (1998) Adapting to and managing diabetes. *IMAGE, Journal of Nursing Scholarship* 30: 57-62.
- Paterson, B.L., Thorne, S.E., Canam, C. and Jillings, C. (2001) *Meta-study of qualitative health research*. Thousand Oaks, CA: Sage.
- Popay, J. and Williams, G. (1997) Qualitative research and evidence-based healthcare. *Journal of the Royal Society of Medicine* 91, Suppl. 35: 32-7.
- Popay, J., Rogers, A. and Williams, G. (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* 8: 341-51.
- Ragin, C.C. (1987) *The comparative method: moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.
- Ragin, C.C. (2000) *Fuzzy set social science*. Chicago: University of Chicago Press.
- Roberts, K.A., Dixon-Woods, M., Fitzpatrick, R., Abrams, K.R. and Jones, D.R. (2002) Factors affecting the uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence. *Lancet* 360: 1596-9.
- Sandelowski, M., Docherty, S. and Emden, C. (1997) Focus on qualitative methods. Qualitative meta-synthesis: issues and techniques. *Research in Nursing and Health* 20: 365-71.
- Schreiber, R., Crooks, D. and Stern, P.N. (1997) Qualitative meta-analysis. In: Morse, J.M. (ed.) *Completing a qualitative project: details and dialogue*. Thousand Oaks, CA: Sage.
- Schutz, A. (1962) *Collected papers*, Volume 1. The Hague: Nijhoff.
- Schwandt, T.A. (1996) Farewell to criteriology. *Qualitative Inquiry* 2: 58-72.
- Seale, C. and Silverman, D. (1997) Ensuring rigour in qualitative research. *European Journal of Public Health* 7: 379-84.
- Sherwood, G. (1997) Meta-synthesis of qualitative analyses of caring: defining a therapeutic model of nursing. *Advanced Practice Nursing Quarterly* 3: 32-42.
- Sherwood, G. (1999). Meta-synthesis: merging qualitative studies to develop nursing knowledge. *International Journal for Human Caring* 3: 37-42.
- Silverman, D. (2001) *Interpreting qualitative data: methods for analysing text, talk and interaction*. London: Sage.
- Snowdon, C., Garcia, J. and Elbourne, D. (1998) Reactions of participants to the results of a randomised controlled trial: exploratory study. *British Medical Journal* 317: 21-6.
- Snowdon, C., Elbourne, D. and Garcia, J. (1999) Zelen randomization: attitudes of parents participating in a neonatal clinical trial. *Controlled Clinical Trials* 20:149-71.
- Sparks, G., Craven, M.A. and Worth, C. (1994) Understanding differences between high and low childhood accident rates: the importance of qualitative data. *Journal of Public Health Medicine* 16: 439-46.
- Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2003) *Quality in qualitative evaluation: a framework for assessing research evidence*. London: Government Chief Social Researcher's Office.
- Stern, P.N. (1994) Eroding grounded theory. In: Morse, J.M. (ed.) *Critical issues in qualitative research methods*. Thousand Oaks, CA: Sage.

- Stiles, W.B. (1993) Quality control in qualitative research. *Clinical Psychology Review* 13: 593-618.
- Strauss, A.L. and Corbin, J. (1990) *Basics of qualitative research*. London: Sage.
- Strauss, A. and Corbin, J. (1998) *Basics of qualitative research: techniques and procedures for developing grounded theory*, 2nd edn. Thousand Oaks, CA: Sage.
- Tallon, D., Chard, J. and Dieppe, P. (2000) Relation between agendas of the research community and the research consumer. *Lancet* 355: 2037-40.
- Tashakkori, A. and Teddlie, C. (1998) *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Thorne, S. (1994) Secondary analysis: issues and implications. In: Morse, J.M. (ed.) *Issues in qualitative research methods*. London: Sage.
- Thorne, S.E. (1997) Phenomenological positivism and other problematic trends in health science research. *Qualitative Health Research* 7: 287-93.
- Thorne, S.E. and Paterson, B. (1998) Shifting images of chronic illness. *IMAGE: Journal of Nursing Scholarship* 30: 173-8.
- Thornton, H. and Dixon-Woods, M. (2002) Recruitment of women to trials (letter). *Lancet* 359: 164.
- West, J. and Oldfather, P. (1995) Pooled case comparison: an innovation for cross-case study. *Qualitative Inquiry* 1: 452-64.
- Wiersma, J. (1988) The press release: symbolic communication in life history interviewing. *Journal of Personality* 56: 205-38.
- Wilson, H.S. and Hutchinson, S.A. (1996) Methodologic mistakes in grounded theory. *Nursing Research* 45: 122-4.
- Yin, R.K. and Heald, K.A. (1975) Using the case survey method to analyse policy studies. *Administrative Science Quarterly* 20: 371-81.
- Zhao, S.Y. (1991) Metatheory, metamethod, meta-data-analysis – what, why, and how. *Sociological Perspectives* 34: 377-90.

Notes

