Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis

Andrew Booth, Chris Carroll, Irene Ilott, Lee Lan Low and Katy Cooper

Authors' Note: This study integrates methodological findings from two nationally funded projects from the National Institute for Health Research namely *the Contribution of nurses, midwives, and health visitors to protocol-based care and its variants* [Principal Investigator: Professor Malcolm Patterson] and *Chemoprevention of colorectal cancer: systematic review and economic evaluation* [Principal Investigator: Mr Paul Tappenden] and one from the Higher Education Academy, *Enhancing the student experience of workplace-based e-learning: a systematic review and best practice framework* [Principal Investigator: Mr Andrew Booth].

Correspondence concerning this article should be addressed to Andrew Booth, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, South Yorkshire UK S1 4DA, e-mail: a.booth@sheffield.ac.uk.

Andrew Booth, MSc is a Reader in Evidence Based Information Practice at the School of Health and Related Research, University of Sheffield. His research interests include qualitative evidence synthesis, systematic review methodology and evidence based practice.

Christopher Carroll, PhD is a Senior Lecturer at the School of Health and Related Research, University of Sheffield. His research interests include qualitative evidence synthesis and the development and evaluation of systematic review methods more generally.

Irene Ilott, PhD is CLAHRC Knowledge Translation Project Lead at Sheffield Teaching Hospitals NHS Foundation Trust. Her research interests include evidence based practice,

knowledge translation for sustaining change to benefit patient, staff and organizational outcomes, research capacity and workforce development.

Lee Lan Low, MA (Med Anthropology) is Research Officer at the Institute for Health Systems Research, Ministry of Health, Malaysia. Her research interests include qualitative research, health systems research methodology and knowledge translation

Katy Cooper, PhD is a Research Fellow at the School of Health and Related Research, University of Sheffield. Her research interests include systematic review methodology and reviewing of complex interventions.

Abstract

Actively seeking the disconfirming or deviant case is properly regarded as a hallmark of trustworthiness in primary qualitative research. The need to subject emergent theory to such testing is no less important within qualitative systematic reviews. There is, as yet, little available guidance on how to implement such strategies. Few researchers describe the practicalities of seeking the disconfirming case. We survey the methodological literature to gain a better understanding of how systematic reviews of qualitative research handle the disconfirming case. We reflect on our own experience from three recent qualitative evidence syntheses. We describe how reviewers might actively manufacture opportunities to identify discrepant or refutational findings. We conclude by outlining possible methods by which a team might integrate active seeking of a disconfirming case within the overall review process.

Keywords

Qualitative research includes "the considered selection of typical, deviant, critical, or otherwise exemplary information rich cases" (Patton, 1990, p. 169). This requirement is shared by analysis of qualitative primary data from informants and by secondary analyses such as qualitative evidence syntheses of multiple study reports (including the specific sub-category of qualitative systematic reviews). "Deviant cases" are particularly important in the context of synthesis because, as its etymology suggests, reviewers are naturally inclined to look for commonalities or similarities (Petticrew & Roberts, 2006):

> It is well known . . . that we select, evaluate, and remember information in a way that supports our individual preferences, we fail to look for evidence that disconfirms our pet hypotheses, and we cannot spot errors in our own reasoning (p. 130).

This quest for congruity, in preference to dissonance, is most clearly encapsulated in the averaging effect sought for quantitative studies through formal meta-analysis (Dong et al., 2008).

In primary research researchers identify the disconfirming case only after they have completed initial data collection and analysis. Identification of the disconfirming case involves selecting cases that serve as examples that do not fit emergent patterns (Patton, 1990). It requires strategies that complement and extend the team's initial sampling strategies. Such cases allow the research team to evaluate rival explanations thereby enabling them to understand and define the limitations of findings from their own research. By seeking out disconfirming cases researchers are able to develop a richer, more in depth understanding of a phenomenon thus lending credibility to the resultant research account. In comparison to the averaging effect so typical of quantitative syntheses, identification of the "exception to the

rule" requires that reviewers engage more extensively with the data to explore both complexity and contradictions.

Researchers often use the term deviant case, almost synonymously, as an alternative to the disconfirming case. Typically use of this phrase is associated with deviant case analysis by which researchers revisit and extend a theory to accommodate and interpret the findings from the deviant case.

In the context of an evidence synthesis the disconfirming case may relate to a specific subgroup of the population for whom an intervention that works on average, or for whom an intervention is broadly acceptable, does not achieve its expected impact. For example a universal telephone reminder system, which in theory should improve attendance at appointments, may not work for adolescents because they frequently change their mobile phone number. Alternatively the disconfirming case may represent a population for whom an intervention or programme works better than expected. So the same reminder system may prove particularly effective for a chronic population who are housebound and therefore always present to receive the telephone call. Such heterogeneity deserves explanation and investigation. At a theoretical level a particular aspect, previously considered significant, may no longer be considered important, perhaps because it is addressed implicitly by an intervention. A reviewer needs to explore and explain why this aspect, either through being completely absent or through being non-visible, does not figure in a published account. Of course such heterogeneity is not only present in differences between populations. It may equally appear in contextual or environmental differences or in preferences for different outcomes.

Although there is not an empirical base for recommendations regarding the optimal number of studies that a reviewer might include in a qualitative evidence synthesis one very important reason for limiting quantities of reports is to reduce the risk of the synthesis and subsequent analysis being considered too superficial (Paterson et al., 2001). Indeed qualitative evidence syntheses that incorporate large numbers of study reports are often characterized as descriptive and aggregative rather than analytical and interpretative (Dixon-Woods et al., 2006a). For interpretative reviews there is a danger, particularly when synthesizing large numbers of studies, that a reviewer is more likely to miss the opportunities presented for identification of the disconfirming case.

Published guidelines for systematic reviews occasionally acknowledge the importance of identifying the disconfirming or deviant case. Nevertheless such guidelines provide relatively little guidance on how a review team might achieve such a goal (Noyes et al., 2008). In this report we seek to address this methodological gap. We undertook a brief literature survey of published examples of qualitative evidence syntheses, identified relevant methodological articles and drew on three examples from nationally-funded research in the United Kingdom. The primary aim was to find methods to maximize the possibility of identifying disconfirmatory findings and then to propose a toolkit from which to select to proceduralize this process. Therefore the purpose of this article is to provide an overview of how the disconfirming case has been handled in the meta-synthesis literature and to inform discussion of this issue. By doing this we hope to help researchers to select appropriately from a range of options according to the specific objectives, purposes and constraints of an individual project. The intention is not to specify a single template by which reviewers should pursue such findings. Indeed within primary research, identification of a disconfirming case is a strategy that researchers often use in iterative

theory-building, theory-testing designs or data-driven research. Such a strategy is likely to prove of most value in evidence syntheses that share these overarching objectives. We outline fourteen such options (including strategies relating variously to the review team, the review methods, the use of theory, and the presentation of findings). We describe our experience from three UK-funded evidence synthesis projects, all with significant qualitative components. We conclude with observations on the usefulness of different procedures and the respective contribution made by each procedure.

Methods

The Disconfirming Case in Qualitative Evidence Syntheses: a literature survey

To gain a better understanding of how systematic reviews of qualitative research have handled the disconfirming case, we conducted a brief literature survey of published examples of qualitative evidence syntheses and relevant methodological articles. We accessed a register maintained by the Cochrane Collaboration Qualitative Research Methods Group of examples of qualitative systematic reviews and methodological discussions, housed on a reference management database. This database contains references to over 300 methodological papers, 250 confirmed examples of published qualitative syntheses, and an additional 1200 references to potential syntheses. In addition, given that the concepts being explored are not typically present in titles or abstracts of such references, a Google Desktop tool was used to search an associated collection of the full text of many hundreds of articles in Portable Document Format and as Microsoft documents, both on the hosting computer and across the wider World Wide Web. All searches were conducted between July 2009 and January 2010. The team identified a variety of relevant terms. Three of these terms used the terminology of primary research (*deviant case,*

disconfirming case. negative case analysis), and another was derived from a specific method of qualitative evidence synthesis (*refutational synthesis*).

Ten reviews reported identification of the disconfirming case (Paterson et al., 1998; Paterson et al., 1999; Thorne & Paterson, 1998; Thomas et al., 2003; Greenhalgh & Peacock, 2005; Dixon-Woods et al., 2006b; Sandelowski et al., 2007; Voils et al., 2008; O'Connell & Downe, 2009; Wong et al, 2010). Sixteen additional items discussed methodological considerations (Noblit & Hare, 1988; Jensen & Allen, 1996; Macintyre & Petticrew, 2000; Author, 2001; Barroso & Sandelowski, 2003; Jones, 2004; Lloyd Jones, 2004; Pawson et al., 2004; Walker & Avant, 2005; Dixon Woods et al., 2006a; May, 2006; Petticrew & Roberts, 2006; Downe, 2007; Weed, 2007; Downe, 2008; Noyes et al., 2008). A process of "snowballing" was used to follow up appropriate references (Papaioannou et al., 2010). Snowballing is an umbrella term, analogous to techniques used in primary interview research, whereby identified respondents become the starting point for identification of additional participants. In the specific context of a literature search snowballing refers to using a known relevant item of literature to identify articles that have been cited, those that cite the article, related articles or subject terms used by the article as a starting point for inquiry. Snowballing was particularly appropriate given that identification of the disconfirming case is not typically the focus for study reports and might only be mentioned incidentally or in passing. Consequently this approach cannot be formally designated as a systematic review, simply a comprehensive search for methodological research to illuminate the chosen case studies.

A single reviewer read and re-read included studies. The same reviewer extracted data relating to dissonance, refutation or the deviant case to a single document, categorising and then coding extracts according to whether they related to characteristics of the team, characteristics of the method or to specific objectives of a review.

General approaches to the disconfirming case.

The meta-ethnography approach to evidence synthesis (Noblit & Hare, 1988) is one of the few methods to give explicit attention to identification of incongruities and inconsistencies. Indeed this approach embodies refutational synthesis as one of three methods for exploration of themes within the data, alongside reciprocal translation and line of argument synthesis. Downe (2008) comments:

> Despite the frequency with which Noblit and Hare's 1988 work is cited, Noblit observed in 2004 that, while most published metasynthesis accounts describe reciprocal findings, few report on the refutational phase of the work. . . .(p. 7)

Writing within the meta-ethnography tradition, and as a precursor to development of the meta-study method, Paterson and colleagues (1998) acknowledge the potential value of diverse approaches to identifying and exploring dissonance in establishing the "trustworthiness" of the synthesis product:

> Meta-ethnography in which trustworthiness was achieved by using multiple researchers, identifying negative or disconfirming cases, and testing rival hypotheses. (p. 57)

They continue by observing that "negative or disconfirming cases were identified when the data did not unanimously support the conclusions of the researcher" (p. 59). Author (2001) identifies that one characteristic of systematic reviews of qualitative research is that "Particular attention is focused on negative or disconfirming cases. This adds to the richness of the insight that the review provides on the phenomenon of interest" (p. 2). Noticeably interpretative methods of synthesis, particularly those methods that trace their pedigree to grounded theory approaches, reflect a greater preoccupation with the disconfirming case (Weed, 2007):

Grounded theory approaches also emphasise the importance of searching for negative or disconfirming cases to challenge emergent analyses . . . and this is also incorporated into the meta-study approach. . . . Similarly, as the iterative process of meta-interpretation develops, theoretical sampling seeks not only to broaden and deepen the analysis, but also to challenge it through seeking alternative points of view and perspectives (p. 19).

Development of a wider choice of methodologies for qualitative evidence synthesis has been accompanied by increasing recognition of the value of identifying the disconfirming case. This stems, at least in part, from a need to demonstrate review findings that go beyond the obvious and that are nuanced to particular contingencies, settings or population subgroups. For example the increasingly popular technique of realist synthesis attempts to explore differences between the outcomes of apparently similar programmes. It seeks to achieve this by unpicking the components of each programme and, indeed, actively exploring dissonance in how, and with whom, they have been implemented (Pawson et al., 2004). Wong and colleagues (2010) describe a realist review where "we deliberately sought out disconfirming data - i.e., data that might refute our provisional candidate theories" (p. 3). May (2006) associates the active searching for deviant or disconfirming cases as characteristic of those formative analyses specifically required where a reviewer is assessing and evaluating complex interventions.

Identification of the disconfirming case may also occupy a place within the specific sub-genre of methodological reviews where vote-counting based approaches are manifestly inadequate (Author, 2001):

This is particularly the case in methodological reviews where a minority approach may have tremendous potential for development but be otherwise overshadowed by the

existence of a substantive corpus of work. This is analogous to the importance of

investigating sources of heterogeneity in quantitative meta-analyses. (p. 2)

Procedures for identifying the disconfirming case

Role of the Review Team

Probably the most accessible method for encouraging the active exploration of dissonance and

the subsequent incorporation of multiple perspectives is in the constitution of the review team.

For example Lloyd Jones (2004) identifies the value of involving a team of researchers

throughout the full range of review processes:

> Ideally, however, a meta-synthesis should be undertaken by a team of researchers: the
>
> application of multiple perspectives . . . may result in additional insights, and thus in a
>
> more complete interpretation of the subject of the review. (Lloyd Jones, 2004, p. 277)

Involvement of multiple team members does not necessarily maximize the prospects that the

team is able to identify the disconfirming case. First, constitution of a review team is often "self-

limiting". Implicitly such a team represents a group of individuals with whom a principal

investigator believes that they could work. The investigator does not typically select a team based

on their variability. A review team might thus share common values or beliefs, whether unspoken

or articulated. Second, the commissioner of a review might require that a review team represent

different disciplines or areas of methodological expertise rather than necessarily reflecting

genuine heterogeneity. In such circumstances the constitution of a team is not specifically a

mechanism to identify and explore differences in findings. Indeed a research team more typically

reflects diversity in the different views each member holds toward certain types of research or

study design (Dixon-Woods et al., 2006a).

Different reviewer backgrounds. Jones (2004) is explicit about the importance of involving

reviewers from different backgrounds:

> The team-building process begins by recruiting participants (two, three or more per team)
>
> from varying backgrounds (professionally as well as demographically) to be immersed in
>
> the selected literature, at times 'line by line' and hypothesise at each new revelation of
>
> dialogic material . . . (Jones, 2004, p. 103).

This description does not address how a review team harnesses such backgrounds nor how best to

deploy a review team to capitalize on such advantages. An additional issue that requires

exploration relates to syntheses of published reports in different languages. The use of

researchers with different first language backgrounds might enhance the richness of subsequent

theoretical thinking as well as assisting in identification of cultural and national nuances.

Reflexivity. The above requirement for different reviewer backgrounds is couched primarily in

quantitative terms (i.e., "professionally as well as demographically" (Jones, 2004, p. 103)) rather

than as the need to engage with different values or belief systems. There is increasing recognition

that the reflexivity of the review team is as important within qualitative evidence synthesis as it is

in primary qualitative research. However to date there are few examples where the team has,

collectively and individually, formally considered how the diverse values and beliefs that they

hold might enhance the richness and complexity of the subsequent analysis.

Team dynamics. Team composition also impacts on the "power dynamic" within the review

team. This phenomenon has only previously been considered in the context of research teams in

general (Gerstl-Pepin & Gunzenhauser, 2002) but is equally critical in review teams. Actual

examples include a proposed review where the aspiring lead reviewer held the joint roles of

senior staff member, subject expert and experienced reviewer. Where such a principal

investigator appears to "hold all the cards" it is difficult to envisage how a single junior co-

reviewer would be empowered to identify dissonance and then, more significantly, to articulate such observations to other team members.

If members of a review team are to harness a positive team dynamic they must pay explicit attention to the dynamics and internal "politics" within their team. All members of the team must be empowered to challenge, and actively resist, interpretations that do not fit comfortably with the data. Indeed team members should recognise such "fresh" insights as potential hypotheses. They should formally charge the original reviewer with re-examining the data specifically for evidence to either "prove" or "disprove" the validity of such insights.

Although the systematic review process conventionally incorporate interaction with steering or advisory groups or consultation with stakeholders it is not usually feasible to ask such groups to fulfil a formal role in looking for contradictions between different data sources. Indeed it is rare for members of steering groups, or other stakeholders, to have either the time or the inclination to look through the supporting data from a major review. It therefore seems likely that the use of supplementary expertise and resource from within the review team is the best route to improving the internal validity of the process. Such procedures should be disassociated from political imperatives to establish the credibility of the review or its subsequent external validity.

Qualitative synthesis does not share with quantitative synthesis the same formal requirement for two reviewers to select articles or extract data from included studies as a protection against bias. Conceivably peer review of a limited random sample of decisions and training and/or piloting of methods might meet the supplementary requirements for quality assurance. This could release this same level of reviewer resource to be harnessed more appropriately in enhancing the quality of analysis and interpretation by maximizing identification of the disconfirming case.

Legitimizing a culture of questioning. A review team must also consider the "culture" within which their review takes place. Typically junior members of a review team are those who

conduct the data extraction and preliminary analysis, for financial and practical reasons. Such junior team members might not be encouraged to identify "problem data" that does not seem to fit within the broad sweeping findings or recommendations of a review or within the prevailing "received wisdom" of topic experts (Macintyre & Petticrew, 2000). The time constraints inherent in the systematic review process provide little incentive for exploring diversions or cul de sacs from the main review findings. There are few incentives for spending less overtly productive time in the backwaters of the review topic. Review teams need to put into place methods and opportunities that legitimize the identification and exploration of incongruity. Perhaps a review team could assign a senior member of the team a formal role as a "champion of incongruity" encouraging the junior team members first to raise and then, with the team's agreement, to explore any contradictory findings and the implications of these for the review product.

Use of Methods for Identification of Studies

A review team typically has relatively little control of factors relating to team composition and how the team operates (as determined by resources, timescales and available expertise). However team members have much more flexibility in choosing methods that are more likely to lead to identification of the disconfirming case. This is particularly true when a team chooses methods for the sampling, identification and subsequent selection of studies for inclusion. Author (2001) recognizes that

> Literature searching for qualitative systematic reviews should exhibit the following characteristics: Identifying major "schools of thought" in a particular area whilst being alert to the identification of variants, minority views and dissenters. It is particularly important to identify negative or disconfirming cases . . .(p. 4)

Although the comprehensive search strategy remains a gold standard for quantitative systematic reviews there is increasing recognition that, for a qualitative evidence synthesis, it is more critical that a search strategy is selected to match the intended purpose of the review. Methods such as critical interpretive synthesis and realist synthesis emphasize identification of specific items of evidence to address an identified need. Other methods too may derive value from the diversity, rather than the comprehensiveness, of their sampling technique. Purposive or theoretical sampling, in a quest to achieve theoretical saturation, may increase the likelihood that reviewers retrieve disconfirming cases. Indeed recent guidance in the Cochrane Handbook (Noyes et al., 2008) acknowledges that:

> a more purposive sampling approach, aiming to provide a holistic interpretation of a phenomenon, where the extent of searching is driven by the need to reach theoretical saturation and the identification of the 'disconfirming case', may be more appropriate (p. 20.7)

Downe (2008) similarly describes having "adopted the techniques of theoretical saturation and of searching for disconfirming data, borrowed from grounded theory" (p. 6). Greenhalgh & Peacock (2005) operationalized their approach to meta-narrative as having "mapped 13 different research traditions, compared their conceptual and theoretical approaches, and synthesised the empirical evidence" (p. 1064). Clearly such search methods, including extensive citation tracking, recognize the relative importance of identifying similarities and differences between the different traditions, in preference to systematic identification of all available evidence.

Such considerations, exemplified by Barroso & Sandelowski (2003) within primary qualitative studies, move the model of searching away from *a priori* identification of all relevant research to more contingent iterative approaches to searching:

As the information value of cases depends on analytic goals emerging in the course of study, researchers can only anticipate— prior to entering the field of study—the kind of purposeful sampling they will conduct and the sample composition and size they will likely obtain . . . researchers may not have anticipated the selection of "deviant" cases until they reached a point in the data analysis where the inclusion of such contrasting cases became vital to the development and validation of a typology. (p. 388)

Indeed such considerations raise the possibility of including a subsequent searching phase to attempt systematically to identify disconfirming cases from the literature. Such searching might employ sampling strategies that aim to maximize diversity (e.g., literature from disciplines or schools of thought not hitherto included). Indeed introduction of an additional review phase that requires maximum variability sampling as a specific validation approach, following initial data analysis, might prove a useful addition to many types of evidence synthesis.

Differential Exposure to the Findings

Typically any reviewer faces the challenging task of moving between the minutiae of data extraction and a more "elevated" stance required for the subsequent synthesis and analysis. Few would argue against the inherent advantages of using someone who has immersed themselves in the data for a period of weeks, or even months, for the ensuing analysis. However it is important not to overlook the associated dangers posed by the cognitive biases of such a reviewer (Petticrew & Roberts, 2006). It could prove advantageous to bring in other members of the review team, with differing levels of prior exposure to the findings, specifically to challenge interim findings and, indeed to co-create new findings. Thorne and Paterson (1998) formalized such a procedure in their description of methods for their synthesis on models of chronic illness.

After recounting how "all researchers met to identify differences in their analyses and arrive at an agreement about the analysis" (p. 175) they describe two specific procedures:

> the research team identified negative or disconfirming cases in the research when the data did not unanimously support the conclusions of the researcher, . . . and the researchers tested rival hypotheses by seeking explanations, other than the initial researcher's original hypothesis, to describe the data . . . " (p. 175)

Multiple Readings

Scientists have observed that the human brain is likely to identify similarities ahead of differences (Petticrew & Roberts, 2006). A review team might therefore expect that exposure to multiple readings of study data, both directly from the source article and repackaged as data within data extraction forms, could help them to identify themes overlooked from a preliminary reading. Indeed O'Connell & Downe (2009) seem to recognize that refutational evidence (citing Noblit & Hare, 1988) is most likely to benefit from within team discussion and repeated reading "The emergent themes were discussed extensively and the studies were reread to consider any evidence that could be considered refutational" (p. 594). Of course there is no inherent virtue in simply re-reading the same data sources. A reviewer needs to employ diverse reading strategies to maximise the likelihood of identifying a finding that does not engage with the prevailing theory.

Different Methods of Analysis

Frequently concerns with the validity of a particular interpretation are addressed within primary studies through the use of *triangulation*. Opportunities to do this are more limited within evidence synthesis, where the reviewer is constrained by the limited data present in published reports. Nevertheless this might prove a useful supplementary strategy. For example O'Connell

& Downe (2009) report using a different type of data (namely observational data) to explore their synthesis:

> Of particular interest here was [Study A]; this, along with the oldest of the studies [Study B] contained observational data. These were particularly explored to disprove the emerging analysis or any prior reflexive assumptions. (p. 594)

In particular, reviewers might juxtapose process and outcome data to examine if factors that study participants believe critical to the success of an intervention are present or absent in studies that have achieved the most significant effects. Conversely reviewers might consider whether any study features, not necessarily identified initially from the qualitative data, are shared across the most effective or least effective studies. "Sibling" studies, our term for associated studies of different types that share a common temporal and geographical "heritage" (e.g., qualitative studies, process evaluations or economic evaluations conducted alongside randomized controlled trials or other outcome based studies) are particularly valuable in this context. As indicated above the benefits of examining studies for congruence do not lie solely in the identification of divergence or the discrepant case. A formal "gap analysis" might inform additional analysis, possible components for candidate interventions and the commissioning of future research (Grant & Booth, 2009). For example a research team from the EPPI-Centre at the Institute of Education in London examined the most effective interventions for promoting intake of fruit and vegetables. They identified an almost complete absence of "healthy messages" i.e., that "eating fruit or vegetables is good for you" (Thomas et al., 2003).

Analytic procedures. Review progress meetings and interim reports are structurally designed to focus on commonalities rather than exceptions. A structured report form might be used to counter this tendency as a mechanism for actively encouraging identification of the disconfirming case.

In the same way as articles in major biomedical journals articulate "What is already known on this topic" and "What this study adds" interim review reports might consider "What we expected to find from this review but have not" (i.e., the gap analysis) and "What we have found that we did not expect" (i.e., the unexpected finding). The review team might use regular team discussions to explore possible explanations for such findings. They might then identify actions needed to explore such deviations. Documentation of such issues, and the subsequent discussions, would contribute to another important attribute for a qualitative systematic review, namely reflexivity by the review team (Dixon-Woods et al., 2006a; Weed 2007). Such reflexivity is considerably enhanced where team members discuss and document a priori beliefs about the topic under review.

A review team should also consider the need to separate formally the synthesis and analysis stages of the review process. Frequently the review team focuses on bringing studies together into a new synthesized product. Subsequently the review team needs to test the robustness of the synthesis. However such a stage may fall victim to time pressures. We could draw an analogy with meta-analysis where production of a forest plot does not represent the endpoint for the process. Rather a meta-analyst uses a variety of techniques such as funnel plots, sensitivity analyses and subgroup analyses to examine the robustness of the synthetic product (Egger et al., 1997). Do analogous processes exist within qualitative evidence synthesis?

Qualitative sensitivity and subgroup analyses

Attention to analysis provides an opportunity to identify exceptions to the rule. It also allows a reviewer to explore and to explain such exceptions. A review team should identify circumstances where findings are only present in particular subgroups (cp., subgroup analysis). They should also identify whether any particular study has had a disproportionate influence on

the themes present in the final synthesis (cp., sensitivity analysis) (Mills et al., 2005). Would a model or framework still be complete if the team was to remove such a study and its findings from the synthesis? Downe (2007; 2008) suggests that such testing might be formative (i.e., iterative and ongoing), through a constant comparison approach:

> We have also adopted the techniques of theoretical saturation and of searching for disconfirming data, borrowed from grounded theory. As we analysed each additional study, we consciously checked if the findings extended or refuted the emerging line of argument synthesis. (p. 6)

Such an approach is undeniably a valuable way for members of a review to explore, and indeed sensitize themselves to, the data. It does not exclude the need for a more summative taking stock of particular groups of studies or groups of findings on completion of the preliminary synthesis. Voils et al. (2008) describe one variant of such an approach where the study finding, not the study report, becomes the unit of analysis:

> Accordingly, we next turned to vote counting, in which a significance level is set as a cutoff, and then each relationship is placed into one of three categories: positive (confirming), negative (disconfirming), or no relationship. The category with the greatest number is then assumed to provide the best estimate of the relationship. (p. 12)

Such an approach differs from others advanced in this article by attempting primarily to quantify dissonance rather than to explain it.

The methodology of qualitative evidence synthesis does not yet include methods for addressing likely publication bias, analogous to use of the funnel plot in quantitative synthesis. Nevertheless members of a review team should consider whether they have captured a rich and diverse range

of study features, preferably identified *a priori*, within their synthesis. In particular they need to consider this if the methodology they have employed utilizes theoretical, opportunistic or purposive sampling as an alternative to the comprehensive searching of formal systematic review. Have they identified studies from all the pre-specified disciplines? Have they excluded coverage of any ages or ethnic groups from their population of included studies? Do the findings only originate from those involved in the delivery of the service or intervention being evaluated? Are low resource and high resource settings represented in the sample of included studies? Few authors of published reviews give adequate attention to such issues or consider the implications for their review findings. Indeed it would be useful if authors of published qualitative evidence syntheses examined such issues in a formal and systematic manner within the Limitations sub-section of their published reports. In addition, where a review team has conducted a mixed method review they could seek to corroborate findings across both quantitative and qualitative included studies.

Testing of existing theories and frameworks

Obvious dangers are present where the purpose of a qualitative systematic review is to validate existing theory. A reviewer might be encouraged, by temperament, team culture or design of the documentation, to privilege squeezing of extracted data into an existing format or framework in preference to creating new dimensions or concepts. This approach may be particularly attractive given that any new concept could, in theory, necessitate re-coding of previously coded studies if not the actual revisiting of the source article. Under such circumstances reviewers might find it valuable to extract data against two or more competing theories or to compare theory-based with atheoretical interpretations of the data. At a technical level reviewers might find it more useful to extract entire articles to qualitative data analysis software to facilitate independent coding of the same data, including *de novo* coding (literally original coding from scratch) for emerging models

or theories, in preference to using a stable set of selective data extracts. Paterson and colleagues (1999) describe how hypothesis testing was essential to developing and refining theory whereby the analytic process:

> included developing hypotheses as the fieldwork progressed and testing these hypothetical relationships by means of further data collection and analysis in the search for confirming or disconfirming evidence to support or negate the emerging theory. (p. 791)

Dixon-Woods and colleagues (2006b) describe a similar process but emphasize that this presents challenges to subsequent reporting:

> much of the later sampling was directed at testing and purposively challenging the theory as we began to develop it. Again, such forms of searching and sampling do not lend themselves easily to reproducibility or indeed auditability" (p. 11).

The same authors conclude that:

> Testing whether the interpretations change in response to different findings will be an important focus for future research, which will also need to evaluate whether apparently disconfirming evidence is the result of methodological flaws or poses a genuine challenge to theory (p. 11).

Similar procedures are well prescribed in concept analysis, a longstanding area of qualitative literature review aimed at the specific generation of theory. Walker & Avant (2005) describe identifying a typical case and a disconfirming case - exactly as a reviewer might seek to do within

the context of interpretative qualitative evidence synthesis. Such methods could conceivably form the basis for a methodology for moving from concepts toward theory in a qualitative synthesis.

Methods for Presentation of Findings

This article focuses on identifying disconfirming findings during synthesis and analysis. Nevertheless similar considerations extend to the shaping process for the final synthetic product. For example some qualitative evidence syntheses provide a clear auditable process linking findings to their originating studies. Formative presentation of findings among members of the research team might provide an opportunity to assess the extent to which individual studies contribute to the synthesis, whether themes are present in multiple studies, whether particular findings are contradictory or whether particular studies are outliers. Such a situation is described by Sandelowski and colleagues (2007):

> we then arranged these abstracted findings to show their topical similarity and thematic diversity, and then referenced each finding with the report(s) from which it was derived. Arranging findings in this way has the advantage of revealing findings that are not there that might theoretically or logically have been expected (p. 107).

Although the authors readily acknowledge that they had not encountered a clear instance of this in their findings they do cite as an example:

> the finding that difficulty accepting HIV was a deterrent to adherence . . . This finding suggests an opposite: that acceptance of HIV would favor adherence." (p. 107)

Subsequently the team reflected on the significance of this process:

> Had we not actually found the one report with this finding, we could not have assumed it even though it might make sense to do so. Although theoretically possible findings may

be derived from actual empirical findings, they do not constitute actual findings in a metasummary. (Sandelowski et al., 2007, p. 107)

Mechanisms for identifying the disconfirming case must not simply be based on examination of what has been found. There are several reasons why it is helpful to introduce some external frame of reference. First, many synthesis methods claim that "the result is greater than the sum of its parts" (Noblit & Hare, 1988, p. 28). Such a stance maintains that bringing findings together might result in the generation of new constructs. Because such new constructs might be either overarching (i.e., aggregative) or manufactured (i.e., synthetic) they might map imperfectly to the originating studies. Indeed if a synthesis simply represents an aggregation or amalgamation of all its component themes it might yield little added value (or, at the very least, return on investment) beyond representing a map of the literature. We should neither expect all themes to contribute equally to the final synthesis nor any model or framework to explain adequately all identified themes. We should therefore seek to identify positively (i.e., not simply through omission) any themes that are not adequately explained by an emerging model or framework. Such a "best fit" approach contrasts with the more exacting demands of qualitative researchers such as Kidder (1981) who state that negative case analysis strives, through progressive interpretations, to produce theory that explains all cases without exception. However a best fit approach is accommodated by the pragmatism of those such as Lincoln & Guba (1985) who acknowledge that the process of thematic analysis always holds apparent exceptions.

A second point relates to the fact that, unlike informants in a primary study, findings from published research articles are not produced independently of one another (Dixon-Woods et al., 2006a). Indeed if a research team has followed good practice and conducted a good quality literature search at the beginning of their primary study, it is almost impossible to achieve such

independence. Perversely academic incentives, in terms of the receipt of funding, likelihood of publication and individual prestige, depend on identifying and reporting new findings or, at least, in providing a new perspective on existing findings. The influence of findings from other authors is typically constrained in a primary research report to the Introduction or Discussion sections of an article, and is correspondingly less visible in the Findings. An even more worrying form of publication bias is where an author attempts to improve the alleged originality of their study report by failing to acknowledge its precursors at all. Because of these various influences on publication a reviewer or reader must not assume that how a topic is represented in its published literature bears any relation to the true state of the topic. Academic incentives for reporting novelty should encourage an author to isolate and identify a disconfirming case in their individual primary study (as something new about which to report). An attendant danger is that an author might place disproportionate emphasis on findings not previously reported (cp., false positives).

Stakeholder or respondent validation

Another possible, and yet equally controversial, mechanism for identification of a disconfirming case when the findings of a synthesis are presented is stakeholder or respondent validation (Jensen and Allen, 1996). Principal concerns relate to the notion of the synthetic construct. If a stakeholder or respondent only recognizes a fragment of their truth in the perspective afforded by a meta-synthesis is it still their prerogative to challenge its findings? Should a reviewer privilege a respondent's individual interpretation of a phenomenon in preference to distilled findings from many such individual accounts? Critics of respondent validation within the context of synthesis focus on the difficulty of identifying an appropriate "respondent". Should the respondent perhaps be the individual researcher who has conducted one or more of the included primary research studies? If so, will such a researcher recognize the "truth" of a synthetic construct if they did not identify it when conducting their original data analysis? Respondent validation is only useful if

reviewers purposefully solicited authors for data from their studies that, those authors believe, specifically challenges a review's conclusions, rather than asking them to agree with or to endorse the findings from that review. Reviewers could then incorporate seeking of the disconfirming case within such formal consultation. Ultimately the synthesis team should retain control over analysis and interpretation of such supplementary data. At the same time reviewers must acknowledge that commissioners often want to be reassured that the findings from such a synthesis reflect the accurate perceptions of those working within a given topic area. Some form of respondent testing, rather than validation might, in fact, be a useful exercise even if it simply serves to anticipate potential objections to the findings or recommendations of the published evidence synthesis.

Findings from exemplar methods for identifying the disconfirming case

A brief survey of the literature illustrates many potential mechanisms to explore the disconfirming case. However few of these mechanisms have received formal evaluation. The next section presents three reviews that opportunistically explored combinations of the above strategies together with observations from such case studies.

The disconfirming case in a study of protocol based care.
This qualitative review examining the development and implementation of protocol-based care was conducted within a larger project examining the contribution of nurses, midwives and health visitors to such protocols (Patterson et al., 2008; Ilott et al., 2010). Two reviewers, separately and independently, extracted data for a common set of identified articles. One reviewer had become familiar with guidance presenting an idealized 12-step development process for producing protocols (NHS Modernisation Agency, 2003). Using this framework for coding allowed identification of the extent to which data from different studies corresponds with the official

guidance. Another reviewer used a grounded theory based approach without reference to the published NHS Modernisation Agency guidance. Both reviewers independently extracted data from the included studies with the former coding deductively using categories derived from the framework and the latter working inductively, using free-form categories assigned on the basis of data extracts. The review team used Joanna Briggs Institute's QARI software (version 2) for organising and compiling data extracts. The team harnessed this software, typically used for aggregative syntheses, in a more interpretive way, providing a ready audit trail for progress from findings to themes and from themes to constructs.

An extensive data report with findings from both reviewers was shared with a visiting scholar who was given several days to digest the findings prior to a review meeting. The two reviewers and the visiting scholar then met to bring together three complementary perspectives. The first perspective derived from the data-driven framework synthesis, the second from the data-driven grounded theory approach whereas the third represented a conceptual, holistic viewpoint. In this way, within the limited constraints of the review, the team maximized the possibility of identifying the disconfirming case.

As expected prominent themes were identified independently by all three reviewers. Reviewer One provided validation for the overarching framework produced by the government-approved model (an idealized account of how the process ought to be). Examination of the published accounts against this framework revealed an absence of unprompted mentions of patient participation in the protocol development process. In the absence of the published framework Reviewer Two had not observed the absence of patient involvement when performing the synthesis. The team considered that patient participation in protocol based care represented government-backed rhetoric rather than the realities of practice.

In addition the published framework portrayed protocol based care as a sequential process. By using a grounded theory approach separated from the published framework Reviewer Two was freed from the "shackles" of such a linear process. The individual steps or components were validated by this independent approach but there was significant variation in the sequencing of the process of protocol development. The second reviewer was also able to identify a much more fluid and iterative approach to the development process. Descriptions in the included reports differed in when, and indeed why, an evidence gathering process took place. Some authors used the literature search process as the starting point for production of their protocol, following prioritization of the topic. Others used the same process as a *post hoc* justification for decisions that they were planning to implement anyway. Single dimensional coding against the published framework as performed by Reviewer One would have missed this important temporal variation. In contrast, in the absence of a frame of reference, Reviewer Two had to identify both the steps of the process and their likely sequence. Furthermore Reviewer Two identified important contextual variables that might prove to be drivers for the protocols. A few protocols were instigated by top down initiatives, most represented a response to a local problem, whereas some could be interpreted as local adaptations of an existing evidence base. The team felt that these contextual differences might moderate the perceived success of each initiative. Such an observation had been masked from Reviewer One who had been directed towards the mechanics of the development process at the expense of the context.

Reviewer Three brought an additional perspective to the process. Stepping back from the detail of the study reports she was able to detect the presence of significant publication bias. She observed that all the included articles were individual case studies. As such they represented success stories, not evaluations, mostly described by those who were instrumental in implementation of the protocol itself. Such authors had few, if any, incentives for candour or

admission of failure, either in acknowledging limited success in reports written for local consumption or in publishing "warts and all" accounts in the peer-reviewed literature. The above account demonstrates that these complementary perspectives do not necessarily contribute to the data synthesis stage but may improve the subsequent analysis by allowing unexpected insights to emerge.

The disconfirming case in a study of attitudes to chemoprevention.
A review team undertook a qualitative evidence synthesis of attitudes toward the taking of potential agents for chemoprevention of colorectal cancer as part of a larger review and economic evaluation (Author, 2010). Two reviewers extracted data and then a third reviewer was required to look independently at the interpretations of the two reviewers. The team convened a meeting at which members could discuss the findings and offer alternative interpretations stimulated by use of an innovative "best fit" model (Author, 2011) as a framework for analysis. Unlike the previous case study, this framework was not purpose-specific but it was similar enough to the review topic to yield comparable insights (Huffman, 2002). The team identified inherent contradictions between the external framework and the chemoprevention data (i.e., the external framework was derived from a context involving young women, whereas the chemoprevention review included both sexes and older age groups). When examining data extracted from included studies against this frame of reference the review team asked: Is this concept likely to transfer to the context of our review? Might this concept be explained away by population, gender or intervention differences from the source study? Although the themes from the existing framework provided a useful starting point, the team could expand on, add to or reduce themes to reflect the chemoprevention data. Such a comparison offered rich opportunities to identify specific additional characteristics present in the older target population.

A second technique involved partial re-analysis of the data against an additional framework identified and suggested by Reviewer Three. The third reviewer examined data from included studies and identified a temporal dimension not captured by the original framework. The team hypothesized that people's attitudes to chemopreventive agents would differ according to whether they were receiving general population advice (precontemplation), were targeted with condition-specific information (contemplation) or were already taking the agents (action). The team reanalysed the data according to this well-recognized transtheoretical model of health behaviour change (Prochaska & Velicer, 1997). This model helped to make sense of otherwise conflicting findings from some of the studies. The fact that important temporal dimensions were initially omitted from both this and the preceding case study suggests that review teams might find it useful to include a prompt such as "Are significant temporal dimensions missing from the model?" whenever they try to validate any logic or process-based model.

Whether such procedures are successful or not depends heavily on the background and conceptual resources of the third reviewer. Such insights only carry the status of hypothesis generation. An additional hypothesis that attitudes might differ between those taking vitamins (associated with being "natural") and chemical substances such as aspirin (associated with being artificial despite its natural origins) received only limited support from the extracted data (Connor, 2004; Huber et al., 2004). In this case there was not enough evidence either to confirm or refute this tentative hypothesis. Nevertheless this independent observation yielded a useful recommendation for exploration and research. The above demonstrates that a review team might find useful lines of inquiry in existing frameworks and models, even when such models imperfectly represent the topic area under consideration.

The disconfirming case in a study of the student experience of e-learning.

A third review, conducted on a modest budget for the Higher Education Academy (Author, 2009a; 2009b, 2009c), nevertheless managed to manufacture opportunities for identification of dissonance. In this case multiple reviewers extracted the data for included studies, reporting the student experience of work-based e-learning, and then a single reviewer completed a preliminary thematic synthesis of data. Other members of the review team, experienced in delivery of e-learning reviewed the data extraction forms and, through a process of memo-ing, made independent observations on the data. The review team considered these notes at a single meeting alongside a preliminary taxonomy of themes. The team discussed the choice of terminology for the themes and sub-themes and the validity of the relationships between sub-themes. Subsequently they undertook minor reorganization of the taxonomy and its associated themes.

In this instance the team felt that the synthesis had not captured the full richness and complexity of the extracted data. A second reviewer revisited the data with a view to identifying potential contradictions or inconsistencies from the original taxonomy. Meanwhile the team validated the initial taxonomy by using it as the basis for an interview schedule with those who deliver work-based e-learning. Observations from these two approaches helped to shape a more sophisticated understanding of the relationship between themes, moving from a taxonomy toward an embryonic model. To cite one such example, work-based e-learners stated a preference for working through materials at their own pace, so that they could adapt their learning schedule to their individual work and personal circumstances. At the same time such e-learners expressed a desire for group interaction and regular feedback. Additional analysis of these themes revealed inbuilt tensions. It is clearly challenging to design a course where individuals work at their own pace and yet interact effectively as a group. It is equally challenging to offer flexibility for students to work at their own pace and yet to plan a programme of regular assessment and

feedback. Again the review team achieved a more nuanced interpretation of potentially

conflicting findings having manufactured opportunities to revisit the data and extend the analysis.

Developing a Disconfirming Case Identification Guide.

The three case studies described above, together with the supporting survey of the literature,

provide an initial basis for a guide to facilitate and document decisions related to identification of

the disconfirming case. Such a toolkit or checklist, from which to select appropriately according

to the scope and purpose of the review, the characteristics of the topic and its associated literature

and the resources (time, funding and so forth) available to the review team, might therefore

include:

- **Strategies relating to the review team:** role of the review team, different reviewer backgrounds, reflexivity, team dynamics and legitimizing a culture of questioning.

- **Strategies relating to the review methods:** use of methods for identification of studies, differential exposure to the findings, multiple readings, different methods of analysis, analytic procedures, qualitative sensitivity and subgroup analyses

- **Strategies relating to use of theory:** Testing of existing theories/frameworks

- **Strategies relating to presentation of findings:** Methods for presentation of findings, stakeholder or respondent validation

Discussion

The approaches described in the three case studies remain tentative and require formal

evaluation. Few reviewers report or evaluate techniques that they currently use when exploring

contradictions within their data.   In these three reviews we employed pragmatic and

opportunistic approaches, constrained by available resources and time limitations. The above

consideration simply records observations on the processes involved. We have not been able to

establish formally the impact of such methods on the overall quality and rigour of the review. Nevertheless it is possible to identify a few common requirements. First it is not sufficient to expect contradictory findings simply to emerge from conventional review processes. The structures and processes of systematic reviews implicitly target consensus and, at least where meta-analysis is concerned, move in the direction of an averaging effect. Instead a review team involved in qualitative evidence synthesis should seek to manufacture a dialectic (Walsh & Downe, 2005). Unlike quantitative reviews, where the role of the dual observer is primarily to counter bias, a second observer in qualitative synthesis could be employed to pursue persistently contradiction and disconfirmation. Such a quest requires that the review team use reflexivity more extensively and explicitly (as in the method of critical interpretive synthesis) acknowledging that findings from reviews are essentially subjective and open to interpretation (Dixon-Woods et al., 2006b):

> CIS demands constant reflexivity on the part of authors of reviews. Authors are charged with making conscientious and thorough searches, with making fair and appropriate selections of materials, with seeking disconfirming evidence and other challenges to the emergent theory, and with ensuring that the theory they generate is, while critically informed, plausible given the available evidence. (p. 9)

A review team might use numerous and diverse methods to create such a dialectic. Where resources permit they might utilize additional reviewers at judicious stages of analysis and interpretation. This is likely to prove particularly fruitful by selecting reviewers with varying levels of prior familiarity with the topic area and different personal inventories of theoretical and practical approaches to synthesis, analysis and interpretation. A review team might use literature searches to operationalize maximum variation sampling by accessing disciplines or schools of

thought that emphasize diversity and dissonance. Reviewers might juxtapose theoretical and atheoretical approaches to synthesis and compare and contrast their findings. They might use external models and frameworks to structure the initial data extraction or to provide an alternative overlay against which to reinterpret and reconcile contradictory findings. Above all review teams need to establish a culture where contradiction and discrepancy are legitimized and, indeed, actively encouraged.

Although the potential for such approaches is more limited where only a single reviewer is involved, for example where a doctoral student is undertaking a synthesis, many of the above techniques are still applicable. A supervisor or mentor may explicitly adopt a role that seeks to identify and explore dissonance, as embodied in the stance of a "devil's advocate". Opportunities to present interim findings to wider groups with diverse backgrounds can also be engineered to be more interactive and less affirmative. Optimally such dissonance should be handled as a property of the diversity of the literature and not as an inadequacy of the analysis or analyst.

Such a "toolkit" for handling contradictory data is most likely to be appropriate to meta-synthesis approaches with a realist orientation (e.g., thematic synthesis) that are intended to generate outcomes that are concrete and definitive. For example, it is likely to be appropriate to use some or most of these procedures where the intention is to generate a list of key dimensions of the phenomenon under study. In those syntheses that feature a more idealist and constructivist orientation the notion of contradictory data is not relevant. Such approaches, including meta-narrative and critical interpretive synthesis (CIS), are conceptual in focus, with the essence of the overall picture being what is considered significant. Nevertheless, even in such cases, the review team should engineer the presence of procedures to identify other types of contradiction and rival explanations. Under such circumstances it is perhaps more accurately to describe this as "the dissenting voice" as opposed to "the disconfirming case". The research community needs to

explore the potential usefulness of some of the individual procedures itemized above for these specific circumstances.

Conclusion

A brief consideration of these issues helps the reader to recognize how many processes codified in conventional systematic review methodology implicitly steer towards consensus. Typically two reviewers, with a shared understanding of the task in hand, review titles or abstracts or complete data extraction with a view to reaching such a consensus. Where such a consensus is not readily apparent the team might refer queries to a third person, acting as arbiter, charged with agreeing a final common position on the specific case or issue. To increase confidence in the process a review team might undertake inter-rater reliability (kappa) tests. The final account for the report of the systematic review thus represents a shared narrative to which all contributors are expected to subscribe. The outcome of such consensual methods contrasts with what might result if two reviewers were required to construct *separate* narratives from the same data sources and then to bring the two narratives together and place them in juxtaposition.

Within qualitative systematic review methodology there is increasing recognition that reviewers choose the type of synthesis based primarily on whether an established and largely accepted framework exists for the literature under consideration (Noyes et al., 2008). The emerging orthodoxy states that where concepts are "secure" it might be most appropriate for reviewers to use a framework or model based approach, for example using framework synthesis as a variant of framework analysis (Dixon-Woods et al., 2006a, 2006b). In contrast, where concepts are still undergoing development and refinement more interpretive approaches, such as those employed by meta-ethnography or grounded theory, are likely to be more appropriate (Dixon-Woods et al., 2006a, 2006b). Our experience suggests an additional consideration when a

review team is selecting the appropriate "methodological mix" – that is the need for reviewers to explore actively any diversity within the data. Certain scenarios, such as the need to challenge an official position or a theory that is seen to monopolize a particular field, might require reviewers to undertake separate framework and grounded theory based analyses and then to compare the findings from each resultant synthesis. Similarly independent use of two competing frameworks or models might result in a richer degree of interpretation, revealing perhaps the inadequacy of a particular model. Such procedures might even contribute to an optimal fusion of two or more models within a "meta-model" (Carroll et al., 2011).

In a multi-author reflection on the methodological agenda Noblit (Thorne et al, 2004) observes that:

> Refutation syntheses seemed to me to hold the most promise, because they invited us to consider that ideas were, indeed, contested via ethnographies, seemingly indisputable to me. Moreover, they invite considerable creative and critical talents to the work of synthesis, yet this form of synthesis is rare, to my knowledge (p. 1349).

However while other authors similarly affirm the benefits of seeking the disconfirming case few propose mechanisms by which a review team might achieve this in practice. We hope that the above discussion leads to more systematic use and evaluation of strategies for identifying dissonance within evidence synthesis through consideration of the constitution of the review team, use of contrasting methods for analysis, use of theoretical models as a catalyst for exploration, and clear and auditable methods for presentation of data. In continuing this quest many proposed solutions are likely to lie, not within the emerging discipline of qualitative evidence synthesis but rather, on the well-trod foothills of primary qualitative data analysis.

References

Author (2001).

Author (2009a).

Author (2009b).

Author (2009c).

Author (2010).

Author (2011).

Barroso, J., & Sandelowski, M. (2003). Sample reporting in qualitative studies of women with HIV infection. *Field Methods*, *15*(4), 386-404. doi: 10.1177/1525822X03257392

Connor, L.H. (2004). Relief, risk and renewal: mixed therapy regimens in an Australian suburb. *Social Science & Medicine*, *59,*1695–1705. doi: 10.1016/j.socscimed.2004.01.030

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D.R., Miller, T., Shaw, R.L., Smith, J., Sutton, A., & Young, B. (2006a). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, *6,* 27-44. doi: 10.1177/1468794106058867

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. & Sutton, A.J. (2006) Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups *BMC Medical Research Methodology*, *6,* 35. doi: 10.1186/1471-2288-6-35

Dong, N., Maynard, R., & Perez-Johnson, I. (2008) Averaging effect sizes within and across studies of interventions aimed at improving child outcomes. *Child Development Perspectives*, *2*(3), 187-197. doi: 10.1111/j.1750-8606.2008.00064.x

Downe, S., Simpson, L., & Trafford, K. (2007). Expert intrapartum maternity care: a

metasynthesis. *Journal of Advanced Nursing*, *57*(2), 127-140. doi: 10.1111/j.1365-

2648.2006.04079.x

Downe, S. (2008). Metasynthesis: a guide to knitting smoke. *Evidence Based Midwifery*, *6*(1), 4-

8. http://www.doctoralmidwiferysociety.org/Portals/c8d3e3f8-9c01-4bf5-abd9-

3fd6b4c510ae/Downe_S.pdf

Egger, M., Smith, G.D., & Phillips, A.N. (1997). Meta-analysis: principles and procedures. *BMJ*,

315(7121), 1533-1537. doi: 10.1136/bmj.315.7121.1533

Gerstl-Pepin, C. & Gunzenhauser, M. (2002). Collaborative team ethnography and the paradoxes

of interpretation. *Qualitative Studies in Education*, *15* (2), 137-154. doi:

10.1080/09518390110111884

Grant, M.J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and

associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91-108. doi:

10.1111/j.1471-1842.2009.00848.x

Greenhalgh, T. & Peacock, R. (2005). Effectiveness and efficiency of search methods in

systematic reviews of complex evidence: audit of primary sources. *BMJ*, *331,* 1064-1065.

doi: 10.1136/bmj.38636.593461.68

Huber, R., Koch, D., Beiser, I., Zschocke, I., & Luedtke, R. (2004). Experience and attitudes

towards CAM – a survey of internal and psychosomatic patients in a German university

hospital. *Alternative Therapies in Health & Medicine*, *10,* 32–36.

Huffman, S. (2002). *Can marketing of multiple vitamin/mineral supplements reach the poor?* The

Vitaldía Project, Bolivia. http://pdf.usaid.gov/pdf_docs/PNACS889.pdf

Author (2010) How do nurses, midwives and health visitors contribute to protocol-based care? A synthesis of the UK literature. *International Journal of Nursing Studies*, *47(*6), 770-780. doi: 10.1016/j.ijnurstu.2009.12.023

Jensen, L.A. & Allen, M.N. (1996). Meta-synthesis of qualitative findings. *Qualitative Health Research*, *6,* 553-560. doi: 10.1177/104973239600600407

Jones, K. (2004). Mission drift in qualitative research, or moving toward a systematic review of qualitative studies, moving back to a more systematic narrative review. *The Qualitative Report*, *9* (1), 95-112. http://www.nova.edu/ssss/QR/QR9-1/jones.pdf

Kidder, L.H. (1981). Qualitative research and quasi-experimental frameworks. In Scientific inquiry and the social sciences, edited by M. B. Brewer & B. E. Collins, 226–256. San Francisco: Jossey-Bass.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills: Sage.

Lloyd Jones, M. (2004) Application of systematic review methods to qualitative research: practical issues. *Journal of Advanced Nursing*, *48*(3), 271–278. doi: 10.1111/j.1365-2648.2004.03196.x

Macintyre, S., & Petticrew, M. (2000) Good intentions and received wisdom are not enough. *Journal of Epidemiology & Community Health*, *54*(11), 802-803. doi:10.1136/jech.54.11.802

May, C. (2006). A rational model for assessing and evaluating complex interventions in health care. *BMC Health Services Research*, *6,* 86. doi:10.1186/1472-6963-6-86

Mills, E., Jadad, A.R., Ross, C., & Wilson, K. (2005). Systematic review of qualitative studies exploring parental beliefs and attitudes toward childhood vaccination identifies common barriers to vaccination. *Journal of Clinical Epidemiology*, *58*(11), 1081-1088. doi:10.1016/j.jclinepi.2005.09.002

NHS Modernisation Agency/National Institute for Clinical Excellence. (2003) *Protocol-based care: underpinning improvement.* London: NSHMA/NICE. http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service _improvement_tools/protocol_based_care.html

Noblit, G.W, & Hare, R.D. (1988). *Meta-ethnography: synthesising qualitative studies* Sage: Thousand Oaks, CA. 88pp

Noyes, J., Popay, J., Pearson, A., Hannes, K., & Booth, A. (2008) Chapter 20: Qualitative research and Cochrane reviews. In: Higgins, J.P.T., & Green, S. (editors), Cochrane handbook for systematic reviews of interventions, version 5.0.1 (updated September 2008). The Cochrane Collaboration. Available from www.cochrane-handbook.org .

O'Connell, R., & Downe, S. (2009). A metasynthesis of midwives' experience of hospital practice in publicly funded settings: compliance, resistance and authenticity. *Health* (London), *13*(6), 589-609. doi: 10.1177/1363459308341439

Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010) Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal*, *27*(2), 114-22. doi: 10.1111/j.1471-1842.2009.00863.x

Paterson, B.L., Thorne, S., & Dewis, M. (1998). Adapting to and managing diabetes. *Image Journal of Nursing Scholarship*, *30*(1), 57-62. doi: 10.1111/j.1547-5069.1998.tb01237.x

Paterson, B., Thorne, S., Crawford, J., & Tarko, M. (1999). Living with diabetes as a transformational experience. *Qualitative Health Research*, *9*(6), 786-802. doi: 10.1177/104973299129122289

Paterson, B.L., Thorne, S.E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research. a practical guide to meta-analysis and meta-synthesis.* Thousand Oaks, CA: Sage Publications.

Patterson, M., Rick, J., Ilott, I., Lekka, C., Dixon, S., & Oluboyede, Y. (2008). The contribution

    of nurses, midwives and health visitors to protocol-based care and its variants, and the

    impact of their contribution on patient and staff outcomes, quality and costs of care.

    Report for the National Institute for Health Research Service Delivery and Organisation

    programme, June 2008. http://www.sdo.nihr.ac.uk/files/project/79-final-report.pdf

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. 2d ed. Newbury Park, CA:

    Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: an introduction.*

    Manchester: ESRC Research Methods Programme. Available at:

    http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf

Petticrew, M., & Roberts, H. (2006). How to appraise the studies: an introduction to assessing

    study quality. In: *Systematic reviews in the social sciences: a practical guide.* Oxford:

    Blackwell; 2006.p. 130

Prochaska, J., & Velicer, W. (1997). The transtheoretical model of health behavior change.

    American Journal of Health Promotion, *12,* 38–48. doi: 10.4278/0890-1171-12.1.38

Sandelowski, M., Barroso, J., & Voils, C.I. (2007). Using qualitative metasummary to synthesize

    qualitative and quantitative descriptive findings. *Research in Nursing & Health*, *30*(1),

    99-111. doi: 10.1002/nur.20176

Thomas, J., Sutcliffe, K., Harden, A., Oakley, A., Oliver, S., Rees, R., Brunton G, & Kavanagh J.

    (2003). *Children and healthy eating: A systematic review of barriers and facilitators.*

    London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of

    London.

    http://eppi.ioe.ac.uk/EPPIWebContent/hp/reports/healthy_eating02/Final_report_web.pdf

Thorne, S., & Paterson, B. (1998). Shifting images of chronic illness. *Image Journal of Nursing Scholarship*, *30* (2), 173–178. doi: 10.1111/j.1547-5069.1998.tb01275.x

Thorne, S., Jensen, L., Kearney, M.H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, *14* (10), 1342-65. doi: 10.1177/1049732304269888

Voils, C.I., Sandelowski, M., Barroso, J. & Hasselblad, V. (2008). Making sense of qualitative and quantitative findings in mixed research synthesis studies. *Field Methods*, *20,* 3-25. doi: 10.1177/1525822X07307463

Walker, L.O. & Avant, K.C. (2005) *Strategies for theory construction in nursing,* 4th edn. Pearson Prentice Hall, Upper Saddle River, NJ.

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*, *50*(2), 204-211. doi: 10.1111/j.1365-2648.2005.03380.x

Weed, M. (2007). A potential method for the interpretive synthesis of qualitative research: issues in the development of 'meta-interpretation'. *International Journal of Social Research Methodology*, *11*(1), 13-28. doi: 10.1080/13645570701401222

Wong, G., Greenhalgh, T., & Pawson, R. (2010). Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Medical Education*, *10,* 12. doi: 10.1186/1472-6920-10-12