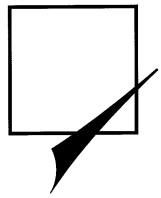
Evaluation

DOI: 10.1177/1356389008097871
Vol 15(1): 047–071



Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews

Effectiveness of Interventions to Promote Smoke Alarm Ownership and Function

MARK RODGERS AND ANGELA SOWDEN
University of York, UK

MARK PETTICREW
University of Glasgow, UK

LISA ARAI AND HELEN ROBERTS
City University, UK

NICKY BRITTEN
Universities of Exeter and Plymouth, UK

JENNIE POPAY
Lancaster University, UK

The objective was to assess the impact of new guidance on the conduct of narrative synthesis in systematic reviews of effectiveness, by means of a blinded comparison of guidance-led narrative synthesis against a meta-analysis of the same study data. The conclusions of the two syntheses were broadly similar. However, differences between the approaches meant that conclusions about the impact of moderators of effect appeared stronger when derived from the meta-analysis, whereas implications for future research appeared more extensive when derived from the narrative synthesis. These findings emphasize that a rigorously conducted narrative synthesis can add meaning and value to the findings of meta-analysis. The guidance framework provided a useful vehicle for structuring a narrative synthesis and increasing transparency and rigour of the process. While there may be risks with overinterpretation of study data, the framework, tools and

techniques described in the guidance appear to increase the transparency and reproducibility of narrative synthesis.

KEYWORDS: guidance; narrative synthesis; research methodology; systematic reviews

Introduction

Systematic reviews have become well established as a means to help inform policies and decisions about the organization and delivery of health and social care. However, making sense of large bodies of evidence drawn from research which uses a range of methods is a challenge. Ensuring that the product of this synthesis process can be trusted is important for policy-makers, for practitioners and for the people research is intended to benefit. There are a number of ways in which research evidence can be brought together to give an overall picture of current knowledge that can be used to inform policy and practice decisions (Dixon-Woods et al., 2005). These methods vary in their objectives and in the forms of evidence to which they can be applied. For instance, Bayesian meta-analysis can be used to incorporate quantified beliefs about effects of variables with evidence from quantitative research (Jones et al., 2003) and realist synthesis can be used to refine theory applied to a particular context using diverse forms of evidence from different contexts (Pawson et al., 2004).

A commonly used method to synthesize research in the context of systematic reviews is that of *narrative synthesis*, a defining characteristic of which is the adoption of a narrative (as opposed to statistical) summary of the findings of studies to the process of synthesis. This may occur either alongside or instead of statistical meta-analysis and does not exclude other numerical analyses. However, usually for reasons of study heterogeneity, most systematic reviews do not actually undertake a formal meta-analysis alone. Instead, they frequently combine the included studies in the way that seems most feasible or appropriate and label the product a 'narrative synthesis'. Consequently, there is currently no consensus on the constituent elements of narrative synthesis and the conditions for establishing a systematic and transparent approach to the synthesis process – with safeguards in place to avoid bias resulting from the undue emphasis on one study relative to another – are frequently absent. This should not be confused with approaches such as meta-narrative mapping which specifically aims to construct a structurally and thematically coherent story to describe a complex body of evidence (Greenhalgh, 2004).

We were funded to develop guidance that would help improve the quality of narrative synthesis in systematic reviews (see Box 1) (Arai et al., 2007; Popay et al., 2006). Like more formally defined methods, narrative synthesis can be used to incorporate diverse forms of evidence within a systematic review, as we have demonstrated elsewhere (Arai et al., 2007). However, the focus here is on the narrative synthesis of effectiveness data where statistical meta-analysis is not possible or advisable

Box 1. Developing the Narrative Synthesis Guidance

We aimed to develop guidance on the conduct of narrative synthesis in systematic reviews, with a particular focus on increasing the transparency and reproducibility of the process.

Identifying Relevant Methodological Literature

We conducted a systematic search of the methodological literature in order to identify relevant articles. These were used to (a) identify common generic elements of the synthesis process and/or (b) identify discrete tools and techniques for the management, manipulation and presentation of data in the narrative synthesis context.

Creating the guidance document

A generic framework was developed to characterize narrative synthesis:

- developing a thery of how the intervention works, why and for whom;
- developing a preliminary synthesis;
- exploring relationships within and between studies;
- assessing the robustness of the synthesis product.

Though each of these elements is essential to the narrative synthesis, they do not necessarily occur sequentially or independently. A number of 'tools and techniques' identified from the methodological literature were placed within the appropriate elements of the framework. Precisely which tools and techniques are used in any given synthesis is likely to vary, depending upon the data being synthesized. Reviewers can choose any tools or techniques they consider appropriate, so long as these decisions are clearly justified and documented.

Objectives

The aim of this article is to demonstrate the way in which narrative synthesis guidance can be used in the context of a review of effectiveness, and to evaluate what the guidance might add (or otherwise) to the findings of a systematic review. In order to do this, we undertook a narrative synthesis of effectiveness studies that had previously been synthesized using meta-analytic methods, and compared the findings/conclusions of the two approaches. The narrative synthesis was carried out according to the recommendations outlined in the guidance.

Our aims are largely methodological. For this reason, the demonstration did not involve all stages of a systematic review, but focused entirely on the synthesis of available data and the comparison of meta-analytic and guidance-led narrative approaches.

Methods

The review selected for comparison was a Cochrane systematic review investigating the effects of interventions for promoting smoke alarm ownership and

function (DiGuiseppi and Higgins, 2001). This review was selected because it was considered methodologically sound, had incorporated a meta-analysis and had analysed a 'manageable' number of primary studies (11 RCTs).

Copies of all original primary studies included in the meta-analysis were obtained and a new narrative synthesis was carried out. The reviewer conducting the narrative synthesis (MR) was blinded to all results and conclusions of the Cochrane review.

For each of the four elements we outline in the guidance framework (see Box 1), a range of related 'tools and techniques' could be applied. The tools and/or techniques that appeared to be useful and relevant to the synthesis described here were selected and applied to the smoke alarm effectiveness studies. The reasons for selecting or rejecting a tool or technique are described. Where tools or techniques proved to be less useful, this is discussed and was used to inform and further strengthen the guidance. A flow chart summarizing the synthesis process is presented in Figure 1.

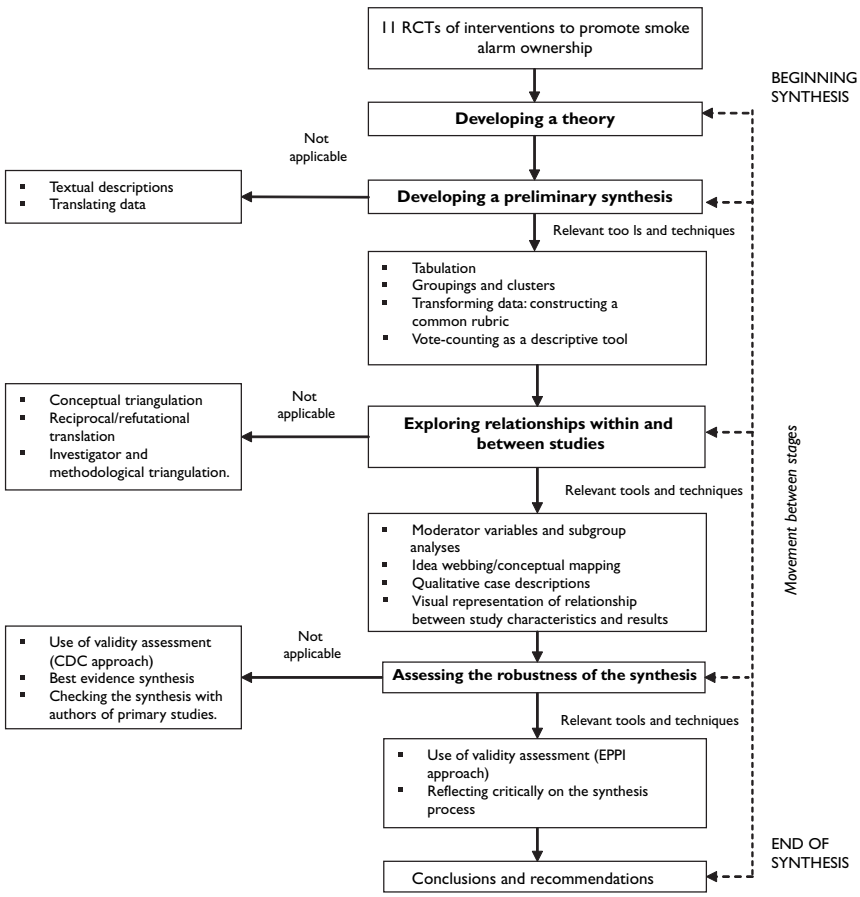


Figure 1. Synthesis Process

Element 1: Developing a Theory

The majority of studies reported on interventions which aimed to increase smoke alarm ownership and function through the use of educational interventions with or without the addition of free or discounted smoke alarms for participants.

In general, reports of the primary studies did not clearly describe the theoretical basis of the evaluated interventions, but the implicit theory underlying most educational interventions was that education can increase recipients' knowledge of potential fire/burns risks, change risk perceptions and lead to behaviour change (i.e. acquisition of smoke alarms). The use of discounted or free smoke alarms as an intervention to increase ownership and function (usually in lower income families) suggests that authors consider cost to be a barrier to smoke alarm acquisition. The determinants appear to have been characterized at the macro level, rather than being targeted at individual behaviour change.

Element 2: Developing a Preliminary Synthesis

The guidance suggests that 'how a reviewer approaches the preliminary synthesis . . . will depend in part on whether the evidence to be synthesised is quantitative, qualitative or both' (Popay et al., 2006). In the case of the review of smoke alarms, the data to be synthesized were anticipated to be predominantly quantitative and, more specifically, derived entirely from RCTs. With this in mind, each of the tools and techniques presented in the 'preliminary synthesis' chapter were assessed as to whether they would be relevant for the synthesis at hand (see Table 1).

Consequently, five of the six tools/techniques described in the guidance were applied to the synthesis and were carried out in the order described here.

Tabulating the data The most natural starting point for the synthesis was to extract data from the primary studies in tabular form. Data were extracted on participants, interventions, outcomes, country of origin, duration and provider of the intervention, number of participants in each group, context in which intervention was delivered, results and comments (see Table 2 for an example).

Study validity/quality was not discussed in detail in this section of the guidance. However, it seemed sensible at this stage of the narrative synthesis (where the papers were being read in detail and some broad judgements about their content were starting to be made) to consider study quality. Consequently, a column including data on methods/quality was included in the table and structured comments were included regarding individual papers, based on Jadad et al.'s scale for evaluating RCTs (Jadad et al., 1996).

It became apparent from the table that the majority of studies were concerned with child safety, and that most included some measure of smoke alarm ownership/function as a main outcome. Only two studies reported on injury rates as an outcome, but neither of these presented separate data on fire/smoke/burn-related injuries (Kelly et al., 1987; Kendrick et al., 1999).

Textual descriptions It was not immediately clear what these might add to the synthesis, over and above the information presented in the tables. At this stage this seemed like an unnecessary duplication of effort, though 'textual descriptions'

Table 1. Selection of Tools and Techniques in Developing A Preliminary Synthesis

Name of tool/technique	*Comments in relation to current synthesis*	*Should this tool/technique be applied here?*
Tabulation	Describe study characteristics and results. Will quality be If so, will this be according to predefined criteria?	Yes
Textual descriptions	Need to determine which aspects of each study will be drawn from the reports. These might be very similar to the table headings.	Yes, but not necessarily as a first step.
Groupings and clusters	Where possible, organize studies by intervention type, context, target population, study design, or outcomes.	Yes.
Vote-counting as a descriptive tool	Would be possible here if all data are converted to odds ratios/relative risks/mean differences.	Yes.
Transforming data: constructing a common rubric	Odds ratios or relative risks for dichotomous data, weighted or standardized mean difference for continuous data.	Yes.
Translating data (integrating themes and concepts reported across studies)	Inappropriate given predominantly quantitative data and the focus on effectiveness outcomes.	No.

might actually be useful for describing the interventions in more depth than can be given in the tables. Consequently, the use of this technique was delayed until the next stage of the synthesis process (exploring relationships).

Groupings and clusters The data extraction tables were examined to determine the presence of dominant groups or clusters of characteristics, by which the subsequent synthesis could be organized. The most obvious difference between studies in terms of the populations included is that all the studies deal with children and/or their families, with the exception of a study by Ploeg that included only participants aged 65+ years. This study was therefore excluded from later comparisons. Second, studies could be clearly be grouped according to which of the four smoke alarm ownership and function outcomes (specified a priori in the Cochrane review) they reported.

Developing a common rubric As mentioned previously, data were only available for the four smoke alarm ownership/function outcomes. As these data were dichotomous, odds ratios and relative risks were calculated. Absolute risk differences and percentage smoke alarm ownership in the control group

Table 2. Characteristics of Included Studies (Truncated)

Intervention	Participants	Setting/context	Outcomes	Results	Methods/quality
Barone (1988), USA. *Content:* I: Usual safety education, plus slides and handouts on burn prevention, motor vehicle safety education and video; bath water thermometer; hot water gauge (n = 41). C: Usual safety education (n = 38). *Duration*: 4 x 2h weekly meetings. *Delivered by*: Unclear	Couples or individuals attending 'Parenting the Toddler' classes	Classes at suburban, hospital, family homes	Home inspection 6 months after class	1) Final smoke alarm ownership I = 32/34 C = 26/29 2) Final functioning smoke alarms I =39/41 C = 34/38 I = 32/34 C = 26/29 No significant difference between groups	Allocation by coin toss within paired classes. Outcome assessment not blinded. Withdrawals: 27% of parents attending randomized classes did not enrol in trial.
Clamp (1998) UK. *Content:* I: Safety advice, leaflets, discount safety devices for low income families (*n* = 83 families). C: Routine child health surveillance and routine consultations without intervention (*n* = 82 families). *Duration:* Unclear *Delivered by:* Health visitors/practice nurses	Families of children <5 yrs on GP list	Delivered during child health surveillance consultations, opportunistically during other consultations, or the family was asked to make an appointment specifically for the intervention.	Telephone/ mail survey 6 weeks after visit	1) Smoke alarms acquired I = 8/83 C = 0/82 2) Functioning smoke alarms acquired I = 7/83 C = 4/82 3) Final smoke alarm ownership I = 82/83 C = 71/82 4) Final functioning smoke alarms I = 80/83, C = 71/82	Allocation by random numbers table, 1–165, the first 83 numbers on the list were the intervention group. Allocation was by a researcher blinded to the number given to each family. Outcome assessment not blinded. Withdrawals: none.

I = Intervention group, C = Control group

were also calculated for each smoke alarm ownership outcome and tabulated (see Table 3 for an example for the 'final smoke alarm ownership' outcome).

These tables showed that the effects of most interventions were generally quite small for most smoke alarm ownership and function outcomes (absolute differences ranged from 0 to 12.4 percent). However, they generally favoured intervention over control (only two of the ten studies that measured final smoke alarm ownership were negative for this outcome and one of the four studies reported a very small negative finding (absolute difference –0.1 percent) for 'smoke alarms acquired'. Smoke alarm ownership in the control groups of each study was generally quite high, with one clear exception (Kelly et al., 1987: 11 percent). This approach proved a useful first step – even in a narrative synthesis like this – in comparing the effects observed across the included studies.

Vote counting as a descriptive tool Although vote counting is generally seen as bad practice when used to summarize evidence (Hedges and Olkin, 1985), it may be a useful descriptive tool. In the current example, tables showing two approaches to vote counting were developed: (i) using ticks where the effect of the intervention was positive and statistically significant (see Table 4); (ii) using colours (superimposed on the rows of the table) to grade both the direction and statistical significance of each outcome (see Table 3, where symbols replace colours for technical reasons).

In terms of the vote count there were no differences between the relative risks and odds ratios calculated previously. The study by Williams reported that there was 'no statistically significant difference' between the experimental and control groups but did not provide data to calculate the measures in this table (Williams,

Table 3. Final Smoke Alarm Ownership (Common Rubric and Vote Count)

Reference	Absolute difference *(%)*	Relative risk (95% CI)	Odds ratio (95% CI)	Vote count RR	Vote count OR	% smoke *alarm ownership* in control group	
Barone (1988)	4.5	1.05 (0.90, 1.22)	1.85 (0.29, 11.89)			90	+
Clamp (1998)	12.2	1.14 (1.04, 1.25)	12.7 (1.6, 100.85)			87	+++
Davis (1987)	5.2	1.08 (0.97, 1.20)	1.27 (0.9, 1.78)			65	+
Jenkins (1996)	-2.8	0.96 (0.78, 1.19)	0.86 (0.39, 1.93)			75	—
Kelly (1987)	3.4	1.31 (0.49, 3.52)	1.36 (0.44, 4.23)			11	+
Kendrick (1999)	3.2	1.04 (0.98, 1.09)	1.49 (0.82, 2.7)			90	+
King (2001)	-1.6	0.98 (0.96, 1.01)	0.59 (0.28, 1.25)			98	—
Mathews (1988)	8.3	1.11 (0.74, 1.68)	1.67 (0.22, 12.35)			75	+
Thomas (1984)	12.4	1.15 (0.95, 1.38)	5.14 (0.53, 49.5)			84	+
Williams (1988)	No stats	No stats	No stats	No stats	No stats	>77	0

$^{+++}$ Significantly favours intervention; + Trend towards intervention; 0 No difference; – Trend towards control; — Significantly favours control.

Table 4. Table (Truncated) Showing Various Components of the Evaluated Interventions

	Burn educ.	Slides	Handouts	Safety advice	Discount devices coupons	First aid training	Home safety inspection	Tailored educ.	Reinforcement	Video	Modelling	Free thermo-meter/choke tube	School fire safety lessons	Child safety educ.
Barone (1988)	✓	✓	✓											
Clamp (1998)			✓	✓	✓									
Davis (1987)			✓										✓	
Jenkins (1996)	✓													
Kelly (1987)			✓				✓							✓
Kendrick (1999)				✓	✓	✓	✓							
King (2001)					✓		✓	✓	✓					
Mathews (1988)			✓				✓			✓	✓	✓		
Thomas (1984)	✓		✓		✓									
Williams (1988)	✓		✓	✓						✓				

Note: All these studies relate to children/familes.

1988). For the subsequent steps, the relative risk and the more 'informative' (colour-coded) vote count were both used.

The vote count supported the observations previously made by looking across the absolute risk values. Where several studies reported the same outcome, most show a tendency to favour the intervention over control, though the relative risk is usually small (see Table 3). Only one study reported any statistically significant differences between intervention and control groups (Clamp reported statistically significant positive effects of intervention on final smoke alarm ownership and final functioning smoke alarms).

In this case, the colour-coded descriptive vote count allows the reader to examine the outcome data either as a simple vote count or as a statistical value, depending upon the 'focus' they adopt when examining the outcome table.

Element 3: Exploring Relationships within and between Studies

As stated in the guidance, at this point in the synthesis the reviewers move beyond identifying, listing, tabulating and/or counting results to explore relationships within and across the included studies. The relationships of interest are of two broad types: (1) those between characteristics of individual studies and their reported findings and (2) those between the findings of different studies. Tools/techniques described in this section of the guidance are described in Table 5.

The four main tools and techniques for exploring relationships within and between studies were conducted in the order described here.

Examination of moderator variables and subgroup analyses It is helpful to identify any variables that might moderate the main effects being examined in the review (Cooper and Hedges, 1994). Two further types of table were drawn up to help investigate whether there were any such moderators of effect. The first table shows the various components that make up the intervention for each study and the overlap in components between different interventions (Table 4).

This indicates that there is little overlap between the studies in terms of the specific components employed within the interventions being evaluated. Seven of the ten studies concerned with children and/or their families used handouts and four used 'burn education', money-off coupons or discounted devices and home safety inspections. However, this apparent lack of overlap might be due to the fact that studies were, on the whole, very poorly described. Even when sufficient information was reported to allow extraction, there was still variation in the terms and definitions used by different authors, making direct comparisons even more difficult.

A second set of tables build on the 'outcomes/vote count' table, by incorporating further information taken from the 'data extraction' and 'intervention components' tables described previously (see Table 6 for an example for the 'final smoke alarm ownership' outcome). Intervention, population and setting columns were included to identify potential subgroups/moderators. These are described as briefly as possible (1–5 words) to simplify visual comparison across the table. The description of the intervention is broken into three separate cells to facilitate such visual comparisons for the complex interventions.

Table 5. Selection of Tools and Techniques for Exploring Relationships between Studies

Name of tool/technique	Comments in relation to current synthesis	Should this tool/technique be applied here?
Examination of moderator variables and subgroup analyses	Effects are likely to be moderated by variations in intervention, population and/or possibly setting.	Yes
Idea webbing/conceptual mapping	This may help structure the planned investigation of moderator variables.	Yes
Conceptual triangulation	This approach would be more appropriate to a synthesis of implementation studies, in which more qualitative information is likely to be available and there is greater scope for model development.	No
Reciprocal/refutational translation	Insufficient qualitative evidence in this review.	No
Qualitative case descriptions	This is similar to the 'textual descriptions' described (but not implemented) earlier. However, here the approach is presented in the context of investigating differences between, rather than simply describing, the studies. Might be worthwhile to revisit the studies and extract detailed data from them, with an eye to any potential moderator variables.	Yes
Visual representation of relationship between study characteristics and results	This is possible given the quantitative data available for each study	Yes
Investigator and methodological triangulation	More applicable to qualitative studies. As all studies here were RCTs, there should not be any systematic difference in results between authors from different disciplines (if there was, bias would be a very serious concern). Data on the disciplinary perspective/expertise of investigators was not available for all studies.	No

Table 6. Final Smoke Alarm Ownership (Potential Moderator Variables)

Reference	*Intervention*			*Population*	*Setting*	*Absolute difference (%)*	*% smoke alarm ownership control group*	
Barone (1988)	Burn education	Slides	Handouts	Parents of toddlers	Hospital, family home	4.5	90	+
Clamp (1998)	Safety advice	Discount devices	Handouts	Parents of children <5yrs	Family home, other	12.2	87	+++
Davis (1987)	Fire safety lessons		Take home material for parents	Children	School	5.2	65	+
Jenkins (1996)	Discharge teaching book on burn care/prevention			Children<17yrs	Hospital burn unit	−2.8	75	–
Kelly (1987)	Child safey education		Home safety inspection Handouts	Families of babies 3-12 mths	Family home	3.4	11	+
Kendrick (1999)	Safety advice First aid training	Discount devices	Home safety inspection	Families of hospitalized children<8yrs	Community	3.2	90	+
King (2001)	Tailored education Reinforcement	Discount coupons	Home safety inspection	Mothers of toddlers (12-18mths)	Family home	−1.6	98	–
Mathews (1988)	Video Modeling re safety	Free thermometers and choke tube.	Home safety inspection Handouts	Parents of infants	Family home	8.3	75	+
Thomas (1984)	Well-baby classes plus burn prevention education lecture.	Discount smoke alarm coupon	Handouts Pamphlet	Pregnant women (last trimester)	Hospital (?)	12.4	84	+
Williams (1998)	Burn prevention lectiure		Handouts		Unclear	No stats	>77	0

+++ Significantly favours intervention; + Trend towards intervention; 0 No difference; – Trend towards control; — Significantly favours control.

Turning to the outcome of 'final smoke alarm ownership' (for which the majority of studies provide data), four studies stand out from the majority of positive but statistically non-significant findings: Williams (no difference), Clamp (significantly positive), Jenkins and King (both non-significantly negative). Williams reports that 'there were no differences between experimental and control groups', though whether this means there was truly no difference between the groups or that any observed differences were not statistically significant is unclear. Either way, it is difficult to determine why the studied intervention had little or no effect based on this one study The intervention studied by Clamp included safety advice, discounted safety devices and handouts and resulted in a significant increase in final smoke alarm ownership and function. However, these particular intervention components were common to other studies that differed from Clamp's study in terms of both magnitude and statistical significance of effect. The two negative studies on the ownership outcome (Jenkins and King) evaluate two different interventional approaches. However, these studies do share a common characteristic that is not present in the 'positive' studies: the intervention was delivered to the families of children who had been previously hospitalized for an injury.

Qualitative case reports/textual descriptions It was decided that writing a short summary of each study at this stage of the synthesis (i.e. having already organized, described and examined them) would provide an opportunity to check the previous stages for accuracy, and allow the reviewer to draw out in detail any aspects of individual studies that may not have seemed relevant at the start of the synthesis, but have become of interest during the subsequent stages of describing and exploring the study data. These summaries were structured such that they provided details of the setting, participants, intervention, comparison and outcomes, along with any other factors of interest (see Box 2).

A number of questions arose from the process of writing these summaries:

- Does the immediate on-site availability of smoke alarms in the intervention setting increase uptake?
- Are lower income families more likely than higher income families to respond to interventions incorporating discounted smoke alarms?
- Does having experienced a child injury prior to intervention increase uptake of the recommendations given in the intervention?
- Do interventions that focus on burn injuries/fire prevention have different effects to interventions that relate to safety more generally?
- Does advice being age-specific alter outcomes? Would advice regarding fire safety always be the same, independent of child age?
- Does attrition have an effect?
- Is length of follow-up an important factor?
- Is sample size important?

Studies may be powered to detect differences on other outcomes. Several studies attribute any lack of effect to the fact that an active effort is required to install smoke alarms. Is there a relationship between intervention effectiveness and amount of active effort required?

Box 2. Textual Descriptions/Qualitative Case Descriptions of Included Studies – an Example (Barone, 1988)

Setting

US suburban hospital.

Participants

Individuals or couples attending a continuing-education series on 'Parenting the Toddler'. Predominantly middle and upper-middle class and well educated.

Intervention

Parenting information, with specific information and materials on burn prevention and child restraints. Included a slide presentation on falls, strangulations, drownings, poisonings and fire hazards, plus additional slides on the hazard of hot tap water, use of smoke detectors and the advantages of child car seats. Four weekly sessions, each of two hours duration. 41 participants.

Comparison

Parenting information, with general child safety information. Included a slide presentation on falls, strangulations, drownings, poisonings, and fire hazards. 38 participants.

Outcomes

A researcher inspecting participants' homes looked for and tested any smoke alarms, six months after the classes.

Other

- Protocol is very similar to that described by Williams (same university and year).
- The author suggests that the very high rate of smoke alarm ownership might be due to previous health promotion efforts.
- The author also suggests that it would have been possible for participants in the control group to be 'warned' in advance what the researchers were looking for and testing during home inspections by other participants whose homes had been inspected.

This suggests that revisiting studies and producing textual descriptions can be a helpful prelude to identifying and assessing impact of moderator variables, building on data extraction and developing conceptual models.

Developing conceptual models/idea webbing/concept mapping The aim here was to make transparent the logic behind the subgroup analyses/investigation of moderator variables (see Figure 2). The resulting figure is also partly a way to link the previously described processes and the resulting issues/ideas together in order to structure the synthesis. It reflects a process whereby variables or patterns were identified in one of the previously described tables or documents and then re-examined from the viewpoint of the remaining tables/documents. For example, the characteristic most fully explored in the

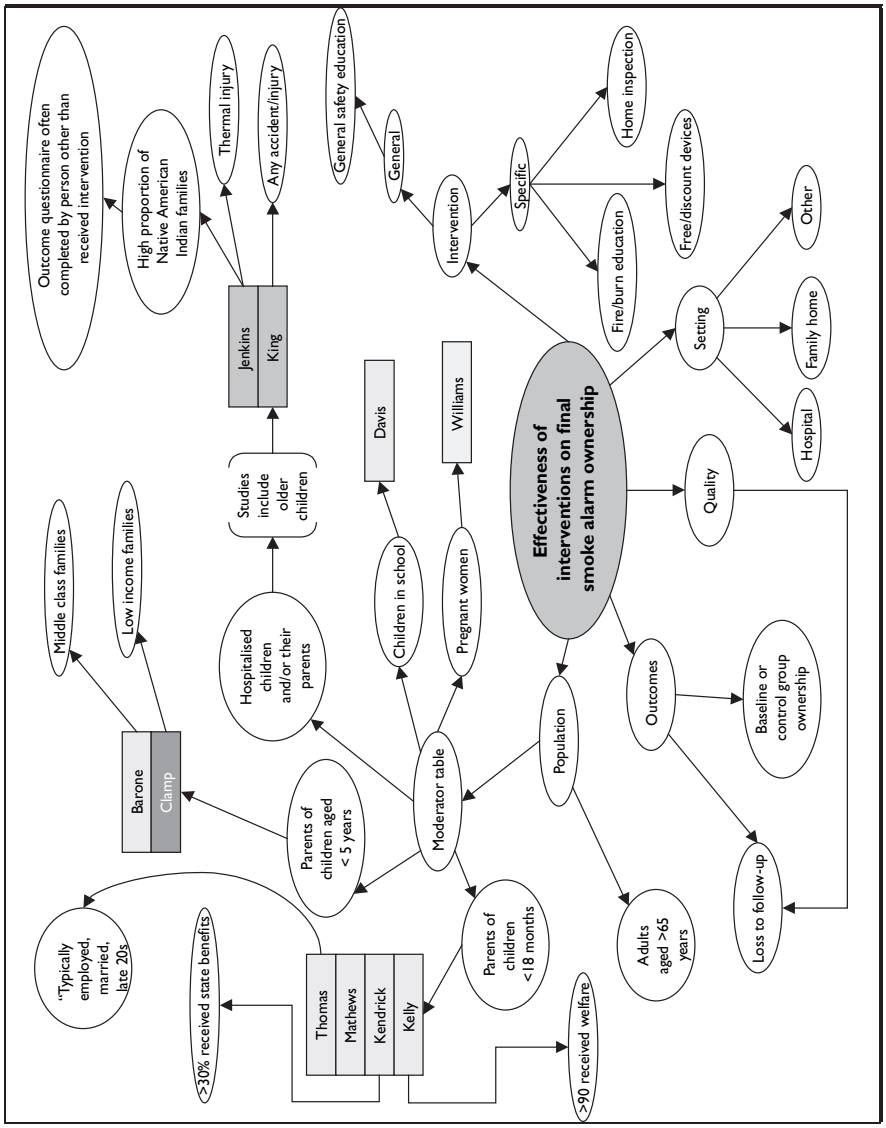


Figure 2. Conceptual Mapping/Idea Webbing

figure is that of study population, as described in the table of potential moderator variables and in the textual descriptions. Studies of children/families were grouped by age of the included children according to the moderator tables. Within these groups, further participant variables such as socioeconomic status were identified using the textual descriptions.

The 'outcomes' and 'quality' nodes are connected to one another via 'loss to follow-up'. The withdrawal rates vary substantially across this group of studies, from 0 to 67 percent. Where high dropout rates are discussed in these studies, it is attributed to non-attendance over time or unavailability of participants at final follow-up.

Though identified as potential moderators, no clear or consistent effect on smoke alarm ownership could be seen across studies for intervention variables such as the use of home inspections or free/discounted devices, or for fire/burn-specific education alone versus general safety information that incorporates fire/burn material.

Visual representation of relationships between study characteristics and results Funnel plots to examine the relationship between study sample size/variance and effect size were constructed by plotting relative risk against standard error (see Figure 3 for an example). Due to the small number of studies reporting data on the outcomes of interest, these proved to be largely uninformative. The plot for 'final smoke alarm ownership' shows that the study with the lowest precision is that with the most strongly positive effect, but this alone does not provide strong evidence for publication bias.

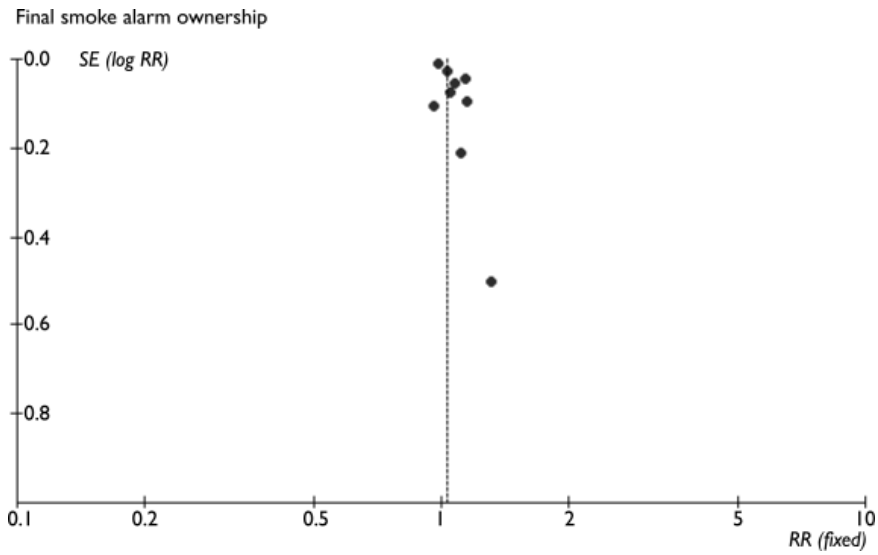


Figure 3. Funnel Plots Showing Standard Error versus relative risk for the Outcome of Final Smoke Alarm Ownership

These proved unhelpful but may be more useful in larger reviews where enough quantitative data are reported to allow a visual display. However this may not be the case for many systematic reviews.

Forest plots showing the point estimates and 95 percent confidence intervals for each study for each of the main outcomes (but without a pooled estimate) were also drawn, as suggested in the guidance (Figure 4). These provide a clear visual representation of the relative risks and associated 95 percent confidence intervals previously presented in Table 3.

Element 4: Assessing the Robustness of the Synthesis Product

The guidance states that, towards the end of the synthesis process, the analysis of relationships within and between studies described should lead into an overall assessment of the strength of the evidence available for drawing conclusions on the basis of a narrative synthesis. Tools and techniques for this purpose are presented in Table 7.

Strength of evidence (EPPI approach) This was summarized in the guidance. Four criteria are used to appraise each study: (1) the study's methodological soundness, (2) the appropriateness of the study design to answering the review question, (3) the study relevance, and (4) an assessment of the overall weight of evidence which the study provides. The first three criteria contribute to the assessment of (4) study 'weight'. These are described elsewhere by EPPI review authors as (1) trustworthiness, (2) appropriateness, (3) relevance, and (4) overall weight. An attempt was made to tabulate these characteristics for the studies included

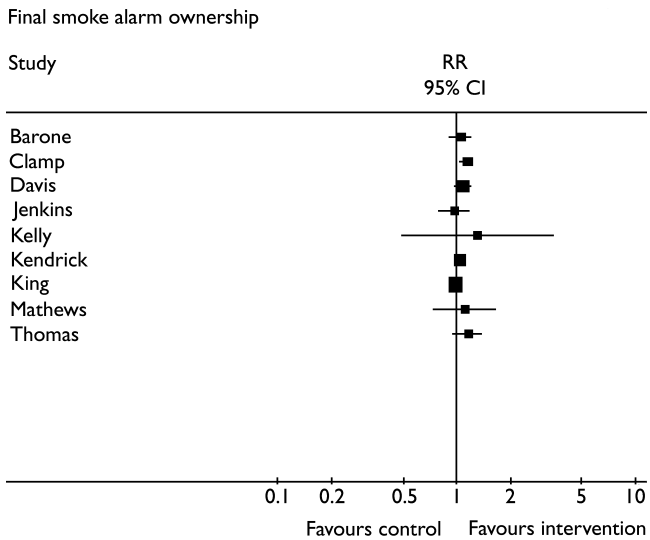


Figure 4. Forest Plot (Without Pooled Data) for Outcome of Final Smoke Alarm Ownership

Table 7. Selection of Tools and Techniques to Assess the Robustness of the Synthesis Product

Name of tool/technique	*Comments in relation to current synthesis*	*Should this tool/ technique be applied here?*
Best evidence synthesis	Not really appropriate since this technique is primarily concerned with the selection of studies, and all studies in this synthesis are RCTs.	No
Use of validity assessment (EPPI-centre approach, CDC approach)	EPPI approach may be possible, using internal validity data presented in the summary tables. CDC approach needs further clarification before it could be applied (e.g. what is a 'sufficient' effect size)	Yes (EPPI), No (CDC)
Checking the Synthesis product with authors of primary studies Reflecting critically on the synthesis process	Not possible given the time available for this synthesis. No Although partly done throughout this process, it might be useful to have a dedicated section discussing issues that arose from the synthesis	No Yes

here (Table 8), with criterion (1) based upon the validity evaluations in the first data extraction table (as these are derived from the Jadad scale, scores of 3–5 are considered 'high' quality. In this example, a score of 2 was described as 'medium' and a score of 0 or 1 as 'low').

Of the ten studies of children or their families, three received an overall 'high' weight, five were classified as 'medium' and two were given an overall weight of 'low'. These 'overall weights' corresponded exactly to the 'trustworthiness' scores that relate to internal validity. This is because there was little to distinguish between the studies in terms of appropriateness (all were RCTs – a design appropriate to this kind of evaluative research) and relevance (studies were selected for relevance early in the review by the application of inclusion criteria). The only study that was not considered 'highly' relevant in its focus was by Davis, as this was delivered to exclusively to schoolchildren, whereas other studies involved parents in the intervention.

It is possible that these 'overall weights' overemphasize the differences between the included studies. All of the studies scored 1, 2 or 3 on the Jadad scale and were consequently labelled 'low', 'medium' and 'high' respectively. All of the studies described themselves as RCTs, and (partly because of the nature of the intervention) none were double-blind or used an indistinguishable control intervention. Therefore, overall study weighting was dictated solely by whether the studies included descriptions of allocation concealment and/or withdrawals.

Three studies received an overall weighting of 'high' (Clamp, Kendrick and King). However, these were conducted in different settings and, for final smoke alarm ownership, reported differing results from one another. Consequently, this quality assessment approach does not greatly impact on the current synthesis,

Table 8. Weighting of Studies by Quality, According to Four Criteria

Study	*Trustworthiness*	*Appropriateness*	*Relevance*	*Overall weight*
Barone (1988)	Medium	High	High	Medium
Clamp (1998)	High	High	High	High
Davis (1987)	Medium	High	Medium	Medium
Jenkins (1996)	Medium	High	High	Medium
Kelly (1987)	Medium	High	High	Medium
Kendrick (1999)	High	High	High	High
King (2001)	High	High	High	High
Mathews (1988)	Low	High	High	Low
Thomas (1984)	Low	High	High	Low
Williams (1988)	Medium	High	High	Medium

though it could prove more useful in syntheses where there is greater variation in the quality of the studies being synthesized.

Reflecting Critically on the Synthesis Process

Methodology of the Synthesis Used There were some limitations to the approach taken in this synthesis, relating to the potential for bias. For example, the selection and arrangement of intervention components included in the moderator table was to some extent subjective. Similarly, the themes emerging from the textual descriptions that seemed most important were chosen at least partly subjectively. This may be an argument for 'downweighting' conclusions based on moderators identified through the extensive examination of the primary studies.

In the case of this particular synthesis, only RCTs were included. Subsequently, there was less methodological heterogeneity than in many narrative syntheses. This precluded the use of several techniques (although it is unlikely that any synthesis would need to make use of *all* the tools and techniques described in the guidance). As all the studies were RCTs, the techniques that were appropriate were often variations on those used when undertaking a meta-analysis. It also meant that the variation in quality between studies was relatively small and difficult to incorporate usefully into the synthesis.

For this example, data were checked and interpretations were agreed by a second and third reviewer. Whether this is the optimal approach to ensuring accuracy and reliability within narrative synthesis is unclear.

Conclusions of the Narrative Synthesis

Interventions that provide safety information directly to families of young children appear to have a small beneficial effect on smoke alarm ownership and function. No conclusions can be made about the effect of such interventions in terms of fire-related injury or burn prevention, as these outcomes were not reported separately. It is unclear from the synthesis of RCTs presented here

how specific fire-related safety education compares with general safety advice. Neither is there a clear relationship between the incorporation of home inspections or discount devices/coupons and the effect of interventions on smoke alarm ownership/function.

However, examination of the studies indicated several implications for the conduct of research in this area:

- Future RCTs of similar interventions should measure relevant fire-related injury and burn outcomes after an appropriately long follow-up, preferably from hospital record review or similar method that reduces the potential for bias and attrition inherent in the questionnaire methods employed in several of the currently published RCTs. However, we acknowledge secondary measures of effectiveness are usually also necessary, given that fire/burn injuries and deaths are relatively rare.
- Any future studies should provide full and detailed descriptions of the intervention being evaluated and each of its components.
- Theory should be incorporated into the design and evaluation of any such intervention. Those designing evaluations of this type of intervention should consider the causal pathways between providing the intervention and the outcomes, and the barriers to its adoption, and ensure that data is provided on each of the steps (or events) in the pathway
- Randomized studies should take into account confounding due to concurrent community-wide initiatives and legislation to increase fire injury awareness and smoke alarm ownership.
- The rates of smoke alarm ownership at baseline might be investigated as a potential variable that influences intervention effectiveness within the target population.
- The only studies with negative findings in this synthesis were those in which participants were children, or the families of children that had been hospitalized for an injury. Whether this was a chance finding or indicative of a true lack of effect for these interventions in families of previously injured children may be of interest.

Comparison of Narrative Synthesis and Meta-Analysis

The Cochrane review (DiGuiseppi and Higgins, 2001) based on the meta-analysis of the same group of RCTs, reached very similar broad conclusions to the narrative synthesis. The Cochrane authors reported that fire-related injury outcomes were not available and the main meta-analyses of RCTs showed that

> smoke alarm ownership at follow-up appeared somewhat more likely in the intervention group (OR = 1.26; 95% C.I., 0.87 to 1.82). Similarly modest positive, statistically non-significant effects on functioning smoke alarms, and on new acquisitions of smoke alarms and functioning smoke alarms, were found.

They summarized that there were 'only modest potential benefits from education to promote smoke alarms'.

As in the narrative synthesis, the apparent lack of effect of intervention in the two trials involving families of injured children was noted in the Cochrane review. The Cochrane authors state that 'exclusion of these trials from the meta-analyses results in a stronger, statistically significant intervention effect on alarm ownership (OR = 1.43; 95% C.I., 1.07 to 1.90) and other alarm outcomes'. On this basis, they suggest 'Having an injured child may lead to safety behaviour changes so large that they obscure any safety education effects', but they do not mention this as one of their implications for future research.

In addition, the Cochrane review concluded that smoke alarms delivered as part of child health surveillance may be more effective. The effects on final smoke alarm ownership were statistically significant (OR = 1.96; 95 percent CI, 1.03 to 3.72), with strong, non-significant effects on the other ownership and function outcomes. The authors state that these subgroup analyses were based on few trials and were heavily influenced by a single trial (Kendrick et al., 1999).

The results of a subgroup meta-analysis suggested that offering discounted alarms had a modestly stronger effect on smoke alarm ownership (OR = 1.83; 95 percent CI, 0.63 to 5.28) than did education alone, but the trial results were significantly heterogeneous ($p = 0.015$). Another subgroup meta-analysis indicated that the removal of the one study in which a research assistant delivered the intervention (King et al., 2001) resulted in a stronger positive effect of intervention on three of the reported outcomes.

The Cochrane authors concluded that the quality of the available evidence is limited, with sensitivity analyses showing that pooled trials with blinded outcome assessment indicated little apparent effect on ownership or function, whereas unblinded studies indicated strong effects.

The Cochrane review made the following recommendations for future research: 'Further trials to evaluate the effect of smoke alarm promotion as part of child health surveillance in primary care . . . should assess their impact on fire-related injuries, using adequate allocation concealment and blinded outcomes assessment.' No recommendations were made in relation to improvements in outcome measurement, description of interventions, use of theory in designing interventions, or adjusting for potential confounding from concurrent fire safety initiatives/policies.

On the whole, the findings of the narrative synthesis and the meta-analyses were very similar. However, the differences mentioned appear to be attributable to two main factors: the impact of sensitivity and subgroup pooled analyses during meta-analyses and the close scrutiny of study reports undertaken in narrative synthesis. Consequently, conclusions about the impact of moderators of effect appeared to be 'firmer' when derived from the meta-analysis, whereas implications for future research appeared to be more extensive and detailed when derived from the narrative synthesis.

However, the Cochrane review authors mention caveats in relation to some of the 'additional' findings derived from subgroup analyses (e.g. that the apparent increase in effect attributable to offering discounted alarms was based on a meta-analysis of highly heterogeneous studies).

Meta-analysis allowed the authors of the Cochrane review to observe the impact of specific aspects of study validity (allocation of concealment and blinded

outcome measurement) on results. In the narrative synthesis, validity was considered more broadly and showed no obvious correlation with study results. Although the differences in the conclusions of the two syntheses were relatively minor, it is unclear whether it would be possible to eradicate them altogether, considering that by definition narrative synthesis precludes statistical pooling.

Implications for the Guidance and Further Research

This demonstration indicates that the guidance successfully contributes to increased transparency and reproducibility in the narrative synthesis process. When comparing these two syntheses, it was possible to check the conclusions derived from the narrative synthesis by examining the synthesis itself and the associated tables and figures, much as it is possible to examine and interpret data presented in a series of forest plots.

Because of the perceived lack of prior knowledge and expertise required, reviewers may select a 'simple' narrative synthesis over potentially applicable but more complex approaches such as meta-ethnography (Campbell et al., 2003) or Bayesian methods (Roberts et al., 2002). Unfortunately, such a decision often results in little more than a brief summary of each of the included studies, with no attempt to truly synthesize the data from these studies. One key benefit of the guidance framework is that, by unpacking the narrative synthesis process into four elements, it explicitly requires the reviewer to move beyond producing a simplistic summary of research findings, toward developing a more reflective and reflexive approach.

In the case of this demonstration, the narrative synthesis added meaning and value to the meta-analysis, explicitly engaging with the theory underpinning the interventions and consequently raising more detailed implications for further research. The true value added to meta-analysis, as perceived by users of systematic reviews, may itself be an area worthy of further investigation.

In some cases, tools and techniques associated with different elements of the guidance (e.g. 'textual descriptions' and 'qualitative case descriptions') appeared to overlap, resulting in an unnecessary duplication of effort. In this demonstration, the application of several tools and techniques ultimately promoted greater engagement with the data being reviewed, so redundancy of these tools/techniques may not be a significant issue. However, the reviewer must maintain a balance between this thorough engagement with the data and the opportunity to undertake 'data dredging' to uncover potentially spurious associations.

A limitation of this particular demonstration (as opposed to the guidance itself) is that the synthesis focuses on a group of homogeneous RCTs. This is because our primary aim was to compare the findings of a guidance-led narrative synthesis against the findings of a meta-analysis of the same data, and for data to be included in a meta-analysis at all, they must be largely homogeneous (Deeks et al., 2006). This comparison allowed us to investigate how the method of synthesis might influence the findings and conclusions of a systematic review of effectiveness. However, in practice, narrative synthesis is frequently considered by systematic reviewers to be the 'fallback' approach when included studies are

too heterogeneous to be combined statistically. It therefore might be of value for future investigations to evaluate the impact of the guidance on the findings of a narrative synthesis conducted under such circumstances (i.e. with a heterogeneous group of primary studies).

Conclusions

We found that unpacking narrative synthesis into four components was helpful in identifying what narrative synthesis 'consists of' and may be helpful in increasing the transparency of narrative methods.

Moreover, it is usually assumed that the choice of 'meta-analysis' or 'narrative review' is an either/or one. Instead we found that quantitative and qualitative methods of summarizing findings appear to make different contributions to a systematic review. We found that meta-analysis offers the opportunity to undertake structured analyses in relation to key moderators of effect, whereas narrative synthesis offers an insight into potential confounders and moderators that might not necessarily be taken into account during a typical meta-analysis.

A meta-analysis may benefit from a detailed narrative review. Conversely a narrative review may benefit from a more systematic approach to description and summary, and we have identified some tools which may be helpful in this regard. Future research may usefully consider their application to other topics, and compare their use in other systematic reviews.

References

Arai, L., N. Britten, J. Popay, H. Roberts, M. Petticrew, M. Rodgers and A. J. Sowden (2007) 'Testing Methodological Developments in the Conduct of Narrative Synthesis: A Demonstration Review of Research on the Implementation of Smoke Alarm Interventions', *Evidence and Policy* 3: 361–83.

Barone, V.J. (1988) 'An Analysis of Well-Child Parenting Classes: The Extent of Parent Compliance with Health Care Recommendations to Decrease Potential Injury of their Toddlers', dissertation, University of Kansas, Kansas City, KS.

Campbell, R., P. Pound, C. Pope, N. Britten, R. Pill, M. Morgan and J. Donovan (2003) 'Evaluating Meta-Ethnography: A Synthesis of Qualitative Research on Lay Experiences of Diabetes and Diabetes Care', *Social Science and Medicine* 56: 671–84.

Clamp, M. and D. Kendrick (1998) 'A Randomised Controlled Trial of General Practitioner Safety Advice for Families with Children under 5 Years', *BMJ* 316: 1576–9.

Cooper, H. and L. V. Hedges, eds (1994) *The Handbook of Research Synthesis.* New York: Russell Sage Foundation.

Davis, J. W. (1987) 'Get Fired up: A Model for Developing In-School 4-H Programs', dissertation, Clemson University, Clemson, SC.

Deeks, J., J. Higgins and D. Altman (2006) 'Analysing and Presenting Results', in J. Higgins and S. Green (eds) *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6* [updated Sept. 2006]; *Section 8.* Chichester: John Wiley & Sons, Ltd.

DiGuiseppi, C. and J. P. T.Higgins (2001) 'Interventions for Promoting Smoke Alarm Ownership and Function', *The Cochrane Database of Systematic Reviews: Reviews 2001 Issue 2* John Wiley & Sons, Ltd Chichester, UK DOI: 10.1002/14651858.

Dixon-Woods, M., S. Agarwal, D. J. Jones, B. Young and A. Sutton (2005) 'Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods', *Journal of Health Service Research Policy* 10: 45–53.

Greenhalgh, T. (2004) 'Meta-Narrative Mapping: A New Approach to the Systematic Review of Complex Evidence', in B. Hurwitz, T. Greenhalgh and V. Skultans (eds) *Narrative Research in Health and Illness.* Oxford: Blackwell Publishing.

Hedges, L. V. and I. Olkin (1985) *Statistical Methods for Meta-Analysis.* London: Academic Press.

Jadad, A. R., R. A. Moore, D. Carroll, C. Jenkinson, D. J. Reynolds, D. J. Gavaghan and H. J. Mcquay (1996) 'Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary?', *Controlled Clinical Trials* 17: 1–12.

Jenkins, H.M., V. Blank, K. Miller, J. Turner and R.S. Stanwick (1996) 'A Randomized Single-Blind Evaluation of a Discharge Teaching Book for Pediatric Patients with Burns', Journal of Burn Care and Rehabilitation 17(1): 49–61.

Jones, D., K. A. Roberts, M. Dixon-Woods, R. Fitzpatrick and K. R. Abrams (2003) 'Bayesian Synthesis of Qualitative and Quantitative Evidence: Overview and Example', *HDA Synthesis Series on Promoting Methodological Development in Evidence Synthesis: Understanding Bayesian Approaches to Synthesis.*

Kelly, B., C. Sein and P. L. McCarthy (1987) 'Safety Education in a Pediatric Primary Care Setting', *Pediatrics* 79: 818–24.

Kendrick, D., P. Marsh, K. Fielding and P. Miller (1999) 'Preventing Injuries in Children: Cluster Randomised Controlled Trial in Primary Care', *BMJ* 318: 980–3.

King, W. J., T. P. Klassen, J. Leblanc, A. C. Bernard-Bonnin, Y. Robitaille, B. Pham, D. Coyle, M. Tenenbein and I. B. Pless (2001) 'The Effectiveness of a Home Visit to Prevent Childhood Injury', *Pediatrics* 108: 382–8.

Mathews, J. R. (1988) 'An Analysis of Dangerous Behaviour in Toddlers', dissertation, University of Kansas, Kansas City, KS.

Pawson, R., T. Greenhalgh, G. Harvey and K. Walshe (2004) *Realist Synthesis: An Introduction*. Manchester: ESRC Research Methods Programme.

Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, N. Britten, M. Rodgers, K. Roen and S. Duffy (2006) *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Final Report.* Swindon: ESRC Methods Programme.

Roberts, K. A., M. Dixon-Woods, R. Fitzpatrick, K. R. Abrams and D. R. Jones (2002) 'Factors Affecting Uptake of Childhood Immunisation: A Bayesian Synthesis of Qualitative and Quantitative Evidence', *Lancet* 360: 1596–9.

Thomas, K., R. S. Hassanein and E. R. Christophersen (1984) 'Evaluation of Group Well-Child Care for Improving Burn Prevention Practices in the Home', *Pediatrics* 74: 879–82.

Williams, G. E. (1988) 'An Analysis of Prenatal Education Classes: An Early Start to Injury Prevention', dissertation, University of Kansas, Kansas City, KS.

MARK RODGERS is a Research Fellow at the Centre for Reviews and Dissemination, University of York, where he researches evidence synthesis and clinical effectiveness. Please address correspondence to: Research Fellow, Centre for Reviews and Dissemination, University of York YO10 5DD, UK. [email: mr14@york.ac.uk]

AMANDA SOWDEN is Deputy Director at the Centre for Reviews and Dissemination, where much of her work has focused on the use of evidence

synthesis to evaluate the effects of public health interventions. Address: CRD, University of York, York YO10 5DD. [email: ajs18@york.ac.uk]

MARK PETTICREW is Professor of Public Health Evaluation at the London School of Hygiene and Tropical Medicine, where he researches the health effects of social interventions. Address: London School of Hygiene & Tropical Medicine, Room 18, Keppel St., London WC1E 7HT. [email: Mark.Petticrew@lshtm.ac.uk]

LISA ARAI is a lecturer in children and young people at the Open University. Her interests are in child public health and methodological innovation. Address: Faculty of Health and Social Care, The Open University, Walton Hall, Milton Keynes, MK7 6AA. [email: l.arai@open.ac.uk]

NICKY BRITTEN is Professor of Applied Healthcare Research at the Peninsula Medical School. She has been involved in the development of meta-ethnography for qualitative synthesis. Address: Peninsula Medical School, St Luke's Campus, Exeter EX1 2LU. [email: nicky.britten@pms.ac.uk]

JENNIE POPAY is Professor of Sociology and Public Health at Lancaster University. Her research interests include health inequalities and qualitative evidence synthesis. Address: Institute for Health Research, Bowland Tower East, Lancaster University, Lancaster LA1 4YT. [email: j.popay@lancaster.ac.uk]

HELEN ROBERTS is Professor of Child Health in the Institute of Education, University of London. Her main research interests are inequalities in child health, and evidence synthesis. Address: Social Science Research Unit, Institute of Education, University of London, 18 Woburn Square, London WC1H 0NR. [email: H.Roberts@ioe.ac.uk]