

SYSTEMATIC REVIEW AND METHODOLOGICAL EXPLORATION OF SYNTHESIS  
METHODS IN PUBLIC HEALTH EVALUATIONS OF INTERVENTIONS CONCLUDES  
THAT THE USE OF MODERN STATISTICAL METHODS WOULD BE BENEFICIAL

Authors: Felix Achana <sup>†1</sup>, Stephanie Hubbard<sup>†1</sup>, Alex Sutton<sup>1</sup>, Denise Kendrick<sup>2</sup>, Nicola  
Cooper<sup>1</sup>

<sup>†</sup> Joint first authors

<sup>1</sup>Department of Health Sciences, University of Leicester, University Road, Leicester LE1  
7RH, UK

<sup>2</sup>Division of Primary Care Community Health Sciences, Faculty of Medicine & Health  
Sciences, University of Nottingham NG7 2RD, UK

*Corresponding author:*

Professor Nicola Cooper

Department of Health Sciences, 2<sup>nd</sup> Floor Adrian Building, University of Leicester,  
University Road, LE1 7RH, UK

Tel: +44 (0)1162297267

Fax: +44 (0)1162297250

Email@ njc21@le.ac.uk

## ABSTRACT

**Objectives:** To review the methods currently used to synthesise evidence in public health evaluations and to demonstrate the availability of more sophisticated approaches.

**Study design and setting:** A systematic review of NICE public health appraisals published between 2006 and 2012 was performed to assess the methods used for the synthesis of effectiveness evidence. The ability of new developments in evidence synthesis methodology to address the challenges and opportunities present in a public health context is demonstrated.

**Results:** Nine (23%) of the 39 NICE appraisals included in the review performed pairwise meta-analyses as part of the effectiveness review with one of these also including a network meta-analysis. Of the remainder, 29 (74.4 %) presented narrative summaries of the evidence only, and 1 (2.6 %) appraisal did not present any review of effectiveness and or cost-effectiveness evidence. Heterogeneity of outcomes, methods and interventions were the main reasons given for not pooling the data. Exploration of quantitative syntheses methods shows that pairwise meta-analyses can be extended to incorporate individual participant data (where it is available), extend the number of interventions being compared by using a network meta-analysis, and adjust for both subject and summary-level covariates. All these can contribute to ensuring the analysis answers directly the policy relevant questions.

**Conclusions:** More sophisticated methods in evidence synthesis should be considered to make evaluations in public health more useful for decision makers.

Keywords:

Public Health Evaluation, Network meta-analysis, Decision making, Systematic review

What is new?

#### Key Findings

- Quantitative synthesis is not carried out in the systematic reviews for the majority of public health evaluations.
- When quantitative synthesis *is* done it tends to use the simplest methods, e.g. a fixed or random effects meta-analysis comparing two groups which potentially limits the scope of the analysis.

What this adds to what is known?

- Demonstrates how more sophisticated synthesis methods can be used in public health (PH) appraisals to more realistically model the data and answer the relevant policy questions.

What is the implication, what should change now?

- Researchers working on PH evaluations should consider expanding their toolbox and using more sophisticated methods many of which have recently been developed, motivated and applied in pharmaceutical evaluations.

## INTRODUCTION

Systematic reviews and economic evaluations conducted within a decision modelling framework are two important tools in healthcare evaluation [1, 2]. Systematic reviews with or without meta-analyses have been accepted as providing a transparent and consistent way of obtaining research evidence on effectiveness of interventions in a way that minimizes bias [3]. Decision analytic models offer an additional framework through which effectiveness evidence, ideally from a systematic review, may be integrated with other relevant evidence and information on resource utilisation in order to derive comparative estimates of cost-effectiveness. By providing a framework for assessing effectiveness and cost-effectiveness, these methods enable policy relevant questions, such as which interventions represent the best use of scarce healthcare resources, to be answered [4].

A key component of a systematic review is how the evidence, on outcomes such as effectiveness and adverse events, is synthesised. Meta-analysis, when used in a systematic review to combine quantitative information from multiple well-conducted randomised controlled trials (RCTs), is considered at the top of the hierarchy of evidence for intervention effectiveness [5]. An alternative approach to evidence synthesis, when meta-analysis is considered inappropriate, is narrative synthesis (also referred to as qualitative synthesis [6]) . In this approach, individual studies identified in the review are summarised using a variety of formats without combining results quantitatively [7].

Meta-analysis is widely applied in reviews of the effectiveness of clinical interventions, treatments and medical device technologies where the interventions and health outcomes are usually well defined and evaluated in well conducted RCTs[8]. In other fields of healthcare

evaluation, however, things may not always be as clear cut. A good example is public health (PH), where interventions are often more complex and less well defined than clinical interventions [9]. There may also be a lack of good quality evidence, particularly from RCTs in PH, for a number of well documented reasons [10, 11] including limited generalisability of the findings of RCTs to the wider population due to highly selected study populations, a narrow definition of intervention strategies and outcomes, and a focus on the individual instead of the community that is of interest in PH. Even when feasible, many have argued that RCTs may not always be possible to conduct in PH for other reasons; for example, ethical concerns may be raised regarding not offering the control population a possibly beneficial intervention [10]. Also, many of the RCTs conducted in PH tend to be cluster randomised trials and hence have more complex designs that need adjusting for in the analysis. In addition, the best available PH evidence may often come from observational non-randomised studies [9], despite the increased risk of bias associated with lack of randomisation. For these reasons, the use of quantitative evidence synthesis methods, such as meta-analysis, in PH raises a number of methodological challenges. These include: a) increased methodological heterogeneity and risk of bias as a result of including studies with different study designs (RCTs, cluster-RCTs, controlled before-and-after studies and other observational non-randomised studies), b) the interventions or 'programme' being evaluated is often described in little detail, c) a wide range of outcomes measures are often used which may be variously defined across studies and d) the use of intermediate and/or surrogate outcome measures.

There are growing calls for public health (PH) decision making to be based on the best available evidence whenever possible. For example, a 2004 Department of Health report [12] on improving health and reducing health inequalities in England called for economic

evaluations of PH interventions to ensure judicious use of scarce resources. Following this report, the remit of the UK National Institute for Health and Care Excellence (NICE), which already evaluated pharmaceutical interventions, was expanded to include the development of guidance for PH based on sound appraisals of intervention effectiveness and cost-effectiveness [13]. Consequently, a number of PH appraisals have been produced by NICE since 2006 on a wide range of issues including smoking cessation, alcohol-use, and, particularly of relevance to the example used in this paper, unintentional injuries in children.

To help address specific methodological challenges and provide advice on the technical aspects of the appraisal development process, NICE published a methods manual for PH evaluation in 2006 [14], which was subsequently updated in 2009 [15] (a further update was published in September 2012[16] after this review was completed but the guidance was not changed). The guidance recommended “Meta-analysis data may be used to produce a graph if the data (usually from RCTs) are sufficiently homogenous and if there are enough relevant and valid data from comparable (or the same) outcome measures. Where such data are not available, the synthesis may have to be restricted to a narrative overview of individual studies looking at the same question”, “Before pooling or combining the results of different studies, the degree of heterogeneity in the data should be assessed to determine how the results have been affected by the circumstances in which studies were carried out”, and “Publication bias [17, 18] should be critically assessed and reported in the interpretation of the meta-analysis results”. These recommendations match well to the challenges in systematic review/meta-analysis in PH highlighted by the Cochrane Collaboration [9] and the 2011 Institute of Medicine (IOM) report on standards for systematic reviews[6].

In view of the aforementioned challenges facing PH evaluations, and recommendations for synthesis of PH evidence contained in the NICE methods manuals, a review of all NICE PH appraisals published since 2006 was conducted. The aim of this paper is twofold: i) to identify the current situation (i.e. what is already done and/or not done) with regards to addressing problems in synthesis of PH evidence; and ii) illustrate the application of new synthesis methods (i.e. beyond those recommended by NICE[14-16] and Cochrane [9]) including methods from other fields, such as health technology assessment to PH evidence that we believe have the potential to address many of the challenges in PH evaluation as outlined above, and thus improve the quality of evidence syntheses in PH interventions.

## **SYSTEMATIC REVIEW OF NICE PUBLIC HEALTH APPRAISALS**

### *Methods*

Completed PH appraisals published between 01/03/2006 and 25/09/2012 were identified for inclusion in the review through the NICE website

(<http://www.nice.org.uk/Guidance/PHG/Published>). Each PH appraisal consisted of a number of articles such as qualitative reviews, epidemiologic reviews, expert opinions; field reports and other similar non-quantitative review reports as well as quantitative systematic reviews of intervention effectiveness, cost-effectiveness and decision analytic modelling reports. These were retrieved from the 'background information' sections and assessed for eligibility. The 'how this guidance was produced' sections were also searched for relevant articles if none were identified under 'background information'. Articles meeting the inclusion criteria were systematic reviews of the quantitative effectiveness and cost-effectiveness evidence and/or decision analytic modelling reports. Qualitative evidence reviews, epidemiologic reviews, field reports, expert opinions and other similar non-

quantitative evidence review reports were excluded. In addition, the final appraisal/guidance documents developed for each PH appraisal area were also excluded as these did not contain relevant information on the conduct of the evidence synthesis and decision modelling, which are of interest in this review. All except two (PH1 and PH2) of the appraisals were published after the 2006 NICE methods manual [14] so should have followed the guidance for quantitative effectiveness evidence synthesis techniques.

Information extracted from the retrieved articles was used to assess the methods used to synthesise the effectiveness evidence and subsequent incorporation of the evidence into the decision models (where developed) that informed the PH appraisal. The assessment criteria for the synthesis methods were:

- Type of systematic review – narrative summary versus meta-analysis;
- Included studies – RCT versus observational (non-randomised) studies;
- Methods used to synthesise the evidence (if undertaken), including specification of the statistical model (including fixed and/or random effects models), heterogeneity and publication bias and the outcome measures used, as well as presentation of results; and
- How evidence from the systematic review was used to inform any cost-effectiveness analysis.

### *Results of systematic review*

Thirty-nine completed PH appraisals published since 2006 were identified from the NICE PH website. Within these 39 appraisals, 371 potentially relevant articles were retrieved and, after



screening the titles and reading the introduction and/or abstract sections, 164 were excluded as they failed to meet the inclusion criteria above. Fifty-two articles, identified as duplicates and supplementary appendices, were combined with the corresponding main report and counted as one article leaving a total of 155 articles for inclusion in this review. The median number of included articles per appraisal was 4 (range 0 to 10)<sup>1</sup>.

*a) Type of review*

Table 1 lists all 39 PH appraisals by summary of the evidence synthesis and cost-effectiveness analyses undertaken to inform each appraisal development. One appraisal (PH36) reported no effectiveness and cost-effectiveness evidence reviews nor a decision model, 2 appraisals (PH33 and PH34) reported reviews of evidence but conducted no cost-effectiveness analysis and a 4<sup>th</sup> appraisal (PH7) reported evidence reviews and decision models, however, no estimates of cost-effectiveness of interventions were presented.

Twenty-nine (74.4 %) of the 39 appraisals contained systematic reviews in which only a narrative summary of the evidence was conducted, another 7 (18 %) conducted both narrative summary and meta-analysis, 2 appraisals (5%) conducted only meta-analysis, and 1 (2.6%) appraisal had no systematic review and hence no evidence synthesis. In the narrative summary approach, the review findings were summarised study by study in the text and through tables. Sometimes forest plots were used to display results of primary studies but no overall mean or pooled result was presented (see PH4 for an example). Eight of the 29 appraisals using only a narrative summary approach did not report the reasons for not pooling

---

<sup>1</sup> No relevant supporting document meeting our inclusion criteria existed for one appraisal (PH36 – prevention and control of hospital infection)

the data, 2 included only review level evidence from overview of reviews, and 19 cited heterogeneity as the reason why meta-analysis was not considered appropriate. The reported causes of heterogeneity are presented in Table 1A of the accompanying web appendix.

*b) Included studies – RCTs versus non-randomised studies*

Two (PH23 and PH38) of the 38 appraisals (containing a systematic review) included evidence from RCTs only in the effectiveness review. The remaining 36 appraisals were informed by reviews of both randomised and observational (non-randomised) evidence identified from individual study reports and/or published systematic review reports. All 38 appraisals (containing a systematic review) graded the quality of primary studies and assessed the applicability of the evidence adhering to the PH appraisal methods guidelines [14, 15].

*c) Quantitative evidence synthesis*

Only 9 of the 39 appraisals (23 %) contained one or more systematic review with a meta-analysis (Table 2). In total, there were 10 systematic review and/or decision analytic modelling reports with at least one meta-analysis within the 9 appraisal areas (Note: PH10 has two systematic review reports in which a meta-analysis was conducted). Four of the 10 meta-analyses included RCTs only and 6 included both RCT and observational (non-RCT) studies. Six of the 10 meta-analyses were conducted on ‘final outcomes’; that is, the main outcome measures on which the corresponding cost-effectiveness analyses were based (e.g. PH10 Smoking abstinence). The remaining 4 meta-analyses were conducted on ‘intermediate outcomes’ (e.g. PH3 Uptake of Chlamydia screening in schools rather than prevention of Chlamydia).

There was evidence that interventions may have been ‘lumped’ [19, 20] into two broad intervention groups to facilitate inclusion of more studies in 7 of the 10 reports with a meta-analysis. For example, in PH23 which investigated the effect of school-based interventions on alcohol consumption, seemingly different interventions (such as lessons delivered by teachers or other professionals as part of the curriculum; peer led education by other pupils; external contributions from for example, the police, life education centre staff; and implementation of school policy type interventions) were lumped together to form one ‘intervention group’ which was then compared to the no intervention control in a pairwise meta-analysis.

Seven of the 10 review reports conducted random effects pairwise meta-analysis, 1 conducted fixed and random effects analysis, 1 conducted random effects mixed treatment comparisons [20] (also referred to as network meta-analysis [21, 22] – see below) alongside the pairwise analysis and another one did not clearly present the statistical model used. Six of the 10 systematic reviews presented forest plots with heterogeneity statistics displayed on them, 2 (PH3 and PH1) presented forest plots without heterogeneity statistics and 1 review (PH35) did not present a forest plot. Only 1 review (PH23) assessed publication bias using funnel plot and Egger’s test for asymmetry.

*d) How the evidence from the systematic reviews were incorporated into the model*

Thirty-five (89.7%) of the 39 appraisals were informed by cost-effectiveness evaluations contained in one or more decision analytic modelling reports (Table 1). Twenty-three (66%) of these used estimates of intervention effectiveness derived from individual studies identified in the systematic review to inform the decision analysis (reasons for using the studies selected given in Table 1), 5 (14%) used previously published systematic review

results, another 5 (14%) used estimates from a meta-analysis of studies identified in the systematic review, 1 used expert opinion / analyst estimate and another one did not clearly report the source(s) of the intervention effect.

## **EXPOSITION OF NEW SYNTHESIS METHODOLOGY APPLIED IN A PUBLIC HEALTH EVALUATION CONTEXT**

In this section we outline new developments in evidence synthesis methodology, many of them motivated by the evaluation of medical interventions as well as others motivated specifically by challenges in PH. We also show how such methods can be applied in a PH context to help address challenges and opportunities that exist in this context and thus, in some situations, raise the quality bar (established in the first part of this paper) for PH interventions.

We use, for illustration, a topic area in which the authors have actively been working for several years - accident prevention in pre-school children in the home. This area of accident prevention in children at home was recently appraised by NICE PH30 (see Table 1) using only narrative summaries for the systematic review of intervention effectiveness and thus using estimates from individual trials to inform the cost-effectiveness analyses. We have found accident prevention to have many of the issues typical of PH appraisals including studies of different designs, heterogeneity in both study design (e.g. specific nature of interventions, level of randomisation (individual or cluster), etc.) and between study results, and interest in differential treatment effects across degrees of population inequality, such as accommodation type, proportion of black and minority ethnicity and proportion of single parent families.

The below account follows an approximately chronological path and details the development and adaptation of methods to synthesise the evidence by making the best use of available data. For the purposes of this paper, we restrict our attention to strategies to reduce falls in children in the home; in particular, to increase the possession of a fitted stair gate(s) in the home.

We start by discussing the analyses performed in a recently updated Cochrane review [23] of interventions to prevent unintentional injuries to children at home - pairwise meta-analysis, subgroup analyses to explore heterogeneity, and meta-regression incorporating individual participant data (IPD). We then present a network meta-analysis which allows the interventions to be ranked and provides more informative evidence for a cost-effectiveness analysis.

### *1) Pairwise meta-analysis*

A random effects meta-analysis was used to synthesise the evidence for the possession of fitted stair gate(s) outcome which comprised 12 studies (10 RCTs (2 clusters allocated) and 2 non-RCTs (1 cluster allocated)). Since the original reporting of the cluster randomized studies had ignored the effect of clustering in their analysis, the meta-analysis was adjusted by using external data to estimate the likely effects of such clustering on the certainty of the results [24]. Figure 1 displays a forest of the results. Intervention arms were more likely to possess fitted stair gate(s) than the control arms (odds ratio (OR) = 1.61, 95% confidence interval (CI): 1.19 to 2.17). Considerable heterogeneity was observed between study results,  $I^2 = 76\%$  [25].

### *2) Subgroup analyses*

Potential sources of heterogeneity were explored using subgroup analyses based on a priori explanations, which were i) whether the intervention included the provision of safety

equipment, ii) follow-up period (up to and including three months and four or more months), iii) whether the intervention was delivered in a clinical setting or in the home or community, iv) use of a randomised or non-randomised design, and v) study quality (allocation concealment, blinding of outcome assessment and at least 80% follow-up in each treatment arm). Some of the heterogeneity was partly explained by different settings and the provision of stair gates but significant heterogeneity remained in the different subgroups.

### 3) *Meta-regression using IPD and Summary data*

In an attempt to explain further variability between study results - and to address whether differential intervention effects could be discerned to be related to indicators of deprivation, and thus try and answer questions relating to inequalities in healthcare, a number of subject level covariates were explored. In order to achieve this, the individual participant data (IPD) were requested from the researchers responsible for all the relevant primary studies. By obtaining IPD, the power of meta-regression to explore subject level covariates (e.g. if subject lived in owned or rented accommodation, etc) is much increased over the use of summary data (e.g. the percentage of subjects living in an owned house in a particular study) [26]; in fact, obtaining IPD is considered the gold standard way to carry out meta-analysis generally [27].

IPD was successfully obtained for approximately half of the studies across all types of injury prevention included in the review; with varying degree of success for the different injury prevention domains. But this partial success presented an analysis challenge. We wanted to not only use the IPD, but also to include the other studies in the analysis for which only summary data were available. This involved using a model developed for the original Cochrane review in this area [28] which essentially “married” summary and IPD meta-analysis models including covariates within a single analysis based on all available data [29].

This approach also accounted for the correct analysis of the cluster allocated studies through appropriate re-analysis of the IPD (where available) as well as through utilisation of adjustment methods for the summary data as outlined above. Importantly, using IPD allowed the use of data on outcomes that had not been reported in the papers; for example, some studies had reported composite measures of home safety and not individual safety practices but the IPD included data on these individual safety practices.

For the possession of fitted stair gate(s) outcome, IPD was obtained for 10 of the 12 studies. Treatment interactions were investigated for child age, ethnic group, gender, family type (single or two parents), housing tenure (rented or owned) and parental unemployment. Most of the findings indicated little difference between the subgroups, except for the analysis of housing tenure which combined the analyses of IPD for two cluster and five non-cluster studies, and one study for which only summary data was available. The OR for intervention effect in non-owner occupied households was 1.98 (95% credible interval (CrI)<sup>2</sup>: 1.48 to 2.66) and in owner occupied households was 1.22(95% CrI: 0.96 to 1.61) providing evidence to suggest the intervention effect was larger in non-owner occupied households (ratio of odds ratios: 1.62 (95% CrI: 1.18 to 2.24)).

It is interesting to note that such covariates could have been investigated without obtaining IPD through the use of meta-regression on summary level covariates (i.e. % of study participants in non-owner occupied households), but such an analysis has much diminished power and is more prone to ecologic/ aggregation biases [30]. Running such an analysis on the same eight studies, but not using any IPD, produces an exponentiated regression coefficient of 1.01 (95% CI: 0.998 to 1.022) indicating that there is no evidence of an increase in the odds of possession of fitted stair gate(s) for a one percentage point increase in

---

<sup>2</sup> A credible interval (CrI) is similar to a confidence interval (CI) generated using Bayesian Statistics

percentage of families living in non-owner occupied accommodation. This result is very different from the findings from the IPD analysis which suggest the odds of possessing fitted stair gate(s) are 62% higher amongst those in non-owner occupied accommodation than those owning their home.

#### 4) *Network meta-analysis*

Our next refinement to the analysis, not included in the Cochrane review, came from concerns with the interpretability of the effect sizes from pairwise analyses of the type presented above. We were aware that the interventions to increase the uptake of safety practices varied between studies (e.g. interventions ranged from educational initiatives, through vouchers to reduce the price of equipment through to the free provision and fitting of equipment) and therefore by fitting the data into a meta-analysis framework of “intervention” versus “usual care” the interpretation of the resulting pooled effect was unclear – exactly what does the pooled effect relate to? This was especially important as the effectiveness results were to be used to inform the cost-effectiveness of injury prevention interventions evaluated via a decision model which would require explicit interventions to be defined and costed. Thus, an analysis in which the different interventions were kept as unique was required. Once this was established, it became possible to include further relevant literature, known about but not used in the initial meta-analysis, in the analysis; namely, studies which compared different interventions to increase safety equipment uptake directly (but which had no “usual care” control group – hence their omission thus far). Further searches of the excluded studies were conducted to identify all such studies. Network meta-analysis, which was being increasingly used in the evaluation of pharmaceuticals for funding bodies such as NICE [31], presented an analysis approach that would both keep interventions distinct and include trials with direct comparisons.



The meta-analysis of possession of fitted stair gate(s) outcome presented in the Cochrane review included all studies which compared a control group to an enhanced intervention group, but these controls and interventions varied considerably as outlined in Figure 2a. In fact, seven distinct controls and interventions (including usual care) were identified across the included studies. To better understand the structure of the evidence base, when interventions are defined in this more refined way, a network diagram of the form presented in Figure 2b [32] can be constructed. Network meta-analysis methods allow a simultaneous analysis of all the comparisons presented on the network. Table 3 [33] presents the odds ratios for the pairwise comparisons between the interventions produced both from the network meta-analysis and the direct comparisons from a trial or, where there was more than one trial, a pairwise meta-analysis of that particular comparison. In the network meta-analysis, the most intensive intervention (education + low cost/free equipment + home safety inspection + fitting) was most effective for the possession of a fitted stair gate outcome when compared to all other interventions. The probability that each intervention is best and the median rank (with uncertainty) of each intervention[31] calculated from the network meta-analysis are presented in Table 4. These show the most intensive intervention clearly had the highest probability (0.97) of being the most effective and a median rank of 1 (95% CI: 1 to 2).

Although we believe such an analysis is more refined, interpretable, complete and thus more helpful than the standard pairwise meta-analysis presented initially above, it only considered summary study data, some of which was obtained from IPD, and did not include any potentially treatment modifying covariates. We had developed models to include covariates in network meta-analysis of summary data [34] (including a special model to deal with the inclusion of the control group event rate as a covariate in network meta-analysis [35] which is not illustrated here but potentially very useful in a PH context where inequalities are of interest, particularly where IPD is not available). We have also extended the network meta-

analysis covariate model to allow the inclusion of IPD and thus subject level covariates where possible [36].

## **DISCUSSION**

This review of completed NICE public health appraisals illustrates the current situation regarding the use of evidence synthesis methods to inform public health decision making in the UK. It identified that effectiveness evidence was mostly synthesised using narrative summaries and that quantitative synthesis was not carried out for the majority of evaluations in PH systematic reviews. Of the 39 appraisals published since 2006, only 9 (23%) appraisals were informed by at least one systematic review with a meta-analysis. The other 30 appraisals may have refrained from meta-analysis due to a lack of randomized trials, or heterogeneity in study design, i.e. a mix of RCTs and non-RCTs. Moreover, systematic reviews opting for a quantitative summary tended to use the simplest methods such as fixed or random effects pairwise meta-analysis which only enables comparison between two interventions at any one time and thus potentially limiting the scope of the analysis and the utility of the findings. These findings would seem to indicate that, despite great advances in quantitative synthesis techniques, application in PH evaluation is still very much in its infancy and appears to lag behind other areas of healthcare such as the evaluation of clinical interventions. There are several reasons for this, not least due to the often heterogeneous nature of PH evidence including variations in many aspects of study design, including i) the exact nature of the interventions; ii) outcome measures; iii) the wider scope of many PH research questions; and iv) the quantitative skills of the researchers involved.

Underlying our desire for public health reviews to become more quantitative, in the face of the challenges encountered, is a firm belief that a structured and transparent description and analysis of the decision question is desirable. Our review found nearly 80% of PH NICE appraisals did not attempt a quantitative synthesis at all due to, what investigators believe but we want to challenge, insurmountable problems owing to the heterogeneous nature of the evidence base. We believe that the more complex synthesis models, described in the previous section, can often more appropriately model the types of data commonly available in PH appraisals than carrying out less focused and detailed reviews of the literature.

NICE guidance states that meta-analysis involves evidence “usually from RCTs” and therefore authors who have non-RCTs are often reluctant to conduct meta-analyses. Often the evidence from RCTs is limited and the best available evidence is from the non-RCTs. If authors attempt to identify the procedures used and choose well-conducted non-RCTs, in order to limit confounding by selection bias, then meta-analysis can be considered.

Heterogeneity between study results can be more likely and results from meta-analyses can reflect bias rather than the intervention effect so authors should be cautious (Cochrane chapter 13, Valentine 2012, Moher Quorum 1999) but not being able to present a quantitative analysis severely restricts the utility of the review; particularly for decision making. Exploring heterogeneity and attempting to account for it should be part of the analysis and greater awareness of modern methods, and greater expertise in using them, will yield fruit for future PH reports. There are several other reasons why conducting a meta-analysis may not be advisable, however, for example: a small number of studies may mean that statistical heterogeneity is underestimated; some studies are too biased to draw a conclusion from them; there is evidence of publication bias; and insufficient reporting of outcomes.

We acknowledge that, whilst software to undertake pairwise meta-analysis is widely available (e.g. Revman, Comprehensive meta-analysis), analyses, such as the most complex ones above, require advanced statistical expertise in evidence synthesis to implement. Our software package of choice is WinBUGS. This is a freely available Bayesian simulation package [37] and is extremely powerful for fitting models not immediately available in other packages (it even allows economic decision models to be included in the same program as the synthesis model allowing a truly comprehensive assessment [38]). With the recent publication of the NICE technical support documents on evidence synthesis methodology [32, 39, 40] including all WinBUGS code to implement the models, together with more widely available specialist training courses and the new introductory WinBUGS book[41] , the time is ripe for getting to grips with the more complex evidence synthesis methodologies currently being embraced by health technology appraisals[1, 42].WinBUGS Code for all our methods papers cited here is available in the papers or on request but the use of WinBUGS requires care and users are advised to acquire a good understanding of the theory and to follow advice given in the WinBUGS manual, and these books and papers, in terms of model checking and sensitivity analyses.

This paper is limited to only considering NICE PH appraisals in the review and does not claim to have all the answers to all evidence synthesis challenges that exist in PH evaluation. For example, none of the above considers directly the influence of the study quality/validity of the individual studies going into an analysis – although others are doing work in other contexts that could be adapted, for example including different, both observational and randomised, evidence [43].

Regarding the specific injury prevention context, for the analyses presented above, even when categorising the interventions into seven distinct groups, there is still residual heterogeneity in intervention definition; for example, education may be a leaflet designed for the prevention of an injury in the home, it may also include a face-to-face interview, or a computer questionnaire producing tailored advice based on the user answers etc. We are developing further modelling extensions including how to extrapolate across a series of evidence networks to allow information sharing on the effectiveness of interventions in promoting other safety practices for the prevention of falls. We hope such analyses will be more efficient and robust than individual analyses of each outcome. Note that all the data considered above only relates to an increase in the *uptake of safety practice* not to reduction in accidents *per se*. Therefore, a further initiative is to develop models which extend those presented to include the direct evidence between safety practices and injury data. This problem is similar to the use of surrogate endpoints in clinical evaluation and we plan to adapt methods developed there.

Public health evaluations are notoriously messy and complex, with many factors to consider. But if a decision has to be made, explicit, transparent and appropriate analysis of the data should be preferred to current alternatives. Just as evaluations of clinical interventions are becoming more sophisticated, we think there is a pressing need to do the same for PH contexts and we hope this paper can contribute to the initiation of such an initiative.

**Funding:**

This study presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10231). The

views expressed are those of the author(s) and are not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health, United Kingdom.

## REFERENCES

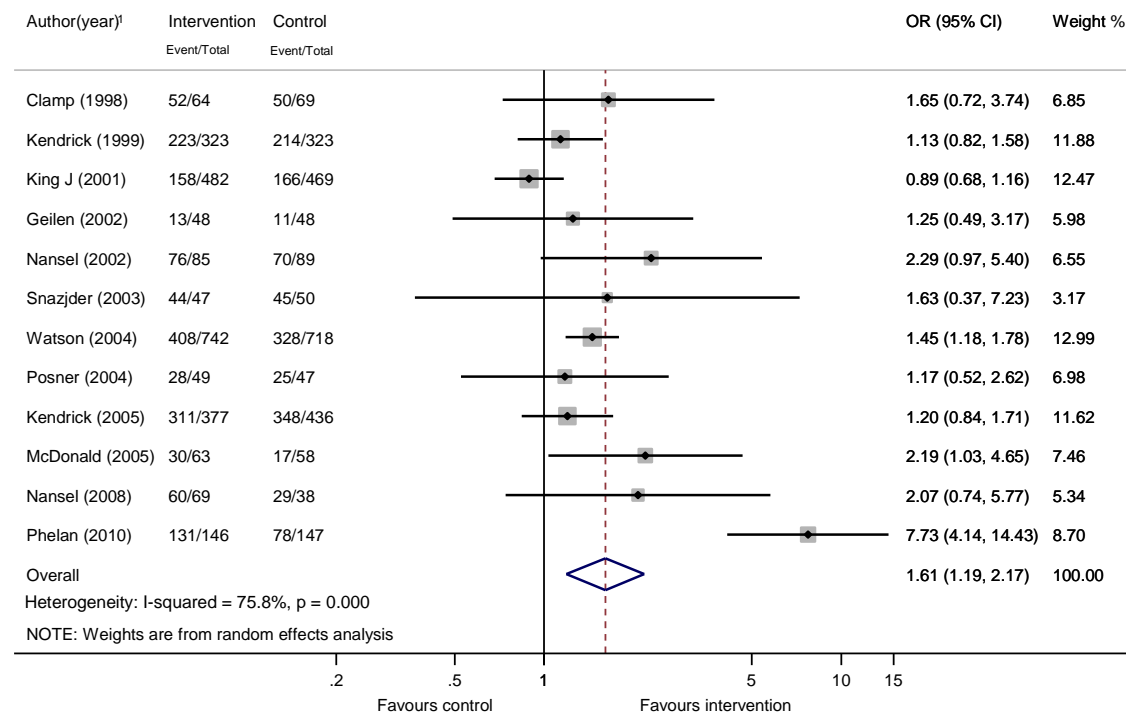
1. Cooper, NJ, J Peters, MCW Lai, P Juni, S Wandel, S Palmer, M Paulden, S Conti, NJ Welton, KR Abrams, S Bujkiewicz, D Spiegelhalter, and AJ Sutton, *How Valuable are Multiple Treatment Comparison Methods in Evidence-Based Health-Care Evaluation?* Value in Health, 2011. **14**(2): p. 371-380.
2. Novielli, N, NJ Cooper, KR Abrams, and AJ Sutton, *How Is Evidence on Test Performance Synthesized for Economic Decision Models of Diagnostic Tests? A Systematic Appraisal of Health Technology Assessments in the UK Since 1997.* Value in Health, 2010. **13**(8): p. 952-957.
3. Higgins, JPT and SE Green, *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration 2011.
4. Drummond, MF, M Sculpher, GW Torrance, B O'Brien, and GL Stoddart, *Methods for the Economic Evaluation of Health Care Programmes* 3ed. 2005, Oxford: Oxford University Press
5. Sutton, AJ, S Donegan, Y Takwoingi, P Garner, C Gamble, and A Donald, *An encouraging assessment of methods to inform priorities for updating systematic reviews.* Journal of Clinical Epidemiology, 2009. **62**(3): p. 241-251.
6. Eden, J, L Levit, A Berg, and S Morton, *Finding What Works in Health Care: Standards for Systematic Reviews.* Available from [http://www.nap.edu/catalog.php?record\\_id=13059](http://www.nap.edu/catalog.php?record_id=13059), 2012.
7. Rodgers, M, A Sowden, M Petticrew, L Arai, H Roberts, N Britten, and J Popay, *Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews.* Evaluation, 2009. **15**(1): p. 49-73.
8. Weatherly, H, M Drummond, K Claxton, R Cookson, B Ferguson, C Godfrey, N Rice, M Sculpher, and A Sowden, *Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations.* Health Policy, 2009. **93**(2-3): p. 85-92.
9. Armstrong, R, E Waters, and Je Doyle, *Chapter 21: Reviews in health promotion and public health.* In Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic reviews of Interventions*. 2008, Chichester (UK): John Wiley & Sons.
10. Rosen, L, O Manor, D Engelhard, and D Zucker, *In defense of the randomized controlled trial for health promotion research.* American Journal of Public Health, 2006. **96**(7): p. 1181-1186.
11. Petticrew, M and M Egan, eds. *Relevance, rigour and systematic reviews. Moving beyond effectiveness in evidence synthesis Methodological issues in the synthesis of diverse sources of evidence* [http://www.nice.org.uk/niceMedia/docs/Moving\\_beyond\\_effectiveness\\_in\\_evidence\\_synthesis2.pdf](http://www.nice.org.uk/niceMedia/docs/Moving_beyond_effectiveness_in_evidence_synthesis2.pdf) (accessed 06 June 2012). ed. J. Popay. 2006.
12. Wanless, D, *Securing good health for the whole population: Final report - February 2004 (Chapter 6).* [http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_4074426](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4074426) (accessed 06 June 2012), 2004.
13. Chalkidou, K, A Culyer, B Naidoo, and P Littlejohns, *Cost-effective public health guidance: Asking questions from the decision-maker's viewpoint.* Health Economics, 2008. **17**(3): p. 441-448.
14. NICE, *Methods for the development of NICE public health guidance.* Available from <http://www.nice.org.uk/media/FB9/59/PHMethodsManual2006.pdf>, 2006.

15. NICE, *Methods for the development of NICE public health guidance (second edition)*. Available from [http://www.nice.org.uk/media/CE1/F7/CPHE\\_Methods\\_manual\\_LR.pdf](http://www.nice.org.uk/media/CE1/F7/CPHE_Methods_manual_LR.pdf), 2009.
16. NICE, *Methods for the development of NICE public health guidance (third edition)*. Available from <<http://publications.nice.org.uk/methods-for-the-development-of-nice-public-health-guidance-third-edition-pmg4>>, 2012.
17. Dickersin, K, S Chan, TC Chalmers, HS Sacks, and H Smith, *Publication Bias and Clinical-Trials*. Controlled Clinical Trials, 1987. **8**(4): p. 343-353.
18. Sutton, A, K Abrams, DR Jones, TA Sheldon, and F Song, *Methods for Meta-analysis in Medical Research*. 2000, Chichester: Wiley.
19. Caldwell, DM, AE Ades, and JP Higgins, *Simultaneous comparison of multiple treatments: combining direct and indirect evidence*. BMJ, 2005. **331**(7521): p. 897-900.
20. Caldwell, DM, NJ Welton, and AE Ades, *Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency*. Journal of Clinical Epidemiology, 2010. **63**(8): p. 875-882.
21. Ades, AE, I Mavranetzouli, S Dias, NJ Welton, C Whittington, and T Kendall, *Network meta-analysis with competing risk outcomes*. Value Health, 2010. **13**(8): p. 976-83.
22. Lumley, T, *Network meta-analysis for indirect treatment comparisons*. Stat Med, 2002. **21**(16): p. 2313-24.
23. Kendrick, D, B Young, AJ Mason-Jones, N Ilyas, FA Achana, NJ Cooper, SJ Hubbard, AJ Sutton, S Smith, P Wynn, C Mulvaney, MC Watson, and C Coupland, *Home safety education and provision of safety equipment for injury prevention*. Cochrane Database of Systematic Reviews, 2012(9).
24. Kendrick, D, A Sutton, S Smith, C Mulvaney, A Woods, M Watson, C Coupland, and J Simpson, *Does the effect of home safety education, with or without the provision of safety equipment, differ by social group? Systematic review and meta-analysis. Report to the Department of Health.*, 2006.
25. Higgins, JPT, SG Thompson, JJ Deeks, and DG Altman, *Measuring inconsistency in meta-analyses*. British Medical Journal, 2003. **327**(7414): p. 557-560.
26. Lambert, PC, AJ Sutton, KR Abrams, and DR Jones, *A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis*. Journal of Clinical Epidemiology, 2002. **55**(1): p. 86-94.
27. Stewart, LA, *Practical methodology of meta-analyses (overviews) using updated individual patient data*. Statistics in Medicine, 1995. **14**(19): p. 2057-2079.
28. Kendrick, D, C Coupland, C Mulvaney, J Simpson, SJ Smith, A Sutton, M Watson, and A Woods, *Home safety education and provision of safety equipment for injury prevention*. Cochrane Database Syst Rev, 2007(1): p. CD005014.
29. Sutton, AJ, D Kendrick, and CAC Coupland, *Meta-analysis of individual- and aggregate-level data*. Statistics in Medicine, 2008. **27**(5): p. 651-669.
30. Riley, RD, PC Lambert, and G Abo-Zaid, *Meta-analysis of individual participant data: rationale, conduct, and reporting*. British Medical Journal, 2010. **340**.
31. Dias, S, AJ Sutton, AE Ades, and NJ Welton, *A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials*. Medical Decision Making, 2012.
32. Dias, S, NJ Welton, AJ Sutton, and AE Ades, *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network*



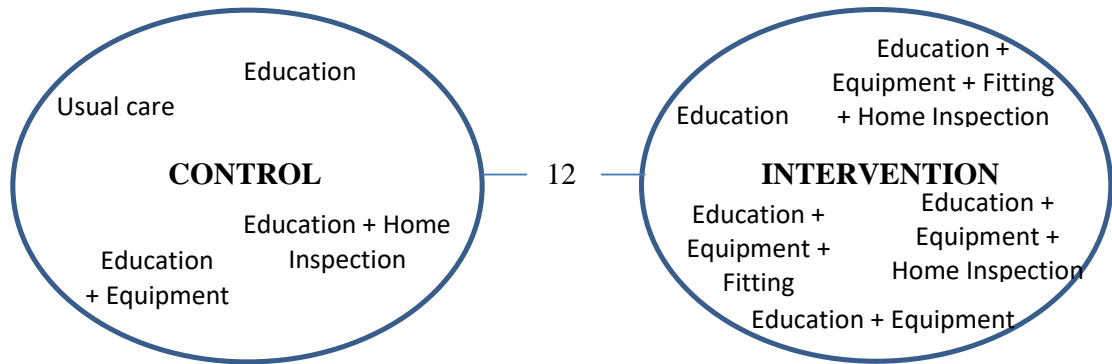
- Meta-analysis of Randomised Controlled Trials: Available from <http://www.nicedsu.org.uk>, 2011.*
33. Hubbard, S, A Sutton, D Kendrick, B Young, F Achana, and N Cooper, *Network Meta-analysis to evaluate the effectiveness of interventions to prevent falls injuries in children under 5*, 2013, University of Leicester, UK.
  34. Cooper, NJ, D Kendrick, F Achana, P Dhiman, Z He, P Wynn, E Le Cozannet, P Saramago, and AJ Sutton, *Network Meta-analysis to Evaluate the Effectiveness of Interventions to Increase the Uptake of Smoke Alarms*. Epidemiologic Reviews, 2011.
  35. Achana, FA, NJ Cooper, S Dias, GB Lu, SJC Rice, D Kendrick, and AJ Sutton, *Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis*. Statistics in Medicine, 2013. **32**(5): p. 752-771.
  36. Saramago, P, AJ Sutton, NJ Cooper, and A Manca, *Mixed treatment comparisons using aggregate and individual participant level data*. Statistics in Medicine, 2012. **31**(28): p. 3516-3536.
  37. Lunn, DJ, A Thomas, N Best, and D Spiegelhalter, *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility*. Statistics and Computing, 2000. **10**(4): p. 325-337.
  38. Cooper, NJ, KR Abrams, AJ Sutton, D Turner, and PC Lambert, *A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2003. **166**(3): p. 389-405.
  39. Dias, S, AJ Sutton, NJ Welton, and AE Ades, *NICE DSU Technical Support Document 3: Heterogeneity: Subgroups, Meta-regression, Bias and Bias-Adjustment: Available from <http://www.nicedsu.org.uk>, 8 November 2011*, 2011.
  40. Dias, S, NJ Welton, AJ Sutton, and AE Ades, *NICE Technical Support Document 4: Inconsistency in networks of evidence based on randomised trials: Available from <http://www.nicedsu.org.uk>, 2011.*
  41. Lunn, D, C Jackson, B Best, A Thomas, and D Spiegelhalter, *The BUGS Book - A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC Texts in Statistical Science, ed. C. Hall. 2012.
  42. Cooper, NJ, D Spiegelhalter, S Bujkiewicz, P Dequen, and A Sutton, *A systematic review of the deliberate and accidental use of Bayesian methods in health technology assessment*. . International Journal of Technology Assessment in Health Care (In press), 2013.
  43. Turner, RM, DJ Spiegelhalter, GCS Smith, and SG Thompson, *Bias modelling in evidence synthesis*. Journal of the Royal Statistical Society Series a-Statistics in Society, 2009. **172**: p. 21-47.

**Figure 1: Forest plot illustrating the findings of random effects meta-analysis of interventions aimed at increasing the uptake of safety equipment for the outcome “possession of a fitted stair gate”**

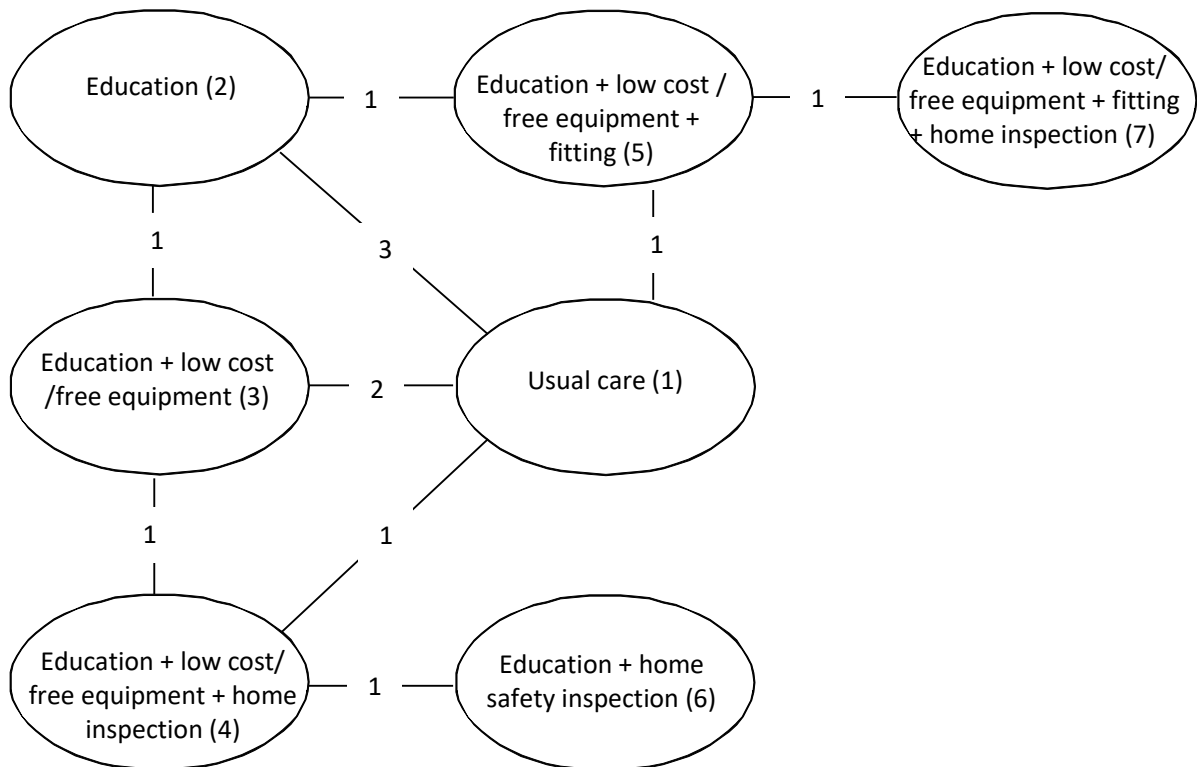


<sup>1</sup>References for all the studies can be found in the updated Cochrane review [23]

a)



b)



**Figure 2: Network diagrams indicating how intervention groups were defined and the number of studies in the a) Cochrane review and b) network meta-analysis**

**Table 1: NICE public health appraisals and summary of evidence synthesis methods and decision modelling used to inform their development.**

NICE public health appraisal title	Review of the effectiveness and cost-effectiveness evidence and decision analysis used to inform each appraisal					
	Systematic review of effectiveness (narrative summary)	Systematic review of effectiveness (At least one M-A) <sup>†</sup>	Cost-effectiveness reviews	Study quality	Decision Model	Source of effectiveness estimate used in decision model <sup>‡</sup>
Brief interventions and referral for smoking cessation (PH1)	✓	x <sup>nr</sup>	✓	✓	✓	Published review
Four commonly used methods to increase physical activity (PH2)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>1</sup>
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	✓	✓	✓	✓	✓	Published review
Interventions to reduce substance misuse among vulnerable young people (PH4)	✓	x <sup>o</sup>	✓	✓	✓	Individual study <sup>1</sup>
Workplace interventions to promote smoking cessation (PH5)	✓	x <sup>m</sup>	✓	✓	✓	Individual study <sup>3</sup>
Behaviour change (PH6)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>4</sup>
School-based interventions on alcohol (PH7)	✓	✓	✓	✓	✓	Individual study <sup>1</sup>
Physical activity and the environment (PH8)	✓	x <sup>i,m,o</sup>	✓	✓	✓	Individual study <sup>3</sup>
Community engagement (PH9)	✓	x <sup>i,m,o,p</sup>	✓	✓	x	Not applicable
Smoking cessation services (PH10)	✓	✓	✓	✓	✓	New Meta-analysis
Maternal and child nutrition (PH11)	✓	x <sup>m</sup>	✓	✓	✓	Individual study <sup>5</sup>
Social and emotional wellbeing in primary education (PH12)	✓	✓	✓	✓	✓	Individual study <sup>5</sup>
Promoting physical activity in the workplace (PH13)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>5</sup>
Preventing the uptake of smoking by children and young people (PH14)	✓	x <sup>m,o</sup>	✓	✓	✓	Individual study <sup>1</sup>
Identifying and supporting people most at risk of dying prematurely (PH15)	✓	x <sup>i,m,p</sup>	✓	✓	✓	Individual study <sup>1</sup>
Mental wellbeing and older people (PH16)	✓	x <sup>i,m,o</sup>	✓	✓	✓	Individual study <sup>1</sup>
Promoting physical activity for children and young people (PH17)	✓	x <sup>i,m,o</sup>	✓	✓	✓	Analyst estimate <sup>4</sup>
Needle and syringe programmes (PH18)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>3</sup>
Management of long-term sickness and incapacity for work (PH19)	✓	✓	✓	✓	✓	New meta-analysis
Social and emotional wellbeing in secondary education (PH20)	✓	x <sup>i,m,o</sup>	✓	✓	✓	Individual study <sup>2</sup>
Reducing differences in the uptake of immunisations (PH21)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>4</sup>
Promoting mental wellbeing at work (PH22)	✓	x <sup>i</sup>	✓	✓	✓	Individual study <sup>3</sup>
School-based interventions to prevent smoking (PH23)	✓	✓	✓	✓	✓	New Meta-analysis
Alcohol-use disorders - preventing harmful drinking (PH24)	✓	x <sup>nr</sup>	✓	✓	✓	Published review
Prevention of cardiovascular disease (PH25)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>5</sup>
Quitting smoking in pregnancy and following childbirth (PH26)	✓	x <sup>i,m,o</sup>	✓	✓	✓	Published review

Weight management before, during and after pregnancy (PH27)	✓	✓	✓	✓	✓	Not clear <sup>5</sup> reported
Looked-after children and young people (PH28)	✓	x <sup>m,o,p</sup>	✓	✓	✓	Individual study <sup>3</sup>
Strategies to prevent unintentional injuries among under-15s (PH29)	✓	x <sup>i</sup>	✓	✓	✓	Individual study <sup>3</sup>
Preventing unintentional injuries among under-15s in the home (PH30)	✓	x <sup>i,o</sup>	✓	✓	✓	Individual study <sup>3</sup>
Preventing unintentional road injuries among under-15s: road design (PH31)	✓	x <sup>i</sup>	✓	✓	✓	Individual study <sup>2</sup>
Skin cancer prevention: information, resources and environmental changes (PH32)	✓	x <sup>i,m</sup>	✓	✓	✓	Individual study <sup>3</sup>
Increasing the uptake of HIV testing among black Africans in England (PH33)	✓	x <sup>m</sup>	✓	x	x	Not applicable
Increasing the uptake of HIV testing among men who have sex with men (PH34)	✓	x <sup>nr</sup>	✓	x	x	Not applicable
Preventing type 2 diabetes - population and community interventions (PH35)	✓	✓	✓	✓	✓	New meta-analysis
Prevention and control of healthcare-associated infections (PH36)		x	x	x	x	Not applicable
Tuberculosis - hard-to-reach groups (PH37)	✓	x <sup>nr</sup>	✓	✓	✓	Individual study <sup>5</sup>
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	✓	✓	✓	✓	✓	New Meta-analysis
Smokeless tobacco cessation - South Asian communities (PH39)	✓	x <sup>s</sup>	✓	✓	✓	Published review

Ticks indicates a systematic review of evidence, meta-analysis or decision model have been conducted whilst x indicate analysis have not been conducted.

<sup>†</sup>Reported reason why meta-analysis was not done (i=heterogeneity of interventions, m=heterogeneity methods, design and settings, o= heterogeneity of outcome measures, p= heterogeneity of study populations, s= heterogeneity of studies (specific cause not stated), nr= not reported or reported that studies do not support a meta-analysis).

<sup>‡</sup>Selection of individual study estimate of intervention effect used in decision model (1= used a pre-specified criteria reported in the decision model report, 2=discussion with NICE or estimates selected based on quality grading of evidence using the methods guide manuals, 3= selected studies based on relevance of the intervention to the decision problem, 4= assumption/analyst estimated based on an assumption, 5 =not clearly reported).

**Table 2: Review of quantitative methods used to synthesise public health evidence for NICE public health appraisal**

Appraisal title	Systematic review report title	Included RCTs only	Main outcome	Description of main outcome	Outcome measure: statistic	Type of synthesis	Model type	Lumping <sup>1</sup> of interventions	Presentat ion of results	Assessed publication bias	Software	Used result of M-A in decision model
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	Review 2 - Review of evidence for the effectiveness of screening for genital chlamydial infection in sexually active young women and men	No	Intermediate	Uptake of proactive chlamydia screening using home-collected specimens	Screening response rate (%)	M-A	Random effects	No	FP/Txt	No	RevMan, Stata	No
School-based interventions on alcohol (PH7)	Alcohol and schools: effectiveness and cost-effectiveness review	No	Final	Alcohol use	Weighted mean difference	M-A	Random effects	Yes	FP/Txt	No	Not stated	No
Smoking cessation services (PH10)	Cut down to quit' with nicotine replacement therapies	Yes	Final	6 or more months' sustained abstinence	Relative risk & Cohen's d	M-A	Random effects	Yes	FP/T/ Txt	No	RevMan	Yes
Smoking cessation services (PH10)	Final report	No	Final	6 or months' sustained abstinence	Cohen's d	M-A	Fixed & random effects	Yes	FP/T/ Txt	No	RevMan	No
Social and emotional wellbeing in primary education (PH12)	Teesside review	Yes	Intermediate	Social problem solving	Standardised mean difference	M-A	Random effects	Yes	FP/T	No	RevMan	No

Management of long-term sickness and incapacity for work (PH19)	PH19 Management of long-term sickness and incapacity for work: Economic analysis report	No	Yes	Number returning to work following sickness	Relative risk	M-A	Random	Yes	FP/T/Txt	No	Revman	Yes
School-based interventions to prevent smoking (PH23)	School-based interventions to prevent smoking: quantitative effectiveness review	Yes	Final	smoking uptake	Odds ratio	M-A	Random effects	Yes	FP/Txt	Yes	Stata	Yes
Weight management before, during and after pregnancy (PH27)	Weight management before, during and after pregnancy: evidence review	No	Intermediate	Number exceeding loM <sup>2</sup> guidelines for healthy weight gain	Relative risk	M-A	Random effects	Yes	FP/T/ Txt	No	RevMan	No
Preventing type 2 diabetes - population and community interventions (PH35)	PH35 Preventing type 2 diabetes - population and community interventions: report on cost-effectiveness evidence and methods for economic modelling	No	Intermediate	Body mass index	Weighted mean difference	M-A	Not reported	Yes	T/Txt	No	Not reported	Yes
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	Prevention of type 2 diabetes: systematic review & meta-analysis of lifestyle, pharmacological and surgical interventions	Yes	Final	Reduce progress to diabetes for people with IGT	Hazard ratio	M-A & NMA	Random effects	No	FP/TxT	No	RevMan (M-A) WinBUGS (NMA)	Yes

Presentation of results (FP = Forest plot, T=Table, Txt=Text), M-A = pairwise meta-analysis, NMA = network meta-analysis;

1 = lumping is a term used in the literature[19, 20] to described the tendency to aggregate or treat seemingly similar but disparate /different interventions as one intervention group in order for example to facilitate inclusion of many studies in a meta-analysis. A classic example is treating different doses of a drug as if they were the same treatment

2 = American Institute of Medicine (IOM) Guidelines on Weight Management in Pregnancy



**Table 3: Results of a network meta-analysis (above stepped line) and pairwise meta-analysis (below stepped line) for possession of a fitted stair gate expressed as Odds Ratios (95% CrI)<sup>a,b</sup>**

	Usual care (1)	Education (2)	Education + Equipment (3)	Education + Equipment + Home inspection (4)	Education + Equipment + Fitting (5)	Education + Home inspection (6)	Education + Equipment + Fitting + Home inspection (7)
<b>Usual care (1)</b>		1.43	1.63	1.28	1.52	1.43	7.80*
		(0.90 , 2.49)	(0.93 , 3.03)	(0.69 , 2.79)	(0.84 , 3.38)	(0.56 , 4.42)	(3.08 , 21.3)
<b>Education (2)</b>	1.48		1.14	0.90	1.07	1.01	5.46*
	(0.97 , 2.25)		(0.56 , 2.23)	(0.41 , 2.07)	(0.51 , 2.41)	(0.33 , 3.25)	(1.75 , 16.12)
<b>Education + Equipment (3)</b>	1.92*	1.17		0.78	0.94	0.88	4.77*
	(1.05 , 3.51)	(0.52 , 2.63)		(0.38 , 1.77)	(0.42 , 2.41)	(0.32 , 2.80)	(1.56 , 15.18)
<b>Education + Equipment + Home inspection (4)</b>	1.13		1.25		1.20	1.12	6.13*
	(0.82 , 1.58)		(0.49 , 3.17)		(0.45 , 3.25)	(0.52 , 2.49)	(1.75 , 18.71)
<b>Education + Equipment + Fitting (5)</b>	1.45*	1.63				0.94	5.07*
	(1.18 , 1.79)	(0.37 , 7.23)				(0.27 , 3.28)	(1.47 , 15.93)
<b>Education + Home inspection (6)</b>				1.12			5.48*
				(0.86 , 1.47)			(1.23 , 20.73)
<b>Education + Equipment + Fitting + Home inspection (7)</b>	7.73*						
	(4.14 , 14.43)						

Abbreviations: CrI, credible interval; NMA, network meta-analysis; OR, odds ratio.

<sup>a</sup> values above the stepped line are results from the NMA; those below the line are direct estimates from a trial or, where more than one was available, a meta-analysis.

Blank cells indicate that no direct evidence on specific pairwise comparisons was available.

<sup>b</sup> Column and row headings signify intervention or comparison (intervention number)

\* significant at the 5% level

**Table 4: Assessment of which intervention is best for possession of a fitted stair gate**

Intervention	Possession of a stair gate	
	Probability treatment is best	Median treatment rank (95% CrI)
Usual care (1)	0.00	7(5, 7)
Education (2)	0.002	4(2, 7)
Education + Equipment (3)	0.004	3(2, 7)
Education + Equipment + Home inspection (4)	0.001	5(2, 7)
Education + Equipment + Fitting (5)	0.008	4(2, 7)
Education + Home inspection (6)	0.013	4(2, 7)
Education + Equipment + Fitting + Home inspection (7)	0.97	1(1, 2)