NHS

Health Development Agency

Grading evidence and recommendations for public health interventions: developing and piloting a framework

Alison Weightman[1], Simon Ellis[2], Adrienne Cullum[2], Lesley Sander[1] and Ruth Turley[1]

[1]Support Unit for Research Evidence (SURE), Information Services, Cardiff University

[2]Health Development Agency, London

This document is also published on the Health Development Agency website at www.hda.nhs.uk

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

Health Development Agency
Holborn Gate
330 High Holborn
London
WC1V 7BA

Email: communications@hda.nhs.uk

ISBN 1-84279-458-2

This work was undertaken by SURE jointly with the Health Development Agency (HDA). SURE forms part of the Wales Collaborating Centre, one of two HDA Evidence and Guidance Collaborating Centres on Obesity.

From 1 April 2005, the functions of the HDA will transfer to the National Institute for Clinical Excellence. The new organisation will be the National Institute for Health and Clinical Excellence (to be known as NICE). It will be the independent organisation responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health.

Contents

Summary

The objective of this work was to develop a practical scale of grades of recommendation for public health interventions, adapted from the current National Institute for Clinical Excellence (NICE) methodology.

A literature review was carried out on the subject of incorporating research evidence into grades of recommendation for public health interventions. The literature search looked at publications from January 2000–May 2004 retrieved from 16 databases. The views of a range of public health experts were also sought for suggestions of other publications to be included in the literature review, and for their comments at various stages of the developing methodology.

The principles for development of the framework were that it should be:

- Adapted from, and clearly linked to, the current NICE methodology
- Based on detailed and transparent reporting and synthesis of *all* relevant supporting evidence (intervention and observation; quantitative and qualitative).

The literature review indicated general agreement that the randomised controlled trial (RCT) has the highest internal validity and, where feasible, is the research design of choice when evaluating effectiveness. However, many commentators felt the RCT may be too restrictive for some public health interventions, particularly community-based programmes. In addition, supplementing data from quantitative studies with the results of qualitative research is regarded as key to the successful replication and ultimate effectiveness of interventions.

Based on the literature review and consultation with experts, a framework was developed that derives grades of recommendation, incorporating:

- Strength of evidence of efficacy based on the research design and the quality and quantity of evidence (the current NICE system)
- Corroborative evidence (from observational and qualitative studies) for the feasibility and likelihood of success of an intervention if implemented in the UK.

The precise methods for combining the results from different types of corroborative evidence and for incorporating the size of effects, including (cost–)benefits and harms for the different outcomes measured, are still in development.

This provisional framework provides a practical and transparent method for deriving grades of recommendation for public health interventions, based on a synthesis of all relevant supporting evidence from research. The methodology is being piloted, alongside the current NICE methodology, within the development of the public health/prevention aspects of the HDA/NICE guidance on overweight and obesity. The lessons learned will help to inform the forthcoming work of the National Institute for Health and Clinical Excellence.

Introduction

In 2003 NICE and the HDA were commissioned by the Department of Health and the National Assembly for Wales to develop guidance on the prevention and management of obesity in children and adults. This was the first time NICE had been tasked to work in collaboration with an external body, and pre-empted the announcement that NICE will take on the functions of the HDA from April 2005. Crucially, it was also the first time that the applicability of existing NICE methodology to public health evidence and recommendations was to be fully considered.

Where possible, the development of the guidance was to adhere to procedures laid down by NICE. However, due to the nature of public health interventions and the associated evidence base, it became clear that further consideration would be needed in adapting the NICE methodology.

The NICE guidelines to date have been based on a well known hierarchy of research designs (NICE, 2004a,b; SIGN, 2001 and website), from which recommendations have been developed for clinical policy and practice. A parallel scale for grading evidence and recommendations for public health policy and practice does not exist at present. NICE is currently developing some broad principles for the methods used to assess evidence and prioritise recommendations that may be applied across all types of question, leading to both clinical and public health recommendations.

In some cases the 'gold standard' RCT cannot be performed in public health interventions for feasibility, cost and practical reasons (Wanless, 2004; Kelly *et al.*, 2005). Furthermore, RCTs tend to be limited to questions of efficacy or effectiveness; they are less useful, and hence less appropriate, when considering external validity and issues of implementation. For example, some public health interventions cannot readily be abstracted from their environment, making context very important. Thus reviews of evidence for public health interventions tend to be dominated by 'lower' levels of evidence, which will in turn receive lower grades of recommendation.

Clearly, a range of grades of recommendation is appropriate to provide guidance for policy makers in deciding which public health interventions might be considered for practice and/or further research. These grades should reflect the (theoretically) most appropriate evidence for the type of intervention, using a clear and transparent methodology.

The objective was therefore to develop a practical public health scale of grades of recommendation adapted from the current NICE methodology. The framework was to relate only to the grading of evidence and recommendations for public health *interventions*. (The types of evidence that are relevant to other (non-intervention) aspects of public health will be included in further developments of the methodology.) Development of the framework has incorporated an analysis of the published literature on deriving grades of evidence and recommendations for public health interventions, and consultation with public health and methodology experts.

Methodology and results

The methodology was designed to answer the following research questions:

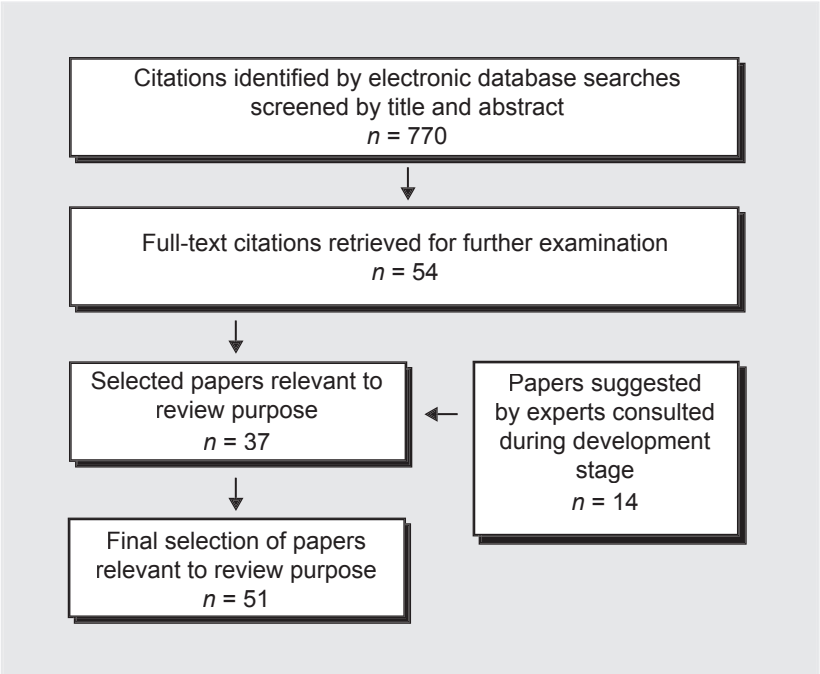
- What are the most appropriate research designs for determining the efficacy of public health interventions?
- How might qualitative research and data about implementation be used to assess whether an intervention is likely to work in the UK?
- How can these different types of evidence be combined to give a grading for public health evidence and help prioritise recommendations?

There were three elements to the development of the framework:

- Literature review
- Consultation with individuals and organisations with expertise in public health and/or grading methodology
- Piloting of the provisional framework.

The methodology and results for each are described below. This was an iterative process – for instance, the consultation with experts at various stages identified further publications for inclusion in the review and other experts to consult. Early versions of the framework formed part of the consultation with experts.

Figure 1 Selection stages for papers included in the review



Box 1 Search strategy

Databases searched
ASSIA, CareData, CINAHL, Cochrane Library, Current Contents, Educational Resources Information Center (ERIC), Embase, EPPI Centre, HDA Evidence Base, HDA HealthPromis, Health Management Information Consortium (HMIC), MEDLINE, PsycINFO, Sociological Abstracts, System for Information on Grey Literature (SIGLE), ZETOC.

Standard search terms
(public health OR health of the public OR health promotion) AND (grade* or level* or type*) AND ((guideline* or guidance or evidence or recommendation*).ti)

Additional search terms
Additional terms were used for the databases ASSIA, CareData, EPPI Centre, ERIC and SIGLE where complex search strategies were not feasible. These are available from the authors.

Website searches
Website searches were conducted for relevant organisations involved in searching and summarising evidence for public health, looking for methodologies for grading public health/health promotion recommendations and reviews/guidelines in the topic area. See Appendix 2 for organisations searched.

Search dates
2000–2004 plus follow-up of reference lists for other relevant publications. The searches were carried out in May 2004.

Literature review

An extended literature review was carried out on grading evidence and recommendations (see Box 1).

Selection and appraisal of relevant publications

Of the 770 abstracts/titles retrieved, 54 publications were examined in full and 37 were found to be relevant to the review question. Other papers were suggested by groups and individuals consulted during the development stages, and 14 additional papers were included, making a total of 51 (see Figure 1). They are marked ✓ in the References (page 17).

Publications were selected for full text review if the abstract (or title if no abstract was available) suggested that the paper included a discussion of the methodology for translating findings from public health research evidence into grades of recommendation for interventions. Papers were read and summarised by one reviewer to determine the authors' views on the most appropriate type(s) of public health evidence that should be taken into account when generating recommendations. Any areas of uncertainty were clarified through discussion with a second reviewer. The purpose was to assess areas of consensus and query. No formal evaluation of the publications included was carried out.

Results of the literature review

Type of evidence (research design)

An established evidence hierarchy of effectiveness is used by NICE, and this has a strong link to the grade of recommendation (NICE, 2004a,b).

The issue of the 'best' evidence for particular types of intervention (individual, group, community, society/socio-political) has been considered by Nutbeam (1998) and by the HDA (Ellis and Grey, 2004). There is general agreement that the RCT has the highest internal validity and, where feasible, is the research design of choice when evaluating effectiveness (Nutbeam, 1998; Kelly *et al.*, 1993; Sorensen *et al.*, 1998; Rimer *et al.*, 2001; Rychetnik *et al.*, 2002; Evans, 2003; Hawe *et al.*, 2004; Victora *et al.*, 2004).

However, RCTs are by nature narrowly defined and typically restricted to single/simple issues (Nutbeam, 1998; Tones, 2000; Truswell, 2001; Kroke *et al.*, 2004; Victora *et al.*, 2004). It is argued that, because of the complexity of interventions in real world settings, RCTs are subject to effect modification in different populations (Victora *et al.*, 2004) and in any event may be too restrictive for community-based programmes (Nutbeam, 1998). Many health promotion programmes draw on political systems and community networks as part of the intervention, rendering random allocation nearly impossible. However, in some circumstances, the design in which geographically isolated populations become the (randomly allocated) units of comparison (ie the cluster RCT) may be appropriate and feasible (Nutbeam, 1998; Sorensen *et al.*, 1998; Rychetnik *et al.*, 2002).

Other commentators are of the view that the RCT design can be appropriate for evaluating complex public health interventions by standardising the function and process of the intervention, but allowing local variations in the individual components. This allows the components to be tailored to local conditions and the needs of specific communities, without threatening the integrity of the intervention (Dane and Schneider, 1998; Hawe *et al.*, 2004).

The importance of supplementing data from quantitative studies with the results of qualitative research to provide depth and insight into people's experiences and social contexts is regarded as central by many commentators (Nutbeam, 1998; Sorensen *et al.*, 1998; Stephenson and Imrie, 1998; Rychetnik *et al.*, 2002; Pawson, 2003; Petticrew and Roberts, 2003; Dixon-Woods *et al.*, 2004; Harden *et al.*, 2004; Jackson and Waters, 2004; Kroke *et al.*, 2004; NSW Centre for Public Health Nutrition, 2004; Swinburn *et al.*, 2004; Thomas *et al.*, 2003, 2004), and of particular relevance to the successful replication and sustainability of interventions. Using each subsequent study to build on the inferences of the others, the likely effectiveness of social programme interventions can be assessed (Pawson, 2003). In one proposed decision-making framework for evidence-based obesity prevention, the RCT sits alongside other forms of evidence and each is judged equally on its ability to contribute to answering different questions (Swinburn *et al.*, 2004).

In a review of children and healthy eating (Thomas *et al.*, 2003) the EPPI Centre cross-matched the findings of qualitative and quantitative studies and looked at interventions based on components matching children's views. The reviewers found a relationship between what children regarded as important and the effectiveness of the intervention.

Petticrew and Roberts (2003) note that RCTs are best for questions of effectiveness (*does it work?*), safety and cost effectiveness; qualitative studies and surveys are best for questions of salience (*does it matter?*), appropriateness and satisfaction; and qualitative studies alone are best for questions concerning process (*how does it work?*) and acceptability. Of course a single study, particularly but not exclusively a systematic review, may provide evidence for several or all of the individual elements – effectiveness, salience, implementation and cost.

Some methodologies consider the type of evidence as one of many factors, and use a single quality assessment/critical appraisal tool for all studies to produce gradings, based on minimisation of potential biases, from poor/weak to good/strong (Millward *et al.*, 2003; EPHPP website).

Consistency

The consistency of study results contributes to the grades of recommendation used in the methodologies of NICE (2004a,b), the GRADE Working Group (2004) and others (Margetts *et al.*, 2001; Kelly *et al.*, 2004), and the importance of combining different study types is widely accepted.

Quality of evidence (critical appraisal)

Single critical appraisal forms are used by a number of public health groups (Briss *et al.*, 2000; Millward *et al.*, 2003; Øvretveit, 2003; EPHPP website), whereas clinical medicine review groups tend to use separate forms for each category of study type or research design (NICE, 2004a,b; Health Evidence Bulletins Wales 'Project methodology'; BMJ Publishing Group 'Clinical evidence'; SIGN 'Guidelines methodology'; Canadian Task Force for Preventive Health Care 'Evidence-based clinical prevention', Centre for Evidence-Based Medicine), other than the GATE method (University of Auckland (a)) which has a generic form for intervention studies.

The conclusions from a review of a large number of grading systems were that different appraisal forms are needed for different study types, and that a single evaluation framework could cause confusion and misleading conclusions (AHRQ, 2002).

Separate critical appraisal forms are used for different types of research study for NICE guidance. The overall assessment of study quality is graded within each study type using a code based on the extent to which the potential biases have been minimised (++, very low risk; +, low risk; –, high risk of confounding, bias or chance) (SIGN, 2001; NICE, 2004a,b). However, NICE does not currently have a critical appraisal form for non-randomised controlled studies, and specific enhancements and adjustments to NICE critical appraisal tools may be required for use with public health research evidence.

The Cochrane Effective Practice and Organisation of Care (EPOC) Review Group has developed critical appraisal forms for intervention studies that have not been adequately randomised and followed up in the randomised groups (EPOC Group, 2002). These are: (1) controlled clinical trials (sometimes called controlled non-randomised trials); (2) controlled before-and-after studies; (3) interrupted time-series studies. Another critical appraisal methodology for non-randomised trials is under development within the Cochrane Collaboration by the Health Promotion and Public Health Field (Jackson and Waters, 2004).

The TREND Group (Des Jarlais *et al.*, 2004) has also proposed a specific appraisal checklist for non-randomised evaluation studies as a companion to the CONSORT statement (Moher *et al.*, 2001) for RCTs.

Salience – does it matter?

Relevance of outcome

Several authors stress the importance of looking at a clearly defined and measured range of relevant health promotion and health outcomes in complex areas such as dietary behaviour and physical activity (for instance, Nutbeam, 1998; Lean, 2000). This includes an assessment of the relevant outcomes and most appropriate methods of evaluation for different types of intervention. Kelly *et al.* (1993) describe four levels of health promotion: environmental, social, organisational and individual, all of which have to be understood and integrated for successful health promotion interventions. It is emphasised that, from the outset of any health promotion project, these four levels should be used as a checklist to consider the likely consequences flowing from the desired intervention (Kelly *et al.*, 1993).

Ellis and Grey (2004) highlight that most reviews of effectiveness focus on health outcomes (eg incidence/prevalence) or intermediate health outcomes (eg behaviour), mainly because they are limited to RCTs and controlled trials which do likewise. These limitations in both the type of research included and the outcome mean that they are severely lacking in evidence about the effectiveness of community and socio-political interventions in addressing the personal and structural determinants of health and health behaviour (eg knowledge, social/peer norms, professional attitudes, discrimination, poverty, availability and accessibility of services). They are particularly unlikely to include any evidence about the effectiveness of 'upstream' (socio-political) interventions (Ellis and Grey, 2004).

Relevance to the UK population – demographic, personal and socio-economic factors

The context in which the intervention is implemented is clearly important. Relevance to the UK population contributes to the grades of recommendation used in the SIGN, NICE and GRADE methodologies (SIGN, 2001; NICE, 2004a,b; GRADE Working Group, 2004), but there is currently a lack of transparency in how this is derived. Specific consideration of socio-economic issues is recognised (Glasgow *et al.*, 1999; Kelly *et al.*, 2004; NSW Centre for Public Health Nutrition, 2004), as is the shortage of relevant evidence in this area (Aldrich *et al.*, 2003; Thomas *et al.*, 2003; Mulvihill and Quigley, 2003).

Implementation – will it work?

Consideration of issues such as feasibility, plausibility, acceptability, transferability and sustainability is suggested by the HDA (Ellis and Grey, 2004; Kelly *et al.*, 2004); the CDC Guide to Community Preventive Services (Briss *et al.*, 2000; Task Force on Community Preventive Services website); and other authors (Glasgow *et al.*, 1999; Evans, 2003; Jackson and Waters, 2004; Pawson *et al.*, 2004). It has previously been highlighted that an intervention should be based on firm theoretical principles using the knowledge of what is likely to work from previous research (Pawson, 2003; NSW Centre for Public Health Nutrition, 2004). In particular, reviewers should question whether the intervention is appropriate in relation to the views and preferences of the target population(s) (Thomas *et al.*, 2003; Pawson *et al.*, 2004).

As noted above, evidence from observational and qualitative research is considered central to informing the assessment of these issues (Nutbeam, 1998; Sorensen *et al.*, 1998; Stephenson and Imrie, 1998; Tones, 2000; Rychetnik *et al.*, 2002; Thomas *et al.*, 2003; Pawson, 2003; Petticrew and Roberts, 2003; Dixon-Woods *et al.*, 2004; NSW Centre for Public Health Nutrition, 2004; Pawson *et al.*, 2004). While there is a need for a transparent and reproducible approach, there is currently a lack of consensus as to how to grade this type of evidence.

Implementation: cost

Estimated cost is considered by the GRADE Working Group (2004), the HDA (Kelly *et al.*, 2004) and the Guide to Community Preventive Services (Briss *et al.*, 2000). However, it is recognised that these data are seldom available from public health interventions undertaken to date.

Synthesis of different types of evidence

There is no consensus from the literature as to how different types of research study might be weighted in terms of their contribution to the overall summary of evidence and/or final grade of recommendation. Significant shortcomings were found in current approaches to grading levels of evidence when six prominent grading systems were critically appraised (GRADE Working Group, 2004). Some reviewers suggest that decisions about quality may require complex, contextualised judgements in combination with existing evaluation methodologies (Pawson *et al.*, 2004). A review of the integrative approaches to qualitative and quantitative evidence concludes that more research is required to resolve the complex theoretical and methodological issues involved in developing the best method for synthesis, although a number of established methods exist, each with advantages and disadvantages (Aldrich *et al.*, 2003). The aim should be to make judgements transparent and to try to protect against bias in the judgements that are made by being systematic and explicit (GRADE Working Group, 2004).

In a completely new model of research synthesis, a 'realist' approach to evaluative research has been suggested (Pawson *et al.*, 2004), where complexity is acknowledged throughout in the task of searching the evidence base. The authors argue that the success of an intervention theory is not simply a question of the merit of its underlying ideas, but depends on the individuals, interpersonal relationships, institutions and infrastructures through which and in which an intervention is delivered.

In summary, a large number of factors should be taken into account in reaching a decision on the likely success of an intervention. The grade of evidence and recommendation should be based on a number of building blocks (individual studies within a topic) and clear, detailed guidance on the type and quality of each relevant study should be provided to steer this process.

Consultation with individuals and organisations with expertise in public health and/or grading methodology

At various stages of development, the findings from the literature review and the proposed framework were discussed with (or circulated for comment to) a large number of public health experts and expert groups within and outside the HDA, including (among others): the HDA Public Health Evidence Steering Group; the HDA Obesity Reference Group; the HDA Evidence and Guidance Collaborating Centres on Obesity; the GRADE Working Group; the Centers for Disease Control's Guide to Community Preventive Services; the EC 'Getting Evidence into Practice' project; the WHO Health Evidence Network; the Cochrane Health Promotion and Public Health Field; the EPPI Centre; the York Centre for Reviews and Dissemination; the Medical Research Council's Social and Public Health Sciences Unit (University of Glasgow); the London School of Hygiene and Tropical Medicine (Interventional Public Health Group); SIGN; and NICE (see Appendix 2 for list of respondents).

The questions posed at various stages are listed below with an indication of the consensus (if any) from those consulted.

- Is it appropriate to class interventions into individual, group, community/environmental and policy/socio-political, or are there other classes/groupings that should be considered?
  Response: Consensus that these are appropriate, however many interventions will cross these groupings.
- What are the 'most appropriate' types of evidence (of effectiveness) for different types of intervention (eg

individual, social structure, environment, organisation, group, community, society/socio-political interventions)?
Response: Narrow consensus to use RCTs whenever feasible, but accepted that this is unlikely to be the case for socio-political interventions.

- How and where do we capture the magnitude of findings for each component of evidence when formulating a grade of recommendation? Furthermore, is the magnitude of the effect size (and implied cost effectiveness) enough to support a recommendation, or should a cost effectiveness analysis be carried out when such evidence is not already available?
  Response: No consensus on how to capture magnitude of findings for each component, not least because outcome measures will vary. Cost effectiveness is difficult to estimate.
- In considering the factors that determine the relevance, generalisability and feasibility of an intervention to UK populations (corroboration), are some more critical than others, and how should they be weighted?
  Response: No consensus on how different aspects should be weighted, but inclusion of corroborative evidence is important.
- How should the combinations of evidence (qualitative and quantitative) be combined to obtain a balanced view of all the important aspects of a public health intervention (effectiveness, appropriateness, sustainability, etc)?
  Response: No consensus.
- In the final grading, is it more helpful to have: (i) an overall grading system (eg A, B) derived from a narrative summary of the different types of evidence; or (ii) a composite grading (eg A3, B1) that reflects the two components of (cost)-effectiveness and corroboration, but may lead to a lack of clarity (eg is B1 a stronger recommendation than A3?)
  Response: No consensus – some argue that a composite grading allows readers to make their own judgements, while others suggest this is too complex.

It is clear that a number of issues have yet to be resolved.

Comments and suggestions of respondents (Appendix 2) were, as far as possible, incorporated into this document and the proposed framework. This does not mean that respondents endorsed the framework – three respondents recommended existing alternative systems as more appropriate for public health evidence (GRADE – GRADE Working Group, 2004; CDC Guide to Community Preventive Services – Briss *et al.*, 2000; realist review – Pawson *et al.*, 2004). As our remit was to develop a system based on the NICE methodology, we have incorporated elements of these three very different systems where relevant and possible. The remainder of respondents were broadly supportive of the developing methodology, and of carrying out a pilot within a practical setting to explore the issues raised.

Developing and piloting the provisional framework

A pragmatic framework was developed, based on the findings from the literature review and the views of experts in the fields of public health and health promotion research. The framework included critical appraisal of individual studies and reviews to assess the strength of evidence, based on the quality and quantity of studies.

In assessing efficacy, the NICE/SIGN (level 1–4) evidence classification (SIGN, 2001; NICE, 2004a,b) was adapted to include non-randomised and quasi-experimental studies, as these are common public health research methods. However, the framework differs further from the NICE system in two significant respects:

- The 'most appropriate' (or highest level of) evidence for efficacy is not necessarily the RCT, in particular for socio-political interventions
- The issue of 'directly applicable to the target population' and 'extrapolated evidence' is separately assessed as 'corroborative evidence' and in so doing the framework draws on sources of evidence above and beyond that found in the studies of efficacy.

The system allows for the grade of recommendation to be promoted where the research design used to demonstrate efficacy is weakened by design or methods, but where there is consistent evidence from corroborative studies to suggest that the intervention is relevant, feasible and could be implemented for the population in question. This kind of approach is consistent with the GRADE methodology (GRADE Working Group, 2004).

Practical guidance was produced for those developing grades of recommendation from the available evidence on efficacy,

Table 1 Pilot public health evidence grading scheme: classification of recommendations	
Class	**Basis for decision***
A [PH]	At least one 1++ study or consistent findings in a body of studies** principally rated as 1+ for efficacy***, with strong or moderate evidence of corroboration OR Consistent findings in a body of 2++ studies for efficacy, with strong evidence of corroboration
B [PH]	At least one 1++ study or consistent findings in a body of studies principally rated as 1+ for efficacy, with limited/no evidence of corroboration OR A single 1+ study for efficacy, with strong or moderate evidence of corroboration OR A single 2++ study or consistent findings in a body of studies principally rated as 2+ for efficacy, with strong evidence of corroboration OR Consistent findings in a body of studies principally rated as 2++ for efficacy, with moderate evidence of corroboration
C [PH]	Consistent findings in a body of studies principally rated as 2++ for efficacy, with limited/no evidence of corroboration OR A single 2++ study or consistent findings in a body of studies principally rated 2+ for efficacy, with moderate evidence of corroboration OR A single 2+ study for efficacy, with strong evidence of corroboration OR A body of level 3 or 4 evidence for efficacy, with strong evidence of corroboration
D [PH]	A single 2++ study or consistent findings in a body of studies principally rated 2+ for efficacy, with limited/no evidence of corroboration OR A single 2+ study for efficacy, with moderate evidence for corroboration OR A body of level 3 or 4 evidence of efficacy, with moderate/limited evidence of corroboration OR Formal consensus
D [GPP]	A recommendation based on experience of best practice by health professionals and expert groups
*See Tables 2 and 3 for key to study type, quality and strength of evidence. **Body of studies = 3 or more, or a systematic review. ***For national environmental/socio-political interventions, a body of 2+ studies is acceptable. [PH] public health; [GPP] Good Practice Point. *Source*: adapted from SIGN (2001).	

Table 2 Evidence of the efficacy of an intervention – did it work?	
Level of evidence	**Type of evidence**
1++	High quality meta-analyses, systematic reviews of RCTs (including cluster RCTs), or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias
1–*	Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias
2++	High quality systematic reviews of, or individual high quality non-randomised intervention studies (controlled non-randomised trial, controlled before-and-after, interrupted time series), comparative cohort and correlation studies with a very low risk of confounding, bias or chance
2+	Well conducted, non-randomised intervention studies (controlled non-randomised trial, controlled before-and-after, interrupted time series), comparative cohort and correlation studies with a low risk of confounding, bias or chance
2–*	Non-randomised intervention studies (controlled non-randomised trial, controlled before-and-after, interrupted time series), comparative cohort and correlation studies with a high risk of confounding, bias or chance
3	Non-analytical studies (eg case reports, case series)
4	Expert opinion, formal consensus
*Studies with a level of evidence (–) should not be used as basis for making recommendations. *Source*: adapted from SIGN (2001).	

corroboration and cost effectiveness. Essentially, the evidence for the efficacy of an intervention (in a particular setting or with a particular population) is first assessed based on the research design, quality and quantity of studies, and a decision is made on the overall strength of the evidence of efficacy for each outcome (eg weight, diet, physical activity). This is then combined with an overall assessment of the strength of evidence of corroboration for the intervention in question, based on evidence from the efficacy studies and from elsewhere.

The framework is being piloted alongside the current NICE system within rapid reviews being carried out for the public health/prevention aspects of the HDA/NICE guidance on overweight and obesity. As a consequence of the initial piloting, it has been further amended as summarised in Tables 1–3. The methods by which the building blocks of evidence from different study types might be appraised and combined to guide an overall grade of recommendation are summarised in Figure 2 on page 12.

Table 3 Evidence for corroboration – will it work? (evidence to support implementation in the UK today) and does it matter? (evidence of salience and relevant outcomes for UK populations today)

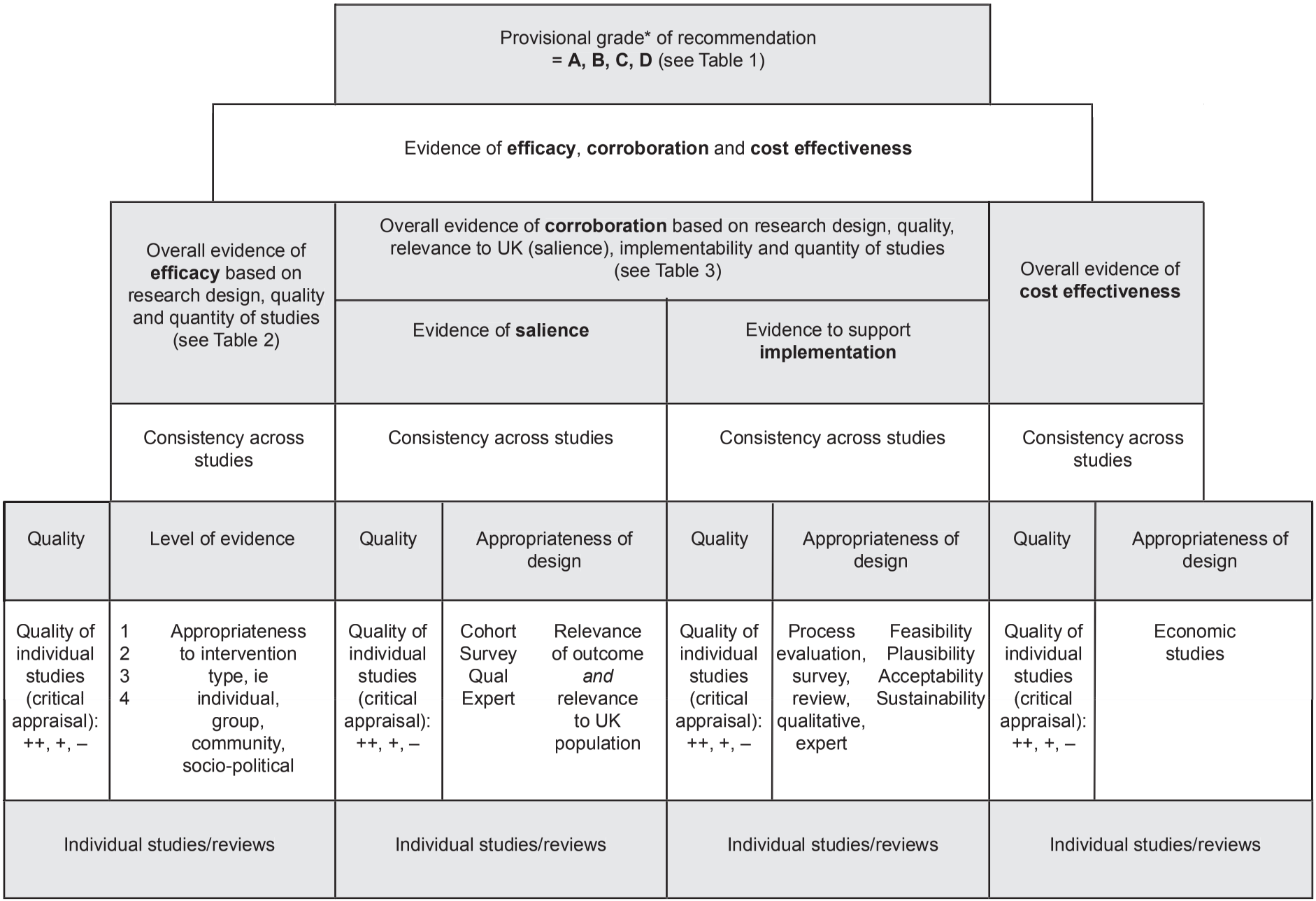
Strength of evidence	Type of evidence
Strong	Consistent findings in two or more studies of ++ quality carried out within the UK and applicable* to the target population, providing evidence on salience and implementation
Moderate	One ++ study or consistent findings in two or more studies of + quality carried out within the UK and applicable to the target population OR Two or more ++ studies from outside the UK but applicable to the target population, providing evidence on salience and implementation
Limited	Only one + study from the UK, two or more studies with inconsistent findings (on balance providing evidence of benefit or harm) or studies of + quality from outside the UK
No evidence	No study of acceptable quality, inconsistent findings (on balance providing no useful evidence) or no relevant research available

*Applicable – in general terms of age, socio-economic status, ethnicity, gender and cultural/religious practices.
Note: there is no established evidence hierarchy for corroborative studies.
Key to quality: ++, very low risk; +, low risk; –, high risk of confounding, bias or chance.

A number of issues were highlighted when the proposed framework was piloted.

- The presentation of corroborative evidence provides valuable information to the developers of recommendations, and aids transparency. There is, as yet, no agreed hierarchy for corroborative evidence (often a combination of observational studies and qualitative evidence), nor is it clear whether corroborative evidence for one outcome, such as diet, can be extrapolated to another outcome, such as physical activity.
- Weighting a body of evidence of efficacy is still under discussion, in particular, how can the evidence be balanced when there is not complete consistency of findings?
- The framework does not, as yet, incorporate the size of the effects including (cost–)benefits and harms for the different outcomes measured.

Figure 2 Pyramid of evidence building blocks on which grades of recommendations may be based



*The final grade would take into account magnitude/effect size(s) (+ve or –ve)

Key to quality: ++, very low risk; +, low risk; –, high risk of confounding, bias or chance.

Discussion and conclusions

This project set out to answer the research questions listed on page 3. While the literature review and consultation with experts suggest that the RCT design is usually the best method to demonstrate the effectiveness of individual and group interventions (and cluster RCTs for many community interventions), there is a dominant (if not universal) view that it may not lend itself to evaluating the effectiveness of many complex public health interventions, such as those involving communities and socio-political (including organisational) 'interventions'. In these cases a non-randomised design may be more appropriate.

In addition, evidence from observational and qualitative research is central to providing the depth and insight required for implementing/replicating appropriate and sustainable interventions. A combination of different study types (quantitative and qualitative) is required to build up a picture of the likely success of an intervention when implemented in a specified context, as is the consistency of research results (where each subsequent study supports the results of the previous studies).

Responses from the expert consultation confirm that this is a complicated area. Various grading systems are already in place, and some respondents did not feel that another system was required. Despite this, the majority of respondents supported this work and its emphasis on the inclusion of corroborative evidence.

The provisional framework presented here aims to provide a practical, but detailed and transparent, method for deriving grades of recommendation for public health interventions, based on a synthesis of all relevant supporting evidence from research. Decisions on the strength of the evidence for efficacy within the framework are, where possible, in line with existing methodologies.

The literature review demonstrates that there is no consensus concerning natural hierarchies for studies looking at corroborative evidence based on salience and implementation. We have proposed a simple, transparent system for assessing the strength of such evidence, while recognising that this results in a considerable increase in the amount of data to be considered within a literature review. However, it is conceded that the appropriateness and ease with which these types of evidence can be combined would benefit from some further clarification.

While the existing NICE methodology does not assess corroborative evidence explicitly, it does consider whether evidence is 'extrapolated' and/or 'directly applicable' (SIGN, 2001; NICE, 2004a,b). Essentially, 'directly applicable' evidence is from studies carried out on populations that are so similar to the target population that applying the same interventions can be expected to have the same effects. Thus evidence from UK studies would normally be considered directly applicable; studies from elsewhere may also be judged directly applicable (Robin Harbour, SIGN, personal communication). The definition of such terms remains open to interpretation and may lead to inconsistencies in the range of issues considered and their implementation. We have therefore attempted to ensure that there is clarity and rigour in the assessment of applicability. This includes taking into account whether the study was conducted in the UK, although it is recognised that this is only a proxy indicator of generalisability: there may be interventions implemented abroad that are more pertinent to some UK populations than those implemented in the UK. This part of the framework would benefit from some further development.

At present the framework does not formally take effect size into consideration as part of the grading. An intervention may have a body of high quality evidence to indicate that it

has been effective in changing some outcome, and there may be strong evidence to suggest it would be implementable in the UK, yet its effect on the population may be negligible. Further work will need to be undertaken so that the final grading and prioritisation of the recommendation will be based on size of effect (including any differential impact on health inequalities), as well as the strength of the underlying evidence, and our confidence in being able to replicate the intervention successfully in a UK setting today.

This framework is being piloted and compared alongside the existing NICE system, within the development of the public health/prevention aspects of the HDA/NICE guidance on overweight and obesity. The lessons learned will help to inform the forthcoming work of the National Institute for Health and Clinical Excellence.

Appendix 1 Organisations involved in searching and summarising evidence for public health

- Campbell Collaboration
  www.campbellcollaboration.org
- Centre for Knowledge Transfer
  www.ckt-ctc.ca/English/Links.htm
- Centre for Reviews and Dissemination – particularly the Wider Public Health project
  www.york.ac.uk/inst/crd/wph.htm
- Centers for Disease Control and Prevention
  www.cdc.gov and www.thecommunityguide.org/default.htm
- Cochrane Collaboration
  www.cochrane.org
- European Project, Getting Evidence into Practice
  www.nigz.nl/gettingevidence
- Hamilton Public Health & Community Services, Effective Public Health Practice Project (EPHPP)
  www.city.hamilton.on.ca/PHCS/EPHPP/default.asp
- Health Evidence Network
  www.euro.who.int/HEN
- International Obesity Taskforce
  www.iotf.org
- National Coordinating Centre for Health Technology Assessment
  www.ncchta.org
- National Institutes of Health
  www.nih.gov
- New Zealand Health Technology Assessment
  http://nzhta.chmeds.ac.nz
- Public Health Association of Australia
  www.phaa.net.au

Appendix 2 Those consulted on the developing framework

The literature review was conducted by SURE and the draft methodology was developed in collaboration with the HDA. In addition, many groups and individuals were consulted during the developmental stages of the methodology and have made contributions during the process. Their participation and contributions are gratefully acknowledged, although their inclusion in the list below in no way signifies their support or endorsement.

- HDA Obesity and Evidence and Guidance teams, including Mike Kelly, Caroline Mulvihill, Hugo Crombie and Daniel Warm
- HDA Obesity Reference Group, including Andrew J. Hill (University of Leeds), Penny Gibson (Royal College of Paediatrics & Child Health), Ken Fox (Bristol University) and Mike Lean (University of Glasgow)
- NICE, including Francoise Cluzeau and Jeremy Wyatt
- Public Health Evidence Steering Group Methodology Subgroup, including Josephine Kavanagh and Sandy Oliver (on behalf of the EPPI Centre) and Ray Pawson (University of Leeds)
- Public Health/Prevention Subgroup of the NICE/HDA Obesity Guideline Development Group
- UK and Ireland Public Health Evidence Steering Group
- Wales HDA Obesity Collaborating Centre, including Eddie Coyle (Wales Centre for Health); Chris Roberts and Nina Parry-Langdon (Health Promotion Division, Welsh Assembly Government)

- Robert Borush, University of Pennsylvania
- Mary Dixon-Woods, Department of Health Sciences, University of Leicester
- Laurel D. Edmonds, Care of Children with Obesity Clinic, Bristol Royal Children's Hospital
- Nick Finer, Centre for Obesity Research, Luton and Dunstable NHS Trust
- Penny Gibson, Blackwater Valley & Hart PCT
- Tim Gill, Australian Society for the Study of Obesity
- Christine Godfrey, Department of Health Sciences, University of York
- Margot Greer, National Public Health Service for Wales
- Peter Hajek, Barts and The London, Queen Mary's School of Medicine and Dentistry
- Robin Harbour, SIGN
- Nicki Jackson, Cochrane Health Promotion and Public Health Field
- Sue Lloyd, Wales Centre for Health
- Anne Ludbrook, Health Economics Research Unit, University of Aberdeen
- Andrew Oxman and the GRADE working group
- Mike Rayner, British Heart Foundation Health Promotion Research Group
- Mary Renfrew, Mother and Infant Research Unit, University of Leeds
- Tim Stokes, National Collaborating Centre for Primary Care
- Carolyn Summerbell, Teesside University (HDA Obesity Collaborating Centre)
- Malcolm Ward, National Public Health Service for Wales

References and bibliography

Papers marked ✓ were included in the literature review.

✓AHRQ (2002) *Systems to rate the strength of scientific evidence*. Evidence Report/Technology Assessment No. 47. Agency for Healthcare Research & Quality, University of North Carolina. www.ahrq.gov/clinic/epcsums/strengthsum.htm

✓Aldrich, R., Kemp, L., Stewart Williams, J. *et al.* (2003) Using socioeconomic evidence in clinical practice guidelines. *British Medical Journal* 327: 1283-5.

Atkins, D., Eccles, M., Flottorp, S., Guyatt, G.H., Henry, D., Hill, S., Liberati, A., O'Connell, D., Oxman, A.D., Phillips, B., Schunemann, H., Edejer, T.T., Vist, G.E. and Williams, J.W. Jr: GRADE Working Group (2004) Systems for grading the quality of evidence and the strength of recommendations. I: Critical appraisal of existing approaches. *BMC Health Services Research* 4: 3844.

✓BMJ Publishing Group (website) 'Clinical evidence.' www.clinicalevidence.com

✓Briss, P.A., Zaza, S., Pappaioanou, M. *et al.* (2000) Developing an evidence-based guide to community preventive services – methods. *American Journal of Preventive Medicine* 18(1S): 35-43.

✓Canadian Task Force for Preventive Health Care (2005) 'Methods: study selection and abstraction forms'. Provided directly from authors, 19/02/05.

✓Clinical Evidence (website) 'How Clinical Evidence is put together'. www.clinicalevidence.com/ceweb/about/put_together.jsp

✓Dane, A.V. and Schneider, B.H. (1998) Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review* 18(1): 23-45.

✓Des Jarlais, D.C., Lyles, C., Crepaz, N. and the TREND Group (2004) Improving the reporting quality of nonrandomized evaluations of behavioural and public health interventions: The TREND statement. *American Journal of Public Health* 94(3): 361-6.

✓Dixon-Woods, M., Agarwal, S., Young, B., Jones, D. and Sutton, A. (2004) *Integrative approaches to qualitative and quantitative evidence*. London: Health Development Agency.

✓Ellis, S. and Grey, A. (2004) *Prevention of sexually transmitted infections (STIs): A review of reviews into the effectiveness of non-clinical interventions. Evidence Briefing.* London: Health Development Agency. www.hda.nhs.uk/documents/prevention_stis_evidence_briefing.pdf

EPHPP (2002) *The effectiveness of school-based strategies for the primary prevention of obesity and for promoting physical activity and/or nutrition, the major modifiable risk factors, for type 2 diabetes: A review of reviews*. Hamilton, Ottawa, Canada: Effective Public Health Practice Project. www.city.hamilton.on.ca/PHCS/EPHPP/Research/Full-Reviews/Diabetes-Review.pdf

✓EPHPP (website) 'Public health and community services', City of Hamilton. Hamilton, Ottawa, Canada: Effective Public Health Practice Project. www.hamilton.ca/PHCS/EPHPP/AboutEPHPP.asp

EPOC Group (2002) *The data collection checklist.* Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. www.epoc.uottawa.ca/checklist2002.doc

✓Evans, D. (2003) Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing* 12: 77-84.

✓Glasgow, R.F., Vogt, T.M. and Boles, S.M. (1999) Evaluating the public health impact of health promotion interventions: the RE-AIM Framework. *American Journal of Public Health* 89(9): 1322-7.

✓Glasziou, P., Vandenbroucke, J. and Chalmers, I. (2004) Assessing the quality of research. *British Medical Journal* 328: 29-41.

GRADE Working Group (2003) 'A critical appraisal of systems for grading evidence and formulating recommendations.' Unpublished report. Informed Choice Research Department, Norwegian Health Services Research Centre, Oslo, Norway: GRADE Working Group.

✓GRADE Working Group (2004) Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490-4.

✓Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G. and Oakley, A. (2004) Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology & Community Health* 58: 794-800.

✓Hawe, P., Shiell, A. and Riley, T. (2004) Complex interventions: how 'out of control' can a randomised controlled trial be? *British Medical Journal* 328: 1561-3.

✓Health Evidence Bulletins Wales (website) 'Project methodology 5'. http://hebw.uwcm.ac.uk/projectmethod/title.htm

International Obesity Taskforce (2004) *Workshop on the prevention of overweight and obesity in children, Prague, 28 May 2004*. Taskforce of the International Association for the Study of Obesity.

Jackson, N. (2004) *Background paper: guidelines for reviewers of health promotion and public health interventions*. Victoria, Australia: Cochrane Health Promotion and Public Health Field.

✓Jackson, N. and Waters, E. (2004) *Guidelines for systematic reviews of health promotions and public health interventions*. Victoria, Australia: Cochrane Health Promotion and Public Health Field. www.vichealth.vic.gov.au/cochrane/welcome/Guidelines%20for%20HPPH%20reviews.pdf

Jackson, N., Waters, E., Doyle, J., Brunton, G., Howes, F., Petticrew, M., Hawe, P., Thomas, H. and Rogers, W. (2004) Guidelines for systematic reviews of the effectiveness of health promotion and public health interventions. Abstract No. 287. In: *Health 2004 – 18th World Conference on Health Promotion and Health Education*, 26–30 April 2004, Melbourne, Australia.

✓Joffe, M. and Mindell, J. (2002) A framework for the evidence base to support health impact assessment. *Journal of Epidemiology & Community Health* 56: 132-8.

✓Kelly, M.P., Charlton, B.G. and Hanlon, P. (1993) The four levels of health promotion: an integrated approach. *Journal of Public Health* 107: 319-26.

✓Kelly, M.P., Chambers, J., Huntley, J. and Millward, L. (2004) *Evidence into practice. Method 1 for the production of Effective Action Briefings and related materials.* London: Health Development Agency.

Kelly, M.P., McDaid, D., Ludbrook, A. and Powell, J. (2005) *Economic appraisal of public health interventions. Briefing paper.* London: Health Development Agency.

✓Kroke, A., Boeing, H., Rossnagel, K. and Willich, S.N. (2004) History of the concept of 'levels of evidence' and their current status in relation to primary prevention through lifestyle interventions. *Public Health Nutrition* 7(2): 279-84.

✓Lean, M.E.J. (2000) Is long-term weight loss possible? *British Journal of Nutrition* 83: S103-S111.

✓MacDonald, G. (2000) A new evidence framework for health promotion practice. *Health Education Journal* 59(1): 3-11.

✓Margetts, B., Warm, D., Yngve, A. and Sjostrom, M. (2001) Developing an evidence-based approach to public health nutrition: translating evidence into policy. *Public Health Nutrition* 4: 1393-7.

✓Millward, L.M., Morgan, A. and Kelly, M.P. (2003) *Evidence briefing: what works in preventing unintentional injuries in children and young adolescents? An updated systematic review.* London: Health Development Agency.

Moher, D., Schulz, K.F. and Altman, D.G. for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357: 1191-4. www.consort-statement.org/statement/revisedstatement.htm

✓Mulvihill, C. and Quigley, R. (2003) *The management of obesity and overweight: an analysis of reviews of diet, physical activity and behavioural approaches.* London: Health Development Agency. www.hda.nhs.uk/documents/obesity_evidence_briefing.pdf

Neuberger, J. (2001) Where's the evidence?: making the case for public health. *Journal of Epidemiology and Community Health* 55(2): 77-8.

✓NICE (2004a) *Guideline development methods. Information for national collaborating centres and guideline developers.* London: National Institute for Clinical Excellence.

NICE (2004b) *Guideline development methods. Information for national collaborating centres and guideline developers.* Appendices. London: National Institute for Clinical Excellence.

✓NSW Centre for Public Health Nutrition (2004) 'Best options for promoting health weight and preventing weight gain in NSW' (draft). Sydney, Australia: University of Sydney. www.cphn.mmb.usyd.edu.au/resources/BestOptionsForPromotingHealthyWeight.pdf

✓Nutbeam, D. (1998) Evaluating health promotion – progress, problems and solutions. *Health Promotion International* 13(1): 27-44.

Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. and Oakley A. (2003) 'An emerging framework for integrating different types of evidence in systematic reviews for public policy.' Unpublished presentation to HDA Methodology Sub-group, February 2003. www.hda.nhs.uk/evidence/sandyo_phesg_0203.pdf

✓Øvretveit, D.J. (2003) 'Grading evidence for decision-makers: issues and methods in assessing the scientific quality of research and summarizing strength of evidence for public health research reviews.' Discussion document for WHO Health Evidence Network (unpublished confidential draft). World Health Organization.

Oxford Centre for Evidence-Based Medicine (website) 'Levels of evidence and grades of recommendation'. www.cebm.net/levels_of_evidence.asp

✓Pawson, R. (2003) *Assessing the quality of evidence in evidence-based policy: why, how and when?* ESRC Research Methods Programme Working Paper No. 1. Manchester: University of Manchester. www.ccsr.ac.uk/methods/publications/Pawson.pdf

✓Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. (2004) *Realist synthesis: an introduction.* ESRC Research Methods Programme Working Paper No. 2. Manchester: University of Manchester. www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf

✓Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57(7): 527-9.

✓Rimer, B.K., Glanz, D.K. and Rasband, G. (2001) Searching for evidence about health education and health behaviour interventions. *Health Education & Behavior* 28(2): 231-48.

✓Rychetnik, L., Frommer, M., Howe, P. *et al.* (2002) Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology & Community Health* 56: 119-27.

SIGN (2001) SIGN 50: *A guideline developer's handbook*. Edinburgh: Scottish Intercollegiate Guidelines Network.

✓SIGN (website) Scottish Intercollegiate Guidelines Network: 'Guidelines methodology'. www.sign.ac.uk/methodology

✓Sorensen, G., Emmons, K., Hunt, M.K. and Johnston, D. (1998) Implications of the results of community intervention trials. *Annual Review of Public Health* 19: 379-416.

✓Stephenson, J. and Imrie, J. (1998) Why do we need randomised controlled trials to assess behavioural interventions? *British Medical Journal* 316: 611-3.

✓Stronks, K. (2002) Generating evidence on interventions to reduce inequalities in health: the Dutch case. *Scandinavian Journal of Public Health* 59(30): 20-5.

✓Swinburn, B., Kumanyika, S. and Prevention Group of the International Obesity Task Force (2004) 'Obesity prevention: a proposed framework for translating evidence into action' (unpublished first draft). Melbourne, Australia: Centre for Physical Activity and Nutrition Research.

Task Force on Community Preventive Services (2002) Recommendations to increase physical activity in communities. *American Journal of Preventive Medicine* 22(4S): 67-72.

✓Task Force on Community Preventive Services (website) 'Guide to Community Preventive Services' (The Community Guide), single data abstraction form for all types of research study. www.thecommunityguide.org/nutrition/default.htm

✓Thomas, J., Sutcliffe, K., Harden, A., Oakley, A., Oliver, S., Rees, R., Brunton, G. and Kavanagh, J. (2003) *Children and healthy eating: a systematic review of barriers and facilitators.* London: University of London.

✓Thomas, J., Harden, A., Oakley, A. *et al.* (2004) Integrating qualitative research with trials in systematic reviews. *British Medical Journal* 328: 1010-2.

✓Thurston, W.E., Vollman, A..R., Wilson, D.R., MacKean, G., Felix, R. and Wright, M.-F. (2003) Development and testing of a framework for assessing the effectiveness of health promotion. *Sozial- und Praventivmedizin* 48(5): 301-6.

✓Tones, K. (2000) Evaluating health promotion: a tale of three errors. *Patient Education and Counseling* 39: 227-36.

✓Truswell, A.S. (2001) Levels and kinds of evidence for public-health nutrition. *Lancet* 357(9262): 1061-2.

✓University of Auckland (a), Faculty of Medical and Health Sciences (website) 'GATE: a graphic appraisal tool for evidence-based clinical practice'. www.health.auckland.ac.nz/population-health/epidemiology-biostats/epiq/newGATEv1.pdf

University of Auckland (b), Faculty of Medical and Health Sciences (website) 'EPIQ (effective practice, informatics & quality improvement'. www.health.auckland.ac.nz/population-health/epidemiology-biostats/epiq/

✓Victora, C.G., Habicht, J.-P. and Bryce, J. (2004) Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health* 94(3): 400-5.

✓Wanless, D. (2004) *Securing good health for the whole population*. Norwich: Stationery Office.

Notes