



*National Institute for
Health and Clinical Excellence*

Issue Date: [March 2006](#)

**Methods for development of NICE public
health guidance**

About this document

This document describes the methods used in the development of NICE public health (intervention and programme) guidance.

This document is available from the NICE website (www.nice.org.uk). A related document on processes is also available:

- National Institute for Health and Clinical Excellence (2006) *The public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public.*

Nothing in this document shall restrict any disclosure of information by the Institute that is required by law (including in particular but without limitation the Freedom of Information Act 2000).

National Institute for Health and Clinical Excellence
MidCity Place
71 High Holborn
London WC1V 6NA

www.nice.org.uk

Produced by the National Institute for Health and Clinical Excellence
First Draft issue March 2006

© National Institute for Health and Clinical Excellence, March 2006. All rights reserved. This material may be freely reproduced for educational and not-for-profit purposes. No reproduction by or for commercial organisations or for commercial purposes is permitted without the express written permission of the Institute..

Contents

Acknowledgements	4
1 Introduction	6
2 Scoping the guidance	12
3 Identifying the evidence	19
4 Reviewing the evidence	24
5 Incorporating health economics	41
6 Creating recommendations	48
7 Testing the recommendations: fieldwork	58
Appendix A – Methodology checklists	65
Appendix B – Guidance scope template	111
Appendix C – NICE review format	115
Appendix D – Data extraction form	122
Appendix E – Glossary	131

Acknowledgements

This document has been prepared by the Centre for Public Health Excellence (CPHE) in NICE. It is partly based on the following publication:

National Institute for Clinical Excellence (2005) *Guideline development methods: Information for national collaborating centres and guideline developers*. London: National Institute for Clinical Excellence.

www.nice.org.uk

It also draws upon work conducted by the Health Development Agency (HDA) to build the evidence base for public health. For further details, see the following:

Kelly MP, Swann C, Morgan A et al. (2002) *Methodological problems in constructing the evidence base in public health*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=508135

Swann C, Falce C, Morgan A et al. (2005) *HDA evidence base: process and quality standards manual for evidence briefings*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=518279

Weightman A, Ellis S, Cullum A, et al. (2005) *Grading evidence and recommendations for public health interventions: developing and piloting a framework*. London: Health Development Agency.

Thanks are also due to staff within other NICE teams including: Information Services (IS), the Implementation and Editorial Teams, Senior Management and the Institute's Board.

The work of the following public health collaborating centres (PHCCs) has also informed development of the CPHE's methods:

- PHCC on obesity, Cardiff University
- PHCC on obesity, University of Teesside
- PHCC on physical activity at Oxford and Loughborough Universities

- PHCC for drug use prevention at Liverpool John Moores University
- PHCC on maternal and child nutrition at the University of York.

1 Introduction

1.1 *Background*

The National Institute for Health and Clinical Excellence (NICE) produces guidance in three areas of health.

- Public health: the promotion of good health and the prevention of ill health for those working in the NHS, local authorities, and the wider public and voluntary sector.
- Health technologies: the use of new and existing medicines, interventions and procedures within the NHS.
- Clinical practice: the appropriate treatment and care of people with specific diseases and conditions within the NHS.

This document describes the methods used by the CPHE in NICE to develop public health programme and intervention guidance. The processes are described in a separate document (see page 2).

NICE's methods and processes of developing public health guidance are located within a clear framework of values and principles as follows.

1.1.1 **Topics and activities**

The range of activities and topics covered is inclusive. Public health activities may be direct (for example, providing family planning or smoking cessation services) or indirect (for example, creating safe open spaces for physical activity as part of general work to upgrade the environment). Traditional public health issues (such as, the welfare of expectant and nursing mothers) and the more implicit issues associated with the wider determinants of health are all covered. The latter might involve, for example, restricting the number of fast food and alcohol outlets in inner city regeneration schemes to discourage people from eating high fat foods or binge drinking.

In summary, NICE public health guidance considers a variety of approaches, from traditional health education and public campaigns to community development.

1.1.2 Determinants of health

NICE public health guidance recognises the wide spectrum of determinants of health – which ranges from social, economic and environmental factors, through to individual choice and ease of access to services.

Recommendations for interventions or programmes may be made at population, community, organisational, group, family or individual level, as appropriate.

1.1.3 Best available evidence

NICE public health guidance is based on the best available evidence drawn from a range of disciplines and research traditions including clinical medicine, epidemiology, health economics, medical sociology, health psychology, medical anthropology, nutrition, sports science, nursing, education and health promotion. Evidence is selected and appraised according to well-defined criteria, based on its appropriateness to answer the research question. It is summarised according to general principles developed by the Institute.

Essentially, there are three issues to tackle.

1. How to identify published and unpublished evidence.
2. How to assess the quality of individual studies.
3. How to synthesise evidence from different kinds of research, and in particular, how to combine quantitative and qualitative data.

The methods used to identify, assess and synthesise the evidence are based on the need to demonstrate the quality and appropriateness of research. Evidence is judged by how systematic, transparent and relevant it is.

Appendix A provides examples of the critical appraisal checklists used to assess evidence from certain types of research design. Further checklists are being developed to help assess evidence from other research methods.

The strength of the evidence is assessed, in part, on the appropriateness of the study design to answer the specific research question. The randomised controlled trial (RCT) is normally the most appropriate source of evidence for judging the effectiveness of individual/group interventions. The cluster RCT is normally the most appropriate evidence for judging the effectiveness of community interventions. Qualitative research is normally the most appropriate evidence source to answer questions about implementation (including the views of the target population).

The review of evidence also assesses the applicability of each study's findings to specific populations and settings in England.

The HDA made some progress in developing innovative methods to address these issues, as have others (including the Medical Research Council Public Health Sciences Unit in Glasgow, the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI Centre)¹, and the Department of Epidemiology, University of Leicester). The CPHE at NICE will build on this work to inform its approaches. NICE publications on its guidance development and other publications on evidence in public health are listed at the end of this chapter.

Methods will be reviewed on a regular basis to ensure that they are fit and appropriate for purpose.

1.1.4 Stakeholders

The role of stakeholders is central to the development of NICE public health guidance. Guidance is subject to scrutiny and validation by stakeholders

¹ The EPPI Centre is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

throughout the development process to ensure that the resulting recommendations are realistic and appropriate.

1.1.5 Quality assurance principles

In addition to the broader values outlined above, the CPHE operates to the Institute's quality assurance principles, which are designed to ensure that guidance and other solutions are credible, robust and relevant.

- Guidance development processes are governed by clear, published statements of methods and process, including a standard timeline, which are developed in consultation with stakeholders and updated at regular, pre-determined intervals.
- Standard operating procedures are prepared for each principal step in the guidance development process. These procedures are developed in consultation with the staff who will operate them and are reviewed at regular, pre-determined intervals.
- Guidance publications are authorised for publication, on behalf of the Board, by the Guidance Executive.
- The Senior Management Team and the Board receive regular reports that identify variations from the planned programme, the reasons for the variations and the remedial action taken.
- Each member of staff is aware of his or her personal responsibility for endeavouring to assure the quality of their work, through the application of standard processes and methods, and through independent thought and action, when necessary and appropriate.
- The risks associated with guidance development programmes are assessed and reported to the Audit Committee together with risk minimisation and handling strategies.
- A set of clear publication standards is applied to the presentation of the work including review, in every case, by a professional editor.
- Dissemination of the publications is undertaken after analysis of the extent and needs of the audience for each piece of guidance.

- Appropriate training is provided for staff to enable them to apply these principles in their daily practice.

1.2 Further reading

National Institute for Health and Clinical Excellence (2005) *Assessing evidence and prioritising recommendations in NICE guidance*. Paper submitted to the meeting of the NICE Board November 2005. Available from:
www.nice.org.uk/pdf/boardmeeting/brdnov05item4.pdf

Dixon-Woods M, Agarwal S, Young B et al. (2004) *Integrative approaches to qualitative and quantitative evidence*. London: Health Development Agency.
www.nice.org.uk/page.aspx?o=508055

National Institute for Health and Clinical Excellence (2005) *Operating model for the centre for public health excellence*. London: National Institute for Health and Clinical Excellence. www.nice.org.uk/page.aspx?o=248187

National Institute for Health and Clinical Excellence (2005) *Social value judgements: draft for consultation*. London: National Institute for Health and Clinical Excellence. www.nice.org.uk/page.aspx?o=250546

Ogilvie D, Hamilton V, Egan M et al. (2005) Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *Journal of Epidemiology and Community Health* 59: 804-808.

Ogilvie D, Egan M, Hamilton V et al. (2005) Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? *Journal of Epidemiology and Community Health* 59: 886-892.

Popay J (editor) (2005) *Moving beyond effectiveness in evidence synthesis: methodological issues in the synthesis of diverse sources of evidence*. London: National Institute for Health and Clinical Excellence.

Weightman A, Ellis S, Cullum A, et al. (2005) *Grading evidence and recommendations for public health interventions: developing and piloting a framework*. London: Health Development Agency.

2 Scoping the guidance

2.1 Introduction

Guidance topics are selected by Ministers. The Department of Health (DH) identifies the main areas to be covered and this remit is referred to NICE. The Institute then defines exactly what the guidance will and will not examine. This process is referred to as 'scoping' and the document containing this information is referred to as 'the scope'.

This chapter describes the methods used to prepare the scope.

2.2 Purpose of the scope

The scope aims to:

- provide a clear definition of the intervention(s)/programme to be addressed
- provide a definition of what the guidance will include and exclude
- identify the settings (for example, the community, primary care or workplace), practitioners and public health delivery systems involved
- identify the population(s) to be included and excluded
- briefly describe the relevant epidemiology
- set the DH referral within a clear policy context
- develop the key questions
- set clear parameters to ensure that the guidance can be developed within the allocated time period
- specify the outcomes, and any comparators, that will be used.

There is an example of a scope template at appendix B.

2.3 Drafting the scope

The following sections describe the procedures that inform development of the draft scope: an initial search of the literature, and development of a conceptual framework to help formulate the key questions.

2.3.1 Preliminary literature search

The purpose of the preliminary literature search is to help define the parameters of the scope. The NICE Information Services (IS) lead carries out a preliminary search of the literature to gain an overview of the issues, help define key areas and to gain an idea of the volume of literature likely to be available (and therefore, the amount of work involved). This includes a search for related NICE guidance (existing or in development), and for guidance published by other agencies in the UK (and abroad, if relevant).

A list of commonly used resources is provided below in figure 2.3.1 (however, not every resource will be relevant to every area of work).

Figure 2.3.1 List of common databases and websites

Ageline

AMED (Allied and Complementary Medicine)

ASSIA (Applied Social Science Index and Abstracts)

British Nursing Index

CINAHL (Cumulative Index of Nursing and Allied Health Literature)

Cochrane Library

Cochrane Central Register of Controlled Trials

Cochrane Database of Systematic Reviews (CDSR)

Current Contents

Database of Abstracts of Reviews of Effectiveness (DARE)

EMBASE

Kings Fund

Medline

National Guidelines Clearinghouse (US) website

National electronic Library for Health (NeLH) – Guidelines Finder website

National Research Register (NRR)

NICE website

Ongoing trials register

Prodigy website

PsycINFO

Research Findings Register (ReFer)

SIGLE (grey literature)

Sociological Abstracts

Trials registers for periods not covered by the highest level of evidence available

Zetoc (Electronic tables of contents)

2.3.2 Developing key questions

Once background documents have been discussed, and issues such as settings and population have been determined, the CPHE Project Team drafts a set of key questions.

The number of key questions required depends on the topic area, whether intervention or programme guidance is being developed, and the breadth of the scope. However, it is important to ensure the total number is:

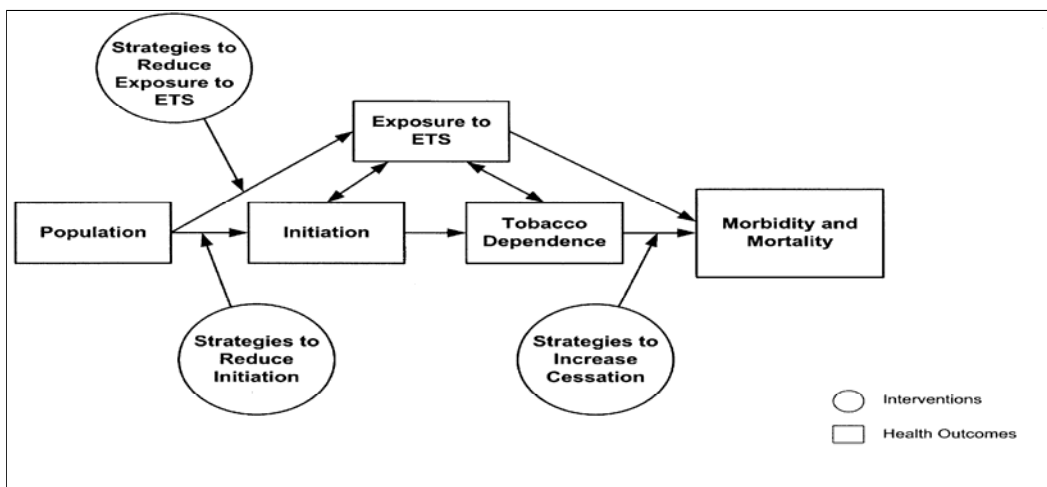
- manageable for the Public Health Interventions Advisory Committee (PHIAC) or the Programme Development Group (PDG)
- appropriate for the allocated budget
- providing a sufficient focus for the guidance.

Key questions cover all areas included in the remit, but should not introduce new aspects. They should be concerned with effectiveness, cost effectiveness, feasibility and acceptability – and may also cover the determinants of health, including risk and protective factors (and the relationships between them). Above all, they should be clear, realistic and focused.

The CPHE Project Team should construct a diagram (a 'conceptual framework') listing important outcomes and other key criteria as a basis for framing the questions, then key words can be identified as potential search terms.

The US Centers for Disease Control and Prevention (CDC) *Community Guide*² (www.thecommunityguide.org) uses a 'logic framework' to illustrate the broad links between social, environmental and biological determinants, potential interventions and outcomes. This is used to develop an 'analytic framework' to demonstrate the relationship between particular interventions and their intended outcomes. (See following figures for an example of this two stage process.)

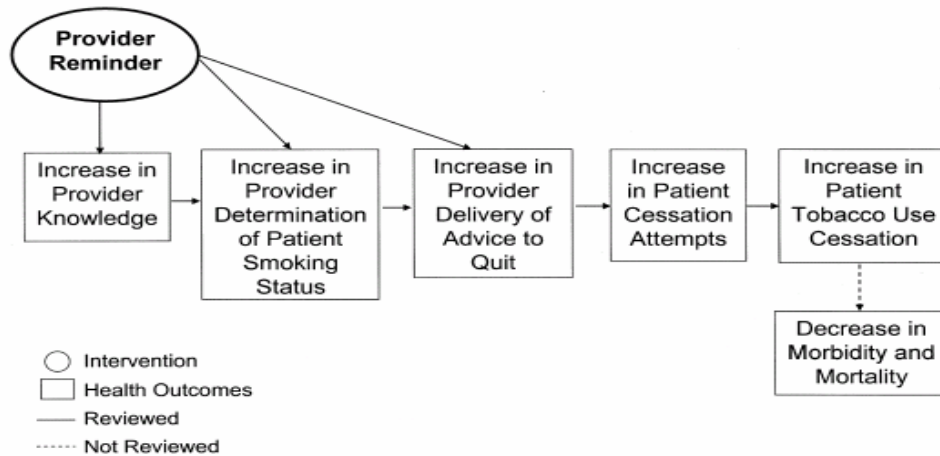
Figure 2.3.2 Developing a conceptual framework for key questions: how the US CDC applied its logic framework to its tobacco reviews



ETS stands for environmental tobacco smoke.

² Task Force on Community Preventive Services (2005) Zaza S, Briss P, Harris K Editors. *The guide to community preventive services. What works to promote health?* Atlanta: Centers for Disease Control and Prevention.

Figure 2.3.3 An analytic framework illustrating the relationships between a smoking intervention and its outcomes



Having developed a conceptual framework using the CDC or a similar approach, the CPHE Project Team should use the population, intervention, comparison and outcome (PICO) framework to define individual questions. PICO divides questions into the following four components.

- Population under study
 - which populations are we most interested in?
 - how can they best be described?
 - are there subgroups that need to be considered?
 - are there any relevant inequality or exclusion issues here?
- Intervention/approach
 - which interventions/approaches should be used?
- Comparison
 - what is/are the main alternative(s)? (including 'usual practice' or 'do nothing')
- Outcome
 - what really matters to the population or individual?

- which outcomes should be considered (for example, mortality, morbidity, relapse rates, physical and social functioning, costs, health status and so on)?

2.3.3 Sub-questions

For each type of intervention covered by the guidance, the various confounding factors that may influence its outcome and effectiveness are taken into account. The scope will normally include the following sub-questions.

- What factors/determinants does the intervention aim to influence (for example, does it aim to change risk factors)?
- How valid and appropriate are the outcome measures used to assess effectiveness? (For example, self-report versus biologically validated measures).
- How does its content influence effectiveness?
- How does the way the intervention/programme is carried out influence effectiveness?
- Does effectiveness depend on the job title/position of the intervenor or factors such as their age, gender or ethnicity? What are the significant features of an effective deliverer/leader?
- Does the site/setting influence effectiveness?
- Does the intensity or length of the intervention influence effectiveness or duration of effect?
- How does effectiveness vary according to the age, gender, class and ethnicity of the target audience? Is there any differential impact on inequalities in health within and between different population groups?
- How much does it cost (in terms of money, people, and time)? What evidence is there on cost effectiveness?
- What are the barriers and facilitators to implementation?
- How acceptable is the intervention to the target population?
- What are the adverse or unintended outcomes?

- What impact does it have on inequalities in health?

2.4 Consulting on the scope

Once the scope has been drafted, it is posted on the website and stakeholders are notified. Stakeholders are then given the opportunity to register their interest, attend a public consultation meeting and provide feedback during the consultation period. Following the consultation, a final scope is prepared and published, along with stakeholder comments and the Institute's responses to those comments.

3 Identifying the evidence

The CPHE and its contractors work closely with the Information Services Team to develop a framework for identifying the best available evidence of effectiveness and cost effectiveness for each piece of guidance. This framework helps to refine and expand the key questions, sets out how evidence will be identified, and highlights any relevant NICE guidance and evidence reviews – either published or under development. It also outlines how areas without evidence will be handled.

3.1 Protocol for searching electronic databases

A protocol for identifying the relevant databases is developed by the Collaborating Centre or contractor in liaison with the CPHE Project Team, taking account of NICE procedures set out in the Guideline Development Manual (NICE 2005) chapters 6 and 7. See: www.nice.org.uk/pdf/GDM_Chapter6_0305.pdf and www.nice.org.uk/pdf/GDM_Chapter7_0305.pdf

It must have signed agreement from the Director of the CPHE or the Director's nominated representative before the literature search proceeds.

See Chapter 2 page 15 for a list of the databases that are usually searched.

If insufficient studies are found, databases of practice may be searched.

3.1.1 Literature search

A comprehensive body of evidence is generated to:

- answer the questions in the scope
- highlight any significant gaps in the evidence base where formal consensus methods may be required.

Developing a search strategy for any type of review is an iterative process and it is not possible to define one that will be appropriate for all scenarios. This section explains the principles of developing such a strategy.

In each case, the strategy will go through several stages of refinement following discussions about the search results. The initial approach (the search undertaken for the scope) may have to be revised to ensure that specific aspects of the questions are adequately covered. There are a number of key factors to take into account.

- **Sensitivity and specificity** A search strategy should be sensitive (able to identify relevant information) and specific (able to exclude irrelevant documents). Both attributes are influenced by the time period covered and by the search terms used. There has to be a trade-off between conducting an exhaustive search (with the additional resources that requires) and undertaking a more modest search (that may miss some studies but would not alter the overall strength of the evidence/findings).
- **Time period** The period that the search covers depends on the guidance topic and when the bulk of the research was published. Time limits are agreed by the CPHE Project Team, in consultation with external experts, as part of the scope consultation. Where adequate published systematic reviews exist, additional searches may be limited to the time since the review was published. If existing reviews do not address all the relevant outcomes, however, additional searches may be required. Review authors may be contacted for updates, particularly for reviews in the Cochrane library.
- **Documentation** The process for identifying the evidence should be repeatable and transparent and the search strategy and terms documented. This provides an important audit trail, describing modes of searching and reasons for changes and amendments. A full description of the documentation process that NICE contractors and collaborating centres should use is provided in the 'Centre for reviews and disseminations report

no. 4' (see 4.7 further reading). Electronic records of the references retrieved are stored in a bibliographic database such as reference manager, along with details of ongoing research.

- **Timetable and updates** The contractors, in liaison with the CPHE Project Team, prioritise searches by topic, according to the material required for PHIAC or the PDG. Additional time may be needed for areas where there are likely to be a lot of papers. Specific searches are carried out for each question and are updated 6–8 weeks before consultation on the guidance. Efforts are made to identify major relevant publications after this date, particularly for programme guidance, which has a longer development timeframe.

3.2 Sources of evidence

The key questions in the scope will be addressed by one or more reviews (see chapter 4). Typically, each review will use one or more of the following sources of evidence:

- an evidence briefing (review of reviews)
- a systematic review (of primary data)
- existing (published) primary research
- new (commissioned) primary research (where appropriate and where time and resources allow).

The type of evidence source used depends on the extent to which it answers the questions in the scope. If evidence briefings and published systematic reviews adequately address them, then it will not be necessary to review primary studies. However, it is likely that a review of primary studies will be needed for some questions.

3.2.1 Evidence briefings

Evidence briefings are 'tertiary' forms of evidence or 'reviews of reviews'. They are produced using methods developed by the Health Development Agency (HDA) – see www.nice.org.uk/page.aspx?o=518279

3.2.2 Systematic reviews

Systematic reviews provide a reliable overview of the research results. They differ from traditional reviews and commentaries in that they use a replicable, scientific and transparent approach that aims to minimise bias. They are often concerned with the effectiveness of interventions, although they can cover reviews of aetiology.

Review teams should use systematic reviews that directly answer questions identified in the scope to make an initial assessment of the evidence. Each scope question should be considered separately.

The guidelines developed by the NHS Centre for Reviews and Dissemination at the University of York should be used to ensure the quality of systematic reviews (www.york.ac.uk/inst/crd/report4.htm). There is also a methodology checklist in appendix A.

3.2.3 Existing primary research

Primary research reports, usually published as journal articles, should normally be used. Selection will depend on the extent to which the research answers the questions in the scope (rather than on the basis that one type of research design is valued more than another). For questions on effectiveness, RCTs will normally provide the best source of evidence. For questions on acceptability and implementability, other research designs will be more suitable. For further details, see chapter 4.

Primary research is assessed using the appropriate methodological checklist in appendix A.

3.2.4 New primary research

It may be necessary to commission primary research when little other evidence exists, or when qualitative or opinion data is needed (for example, from good quality market research). Time and budgetary constraints are taken into

consideration when planning any primary research to inform development of the guidance.

3.3 Evidence submitted by stakeholders

Stakeholders are invited to submit potential evidence during consultation on the synopsis report. The references are entered into a bibliographic database and the details are cross-checked with evidence identified through database searching. Items that may be included as evidence are listed below (not in order of priority):

- systematic reviews
- RCTs
- other guidelines on the same topic
- epidemiological studies
- quantitative and qualitative studies relevant to the scope
- published and unpublished economic models
- research in a current academic journal
- good quality process or outcome evaluations of local interventions where relevant information is provided about delivery, staff, public health structures and systems, and any other contextual factors that impact on effectiveness.

Stakeholders are asked not to submit the following types of evidence:

- studies with weak designs when better designed studies are available
- promotional literature
- papers, commentaries and editorials that interpret the results of a published paper
- representations and experiences of individuals (unless included as part of a good quality study).

4 Reviewing the evidence

Once the literature searches are complete, the identified studies and reviews are reviewed. The aim is to select the most appropriate data to help answer the key questions in the scope, and to ensure that the recommendations are based on the best available evidence. This chapter describes how one or more reviews (typically two for intervention guidance and five for programme guidance) are produced by NICE as part of an explicit and transparent process for identifying, assessing and summarising the evidence. Each review involves the following steps:

1. selecting relevant studies or reviews, including evidence briefings, systematic reviews and/or primary research
2. assessing their quality
3. extracting and synthesising the evidence
4. making evidence statements.

A NICE review template is given in appendix C.

4.1 *Selecting relevant studies*

Separate searches are conducted for each question or group of questions (see chapters 2 and 3), using the inclusion and exclusion criteria set out in the scope.

Before acquiring papers for assessment, preliminary screening of retrieved items is carried out to discard irrelevant material. In the first instance, titles are scanned and those outside the topic area are eliminated. A quick check of the remaining abstracts should identify others that are clearly not relevant to the research questions – and they are also excluded.

The remaining abstracts are scrutinised against the parameters agreed by the CPHE Project Team. Abstracts that do not meet the inclusion criteria are eliminated. If there is any doubt, this is resolved by discussion with the CPHE Project Team.

Once the sifting is complete, paper copies of the selected studies or reviews are acquired for assessment. Those that fail to meet the inclusion criteria are excluded; the others are assessed. As there is always an element of bias in selecting evidence, periodically a random selection of abstracts is double sifted (considered by two people). Any differences should be resolved by discussion or recourse to a third reviewer and the 'inter-rater' reliability scores should be reported.

The study selection process is clearly documented, including details of the inclusion criteria applied.

4.2 Assessing the quality of evidence

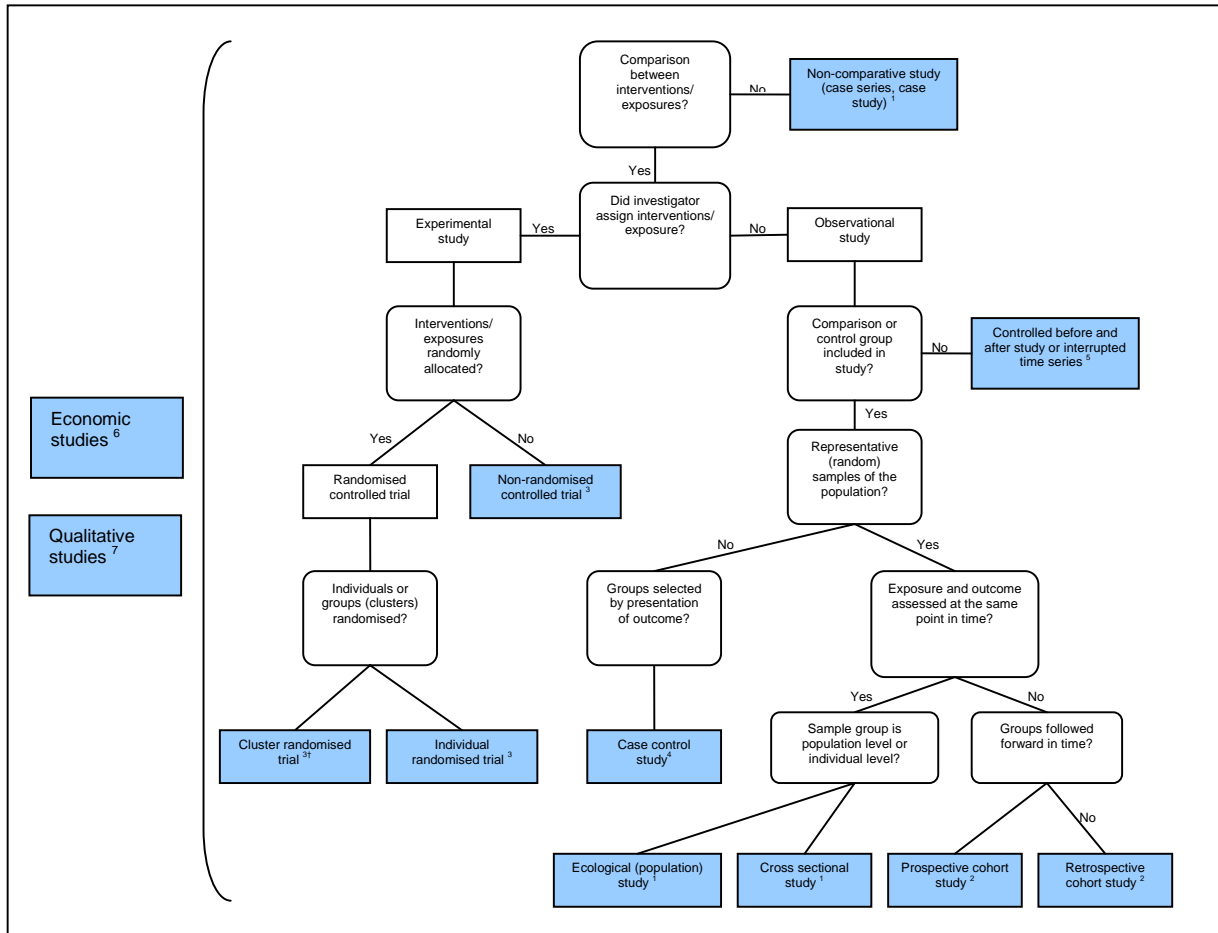
Studies or reviews are evaluated by assessing the methods used in relation to the question(s) being addressed.

This is a key stage in the guidance development process, since the result will affect the strength of evidence ascribed to each evidence statement (see section 4.5). Along with other key factors, such as public health impact and cost effectiveness, this will inform how the recommendations are formulated (see section 6.1) and prioritised (see section 6.4).

Before assessing the quality of each study, its design must be determined so that the appropriate criteria can be applied (see section 4.2.1). An algorithm developed by NICE for questions related to clinical efficacy has been adapted for the questions about effectiveness in public health guidance (see figure 4.1).

The qualitative aspects of a programme or intervention (for example, its acceptability and appropriateness) impact on its success. Likewise, economic evaluations (which determine cost effectiveness) play a key part in determining its feasibility. It is therefore recommended that, wherever possible, the qualitative and economic aspects of the study designs (as depicted in Figure 4.1) are assessed and extracted.

Figure 4.1 Algorithm for classifying primary study designs about effectiveness



Key

- 1 Currently no checklist
- 2 Cohort study checklists (retrospective and prospective), see appendix A.5
- 3 RCT checklist, see appendix A.2
- 4 Case-control study checklist, see appendix A.3
- 5 Interrupted time series (ITS) and controlled before and after (CBA) checklist, see EPOC (appendix A.6, A.7)
- 6 Economic evaluation checklist, see appendix A.8
- 7 Qualitative study checklist, see appendix A.4
- † Requires modification to existing NICE checklist

4.2.1 Published studies

Published studies selected from the search are assessed for their methodological rigour and quality against a number of criteria. These criteria differ according to the study type, so a range of methodology checklists are used to provide a consistent approach.

NICE uses methodology checklists developed originally by the MERGE (Method for Evaluating Research and Guideline Evidence) Group in Australia and modified by the Scottish Intercollegiate Guidelines Network (SIGN). These are supplemented by the EPOC checklist for interrupted time series and CBA studies, and the Tooth et al. (2005) checklist for cohort studies. Health economics studies are assessed using the Drummond checklist (Drummond 1997).

To minimise any potential bias or subjectivity in the assessment, each paper should be assessed by two reviewers and differences resolved by discussion or recourse to a third reviewer. Any remaining differences should be reported in full in the review.

Each study is categorised by study type (categorised as type 1–4) and graded using a code, ‘++’, ‘+’ or ‘–’, based on the extent to which the potential sources of bias have been minimised:

Study type

- 1 Meta-analyses, systematic reviews of RCTs, or RCTs (including cluster RCTs).
- 2 Systematic reviews of, or individual, non-randomised controlled trials, case-control studies, cohort studies, controlled before-and-after (CBA) studies, interrupted time series (ITS) studies, correlation studies.
- 3 Non-analytic studies (for example, case reports, case series studies).
- 4 Expert opinion, formal consensus.

Study quality

- ++ All or most of the criteria have been fulfilled. Where they have not been fulfilled the conclusions are thought very unlikely to alter.
- + Some of the criteria have been fulfilled. Those criteria that have not been fulfilled or not adequately described are thought unlikely to alter the conclusions.
- Few or no criteria have been fulfilled. The conclusions of the study are thought likely or very likely to alter.

Study design and quality are combined. For example, a type 1 study fulfilling most criteria and a type 2 study fulfilling very few criteria would appear in the format (1++) and (2-) respectively. Each review includes a number of evidence statements that reflect the strength (quantity, type and quality) of evidence – see 4.4.

Study design and quality (that is, internal validity) inform decisions about recommendations – alongside data on impact and cost effectiveness (see chapter 6). There is a separate assessment of external validity: that is, how well it can be applied to the particular question or intervention/population in question.

The methodology checklists are presented in appendix A, together with explanatory notes on their use. They have a number of limitations and may not be entirely fit for purpose for all the study designs that reviewers may encounter. For example, although the checklist for RCTs is appropriate for RCTs that randomise by individual, it is less applicable to RCTs that randomise by cluster (where issues of inter- versus intra-cluster correlation and unit of analysis also need considering as potential sources of bias). It is envisaged that future developments will address such limitations.

4.2.2 Unpublished data and studies in progress

Unpublished data may be obtained in the course of the review, particularly from stakeholders (although reviewers are not routinely expected to search the grey literature). Any unpublished data is subjected to a quality assessment in the

same way as published studies. Authors are requested to provide information so that the reviewers can complete the relevant quality checklist.

4.2.3 Published guidance

Relevant published guidance (from NICE and other agencies) should be identified in the search, as well as NICE guidance in development.

4.2.3.1 NICE guidance

NICE guidance is fully referenced and the evidence underpinning the recommendations is left unchanged, provided it is not out of date. If there is new published evidence that would significantly alter the existing recommendations, this should be brought to the attention of the relevant NICE team so that the guidance can be considered for update.

4.2.3.2 Other guidance

Other relevant guidance is assessed for quality using the AGREE instrument (AGREE Collaboration 2003 – see 4.7 further reading) – to ensure it has sufficient documentation. There is no cut-off point for accepting or rejecting guidance and PHAC and each PDG sets its own parameters, which should be documented in the methods section of the guidance, along with a summary of the assessment. The results should be presented as an appendix in the guidance.

Reviews of evidence from other guidance that cover the relevant research questions can be considered, provided:

- the review of evidence is assessed using the appropriate checklist from the methods manual and is judged to be of high quality
- it is accompanied by the evidence statement and evidence table(s).

The NICE review(s) should include their own evidence summaries and statements. Evidence tables from other guidance should be directly linked to the source website address or a full reference to the published document should be given.

4.3 *Extracting and synthesising the evidence*

4.3.1 *Data extraction and evidence tables*

A standard form (see appendix D) is used to extract data about each study and the intervention (if applicable), including its aim, objectives, setting, target population, intervener, content, method and duration. It also gathers information about the study design, outcomes and conclusions, noting:

- what type of question it answers (for example, effectiveness)
- which specific questions from the scope it answers.

All effects/outcomes (significant or not) should be reported for each study, along with confidence intervals and p-values, if appropriate. Meta-analysis should be performed, if appropriate (see 4.3.1.3).

4.3.1.1 *Assessing applicability*

Each study is assessed on its external validity: that is, whether or not it is directly applicable to the target population(s) and setting(s) in the scope. Normally, this takes account of whether the study was conducted in the UK (and when).

However, specific issues will often be determined on a case-by-case basis for each question. NICE is developing a framework to guide judgements about applicability and, in time, there may be a more systematic method for doing this.

In the interim, it is useful to use one of the following phrases to describe the body of evidence in relation to each question:

- likely to be applicable across a broad range of populations and settings
- likely to be applicable across a broad range of populations and settings, assuming it is appropriately adapted
- applicable only to populations or settings included in the studies – the success of broader application is uncertain
- applicable only to settings or populations included in the studies.

In the case of effectiveness studies, the data extraction form allows the reviewer to make a comparison between the parameters of the review and each question.

4.3.1.2 Evidence tables

Evidence tables help identify the similarities and differences between studies, including key characteristics of the study population, interventions or outcome measures. Apart from providing a basis for comparison, they also help determine if it is possible to calculate a mean estimate of effect.

The completed methodology checklist and data extraction forms should be used to produce evidence tables. Separate evidence tables should summarise the evidence for:

- different types of question (for example, about effectiveness, implementation or cost effectiveness)
- individual questions and outcomes (for example, about individual interventions).

The information included depends on the question and the level of detail and analysis needed. The NICE review template in appendix C shows the type of data that should be included in evidence tables related to efficacy studies.

All evidence tables should include columns relating to the nature of the intervention (that is, objectives, content, intervener, duration, method/mode of delivery, population) and the key findings. Crucially, they should also include an assessment of the study's applicability to the target population (see section 4.3.1.1), the type of evidence (that is: 1, 2, 3, 4) and the quality appraisal (that is: ++, +, -). See section 4.2 for further details.

Each discrete intervention has its own set of evidence tables to answer the different types of questions. For studies about effectiveness, it is usually appropriate to further sub-divide the evidence in these tables by different types of outcome – for example, health outcomes, intermediate health outcomes

(including behaviour) or health promotion outcomes (including knowledge, attitudes, service availability).

All reported outcomes, effect sizes and confidence intervals should be included and the tables should make clear where outcomes are self-reported or objectively validated. If the effects are significant, the p-value should be given.

In some circumstances – and if the necessary data are available – it may be appropriate to carry out a meta-analysis (see following section). Even where this is not possible, it is useful to present the data in a forest plot or summary table/chart as well as in the narrative, so that it is clear whether the results from the non-significant studies show similar results.

4.3.1.3 Conducting a meta-analysis

It is appropriate to use meta-analysis to synthesise outcome data (usually from RCTs) provided there is enough relevant and valid data containing comparable outcome measures – and where the data are sufficiently homogenous. (Where such data are not available, the analysis may have to be restricted to a narrative overview of both significant and non-significant individual studies). Forest plots are a useful way of illustrating the results of different studies in a meta-analysis looking at the same question. The characteristics and limitations of the data (that is, population, intervention, setting, sample size and validity of the evidence) should be fully reported.

Before pooling (or combining) the results of different studies, an assessment should be made of the degree of heterogeneity in the results to determine if this reflects the different circumstances in which individual studies were carried out. The results of any homogeneity tests should be reported.

Statistical heterogeneity of study results can be addressed using a random (as opposed to fixed) effects model. Known research heterogeneity (for example, population characteristics or the intensity or frequency of an intervention) can be

managed by judicious use of methods such as subgroup analyses and meta-regression.

For methodological heterogeneity (for example, where different trials of varying quality are involved), sensitivity analyses should be carried out (by varying the studies in the meta-analysis). Forest plots should include lines for studies that are believed to contain eligible data, even if the data are missing from the published study. An estimate of the proportion of missing eligible data is needed for each analysis (as some studies will not include all relevant outcomes).

A full description of data synthesis, including meta-analysis and extraction methods, is available in NHS Centre for Review and Dissemination (2001) (see 4.7 further reading for complete reference). Sensitivity analysis can be used to investigate the impact of missing data.

4.3.2 Levels of evidence

Studies meeting the minimum quality criteria should be categorised according to the 'level' of evidence they provide. This helps the guidance developers (and users) understand the type of evidence on which the recommendations have been based.

There are many ways of assigning levels to the evidence. NICE, in collaboration with the National Collaborating Centres (NCCs), the PHCCs and academic groups throughout the world, is examining a number of systems to identify the most appropriate one. Until a decision is reached, the Institute uses the system shown in table 4.1.

For each type of question (for example, about effectiveness, cost effectiveness and implementation), study designs are ranked in terms of the 'most appropriate' and most feasible.

For questions of efficacy or effectiveness, the most appropriate study design is normally an RCT, although for some public health interventions (for example, those aimed at changing policy), RCTs may not be possible or appropriate. In

these cases, a CBA study might be the best way of establishing cause and effect. (See table 4.1 for the levels of evidence that NICE applies to studies on the efficacy of interventions.)

Qualitative data may provide the most appropriate answers to other questions – for instance, about what is acceptable to the target audience, or how the intervention is delivered. As there are no established hierarchies for this kind of evidence, a judgement has to be made about the level it is assigned. The quality of the method used is separately assessed (see section 4.2).

For questions about implementation, the most appropriate evidence might come from a process evaluation. NICE is developing a critical appraisal checklist for assessing this type of evidence.

Table 4.1 Type and quality of evidence for studies on the efficacy of interventions

Adapted from the Scottish Intercollegiate Guidelines Network (2001); for further information, see further reading.

Type and quality of evidence	
1 ⁺⁺	High quality meta-analyses, systematic reviews of RCTs, or RCTs (including cluster RCTs) with a very low risk of bias
1 ⁺	Well conducted meta-analyses, systematic reviews of RCTs, or RCTs (including cluster RCTs) with a low risk of bias
1 ⁻	Meta-analyses, systematic reviews of RCTs, or RCTs (including cluster RCTs) with a high risk of bias
2 ⁺⁺	High quality systematic reviews of these types of studies, or individual, non-RCTs, case-control studies, cohort studies, CBA studies, ITS, and correlation studies with a very low risk of confounding, bias or chance and a high probability that the relationship is causal
2 ⁺	Well conducted non-RCTs, case-control studies, cohort studies, CBA studies, ITS and correlation studies with a low risk of confounding, bias or chance and a moderate probability that the relationship is causal
2 ⁻	Non-RCTs, case-control studies, cohort studies, CBA studies, ITS and correlation studies with a high risk – or chance – of confounding bias, and a significant risk that the relationship is not causal
3	Non-analytic studies (for example, case reports, case series)
4	Expert opinion, formal consensus
NB: for policy interventions, then CBA can be awarded level 1 evidence.	

PHIAC or the PDG is responsible for endorsing the final levels ascribed to the body of evidence.

4.3.3 Evidence of implementation

Evidence to inform implementation (for example, to determine the acceptability or feasibility of an intervention) may be derived from a number of sources and research designs, including efficacy studies, process evaluations, qualitative research and surveys. There is no established hierarchy and the strength of this evidence depends on its quality and quantity, and its relevance to the target UK populations and settings.

4.4 Making evidence statements

4.4.1 Summaries of the evidence

Each review should include an executive summary of the evidence and how/whether it answers the individual questions. There should also be a summary for each intervention or setting (if appropriate) and each question and sub-question. Each summary should cross-refer to the evidence tables, where all effects/outcomes (significant or not) should be presented for each study, along with their confidence intervals and p-value, if appropriate. Each summary should highlight key confounders observed in the studies and include comments on the nature of the evidence base. Where it relates to the effectiveness of an intervention, a statement about the typical size of effect should normally be included.

4.4.1.1 Evidence statements

The narrative summary related to each question (and sub-question) should conclude with an 'evidence statement' reflecting the strength (quality, quantity and type) of evidence and its applicability to the target population.

In the quality assessment, each paper is rated as: '++', '+' or '-'. Studies rated as '-' should not be used as a basis for making an evidence statement if good quality studies are available to help answer the question – and their outcomes are consistent. If there is a body of reasonable, but fairly weak evidence showing a consistent effect and there are '-' studies that show the same effect, the '-' rated studies should be included. If the '-' studies suggest a different outcome, they should be left in the evidence table for further discussion with PHIAC or the PDG.

For each question, intervention and outcome there should be an evidence statement about:

- the **strength** of evidence (reflecting the appropriateness of the study design to answer the question, the quality and quantity of evidence)

- the **applicability** of the evidence to the question/target population/setting.

For example:

- a body of 1++ evidence of efficacy offers consistent findings about the impact of intervention X on outcome Y
- it is directly applicable to the target population in terms of, for example, ethnicity, age or gender.

These evidence statements are just part of the process of formulating and prioritising recommendations. They reflect what is plausible, given the evidence available about what has worked in similar circumstances. They should help PHIAAC or the PDG to form a judgement about:

- whether or not there is sufficient evidence (in terms of quantity, quality and applicability) to form a judgement
- whether, on balance, the evidence demonstrates that the intervention is effective, ineffective or the evidence is equivocal
- typical size of effect (where there is one).

Depending on the volume of literature, the statement may simply describe the evidence. For example: 'There are two [1+] studies, one [1-] study, and one [2-] study that showed an effect'.

The CPHE is currently developing a system to describe the evidence in a consistent way. In the interim, evidence statements should use terms that convey the strength of the evidence and allow for the following distinctions:

- the evidence base is consistent and in the same direction (for example, it suggests that an intervention is effective or ineffective)
- the evidence base is mixed and therefore equivocal (for example, some good quality studies report a significant effect, while a similar number of good quality studies report no effect)

- there is limited evidence
- there is evidence of limited efficacy
- there is no evidence
- there is evidence of harm.

The evidence statements are included in the executive summaries of the reviews and incorporated into the synopsis – see 4.5.

4.5 Synopsis of evidence

Executive summaries of the reviews of the evidence and the economic appraisal are incorporated into a synopsis that is released to stakeholders, who are invited to submit additional evidence that meets the inclusion criteria. Any new evidence should be assessed using the same methods. The synopsis is also considered by PHIAC/the PDG and used as the basis for drawing up draft recommendations. These recommendations are then tested during fieldwork (see chapter 7).

4.6 References and further reading

AGREE Collaboration (2003) Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care* 12(1):18-23.

Dixon-Woods M, Agarwal S, Young B, et al. *Integrative approaches to qualitative and quantitative evidence*. London: Health Development Agency.

Drummond MF, O'Brien B, Stoddart GL et al. (1997) Critical assessment of economic evaluation. In: *Methods for the economic evaluation of health care programmes*. 2nd edition. Oxford: Oxford Medical Publications.

Edwards P, Clarke M, DiGuiseppe C et al. (2002) Identification of randomized trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine* 21:1635–40.

Eccles M, Mason J (2001) How to develop cost-conscious guidelines. *Health technology assessment* 2001 vol 5. no.16. Southampton: University of Southampton.

GRADE Working Group (2004) Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 490-4.

Harden A, Garcia J, Oliver S et al. (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.

Jackson N, Waters E, for the Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce (2005) *Guidelines for systematic reviews of health promotion and public health interventions. Version 1.2*. Australia: Deakin University: (www.vichealth.vic.gov.au/cochrane/)

Kelly MP, Swann C, Morgan A et al. (2002) *Methodological problems in constructing the evidence base in public health*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=508135

Khan KS, Kunz R, Kleijnen J et al. (2003) *Systematic reviews to support evidence-based medicine. How to review and apply findings of healthcare research*. London: Royal Society of Medicine Press.

NHS Centre for Reviews and Dissemination (2001) Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. *CRD report number 4* 2nd edition. NHS Centre for Reviews and Dissemination, University of York. www.york.ac.uk/inst/crd/report4.htm

Ogilvie D, Hamilton V, Egan M et al. (2005) Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *Journal of Epidemiology and Community Health* 59: 804-808.

Ogilvie D, Egan M, Hamilton V et al. (2005) Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? *Journal of Epidemiology and Community Health* 59: 886-892.

Oxford Centre for Evidence-Based Medicine 'Levels of evidence and grades of recommendation' [online]. Available from: www.cebm.net/levels_of_evidence.asp

Petticrew M, Roberts H (2003) Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57: 527-529.

Popay J, editor (2005) *Moving beyond effectiveness in evidence synthesis: methodological issues in the synthesis of diverse sources of evidence*. London: National Institute for Health and Clinical Excellence.

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

Scottish Intercollegiate Guidelines Network (2002) *SIGN 50. A guideline developer's handbook*. Edinburgh: Scottish Intercollegiate Guidelines Network.

Swann C, Falce C, Morgan A et al. (2005) *HDA evidence base: process and quality standards for evidence briefings – third edition*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=518279

Tooth L, Ware R, Bain C et al. (2005) Quality of reporting of observational longitudinal research. *American Journal of Epidemiology* vol 161(3):280-288.

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

Weightman A, Ellis S, Cullum A, et al. (2005) *Grading evidence and recommendations for public health interventions: developing and piloting a framework*. London: Health Development Agency.

5 Incorporating health economics

5.1 Introduction

Economic analysis and evaluation forms an integral part of the public health guidance development process. This chapter describes how the economic evidence should be collated and analysed.

The first Wanless report (2002) assessed the resources required to provide high-quality health services. It argued that government should refocus on improving the public's health by turning the NHS into a health service not a sickness service. It concluded that, 'thanks to the health outcome benefits associated with investment in public health, the UK would find itself much better placed to deal with the [financial] pressures [if it adopts] a step change in the way public health is viewed, resourced and delivered nationally'.

The second Wanless report (2004) investigated the cost effectiveness of improving the health of the whole population and reducing health inequalities (using the 'fully engaged' scenario set out in the first report). It explicitly stated that 'to achieve the objective of an efficient allocation of NHS funding between health care and public health, a similar method of cost-effectiveness analysis needs to be applied to public health and clinical interventions'. Wanless recommended that 'a consistent framework should be used to evaluate the cost effectiveness of interventions and initiatives across health care and public health'. This framework should allow a comparison of the cost effectiveness of public health interventions:

- within and between risk factors and disease areas
- involving screening and treatment within and across disease areas
- directed towards tackling the wider determinants of health.

Ultimately, Wanless envisaged that such a framework 'would demonstrate the respective merits of investment in public health measures compared to clinical

services, and so enable policymakers and practitioners to prioritise the use of scarce resources’.

5.2 *Perspective*

As defined by the Institute’s statutory instruments, economic analysis relating to public health guidance should adopt the public sector, NHS and personal social services perspectives. When appropriate, results may also be presented from other perspectives.

5.3 *Economic evaluation method*

Cost-effectiveness analysis, with the quality-adjusted life year (QALY) as the health-related outcome measure, is the primary measure used. This ensures baseline comparability within the UK healthcare sector and across the Institute’s programmes.

In addition, a ‘cost-consequence’ approach may be adopted to take into account the complexity and multidimensional character of public health interventions and programmes. This ensures the guidance development process remains fair and transparent and allows explicit consideration of multiple, non-health related and/or non-quantifiable outcomes by the CPHE committee(s). It also means that issues such as equity and distribution, which are key to public health policy, can be used to inform the analysis.

5.4 *Identifying topics for economic analysis*

Given the broad scope of many programmes and interventions, it will not be possible to conduct original analyses for every component. If priority areas need to be identified, the following criteria should be used:

- the proposed topic has major resource implications
- current public health practice may be challenged
- sufficient data of adequate quality are available to allow useful modelling
- there is a lack of consensus among public health professionals.

Economic modelling may not be warranted, for example, if the published evidence is so reliable that further analysis would be superfluous – or if the public health evidence is so uncertain that it is not possible to make even a broad estimate of cost effectiveness. However, in cases where the public health evidence base is weak, it may be necessary to model changes in outcomes before deciding whether or not to carry out further economic analysis.

5.5 *Reviewing the economic literature*

A systematic review of the published economic literature should be carried out to ensure that no economic evaluations are missed during searches undertaken in the review of effectiveness.

Searches should be carried out using economic search filters developed by the Centre for Reviews and Dissemination and used extensively (see 5.10 further reading). A review and summary of the evidence should be carried out by a reviewer trained in health economics.

A thorough systematic review should be attempted, but if there is a large amount of economic evidence, it may be necessary to limit the search (for example, to UK-based studies, a specific date range, or to full economic evaluations carried out alongside RCTs). The health economist should be involved at the earliest opportunity, usually at the scoping stage, to estimate the amount and quality of the economics evidence.

Papers identified for inclusion should be critically appraised using a validated checklist that includes the economic evaluation criteria detailed in NICE methods guidance (see appendix A). In addition, a commentary should be presented on the quality of each paper.

5.6 *Cost-effectiveness analysis*

Additional economic analyses are likely to be required for interventions identified in the effectiveness review (as the health-economic literature is unlikely to be comprehensive and conclusive enough). The CPHE Project Team should work

with PHIAC or the PDG to select topics for further economic analysis, in liaison with the contractors undertaking the effectiveness review.

If an existing analytical model cannot be easily adapted to answer the question, new models should be developed. The models used to synthesise evidence to generate estimates of outcomes and cost effectiveness should follow accepted guidelines.

5.7 Modelling approaches

NICE uses cost-effectiveness analysis – specifically, cost-utility analysis (with the QALY as the health-related outcome measure) – to produce an economic evaluation of public health interventions and programmes. This ensures baseline comparability within the UK healthcare sector and across the Institute’s programmes. It also assists with the prioritisation of recommendations for local implementation.

QALYs weight the life expectancy of a person against an estimate of their health-related quality of life (measured on a 0–1 scale). There are methodological problems with them and these are well documented (however, the same is true of other approaches). Specific guidance is provided in the Institute’s ‘Guide to the methods of technology appraisal’ (available from the NICE website: www.nice.org.uk?page.aspx?o=201974). In particular, the manual recommends a ‘reference case’ analysis which:

- includes all health effects on individuals
- measures costs from the public sector, NHS and personal social services perspectives
- does not apply equity weightings
- discounts costs and health outcomes at 3.5%
- uses choice-based elicitation methods, a representative sample of the general population, and validated, generic, health-state instruments to value health-

related quality of life. (There is unlikely to be time to collect original quality of life valuations, so data collected by alternative methods may be used)

- covers a time span that incorporates all important costs and effects.

Departures from this reference case should be highlighted and reasons given.

If there is not enough good quality data to estimate QALYs gained, an alternative measure might be considered for the cost-effectiveness analysis (such as life years gained, cases averted or some more disease-specific outcome).

In addition, a cost-consequence approach may be used with the cost per QALY calculations to allow PHIAAC or the PDG to consider multiple, non-health related and/or difficult to quantify factors (for example, equity or distributional considerations). These factors can be considered when they may:

- modify a decision based solely on cost per QALY considerations
- help communicate decisions to stakeholders and the public.

The cost-consequence approach also ensures that guidance development is transparent.

5.8 General principles

Regardless of the approach taken, the following principles are observed:

- economic analysis is underpinned by the best public health evidence
- there is the highest level of transparency in the reporting of methods
- any uncertainty (around both internal and external validity) is discussed fully and explored by sensitivity analysis (and, where data allow, statistical analysis)
- the limitations of the approach and methods are fully discussed
- conventions on reporting economic evaluations are followed (see Drummond and Jefferson 1996)

- analysis is carried out in collaboration by the health economist and the CPHE guidance groups.

5.9 Economic evidence and guidance recommendations

NICE encourages PHIAC or the PDG to consider recommendations that may be slightly less effective than current practice, provided they free up a substantial amount of resources that could be reinvested. They are also encouraged to consider recommendations that increase public health effectiveness at an acceptable level of increased cost.

Cost effectiveness and intervention effectiveness are discussed in parallel when formulating recommendations.

As most members of PHIAC or the PDG are not experts in economic analysis, it is important that the authors of the economic analysis communicate pertinent issues on cost effectiveness (such as marginal or incremental analysis) clearly and concisely.

5.10 References and further reading

Briggs AH, Goeree R, Blackhouse G et al. (2002) Probabilistic analysis of cost effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Medical Decision Making* 22:290–308.

Drummond MF, Jefferson TO (1996) Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal* 313:275–83.

Drummond MF, O'Brien B, Stoddart GL et al. (1997) *Methods for the economic evaluation of health care programmes*. 2nd edition. Oxford: Oxford Medical Publications.

National Institute for Clinical Excellence (2001) *Technical guidance for manufacturers and sponsors on making a submission to a technology appraisal*. London: National Institute for Clinical Excellence.

NHS Centre for Reviews and Dissemination (2001) Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. *CRD report number 4*. 2nd edition. NHS Centre for Reviews and Dissemination, University of York.

www.york.ac.uk/inst/crd/report4.htm

Palmer S, Raftery J (1999) Economics notes. Opportunity cost. *British Medical Journal* 318:1551–2.

Spiegelhalter DJ, Myles JP, Jones DR et al. (2000) Bayesian methods in health technology assessment: a review. *Health Technology Assessment* 4:1–136.

Wanless D (2002) *Securing our future health: taking a long-term view*. London: HM Treasury.

Wanless D (2004) *Securing good health for the whole population*. London: HM Treasury.

6 Creating recommendations

Some users may focus on the recommendations, rather than reading the full guidance. It is therefore vital that the recommendations are:

- clear
- based on the best available evidence of effectiveness
- based on the best available evidence of cost effectiveness
- implementable.

This chapter describes the four key stages involved in creating clear, practical recommendations:

1. translating the evidence into draft recommendations
2. revising them following fieldwork
3. wording the recommendations
4. prioritising them.

It also describes a framework for selecting and formulating research recommendations.

6.1 *Translating the evidence into recommendations*

NICE public health recommendations are not graded, but they are formulated and prioritised based on the:

- strength (quality and quantity) of the supporting evidence and its applicability to the populations and settings in question (that is, the evidence statements – see chapter 4)
- importance of the outcomes (including impact on inequalities)
- size of effect and potential impact on individual and population health
- cost effectiveness
- any other considerations (for example, risks to health and implementability).

PHIAC or the PDG determines the specific criteria for deriving and prioritising recommendations for each piece of guidance and the process is made explicit. The following documents are available to members to help them develop draft recommendations.

- NICE review(s) including evidence tables, draft evidence statements and the economic analysis (including modelling).
- Synopsis report (a summary of the review(s) and economic review) including draft evidence statements.

Members of PHIAC or the PDG examine the evidence and discuss its suitability to answer the key questions, before drafting recommendations with the help of the CPHE Project Team. If the evidence is very strong (that is, consistent and of good quality), directly applicable and there is good evidence to suggest the intervention is implementable, then it should translate directly into a draft recommendation. Often, however, this is not possible (see table 6.1).

The recommendations may be prioritised (see section 6.4). PHIAC or the PDG may also determine that research should be a condition for implementing an intervention, where evidence is lacking.

PHIAC or the PDG makes explicit the factors that have been considered and the methods used to achieve consensus, so that the process is as transparent as possible. Recommendations from other guidance are not to be quoted verbatim. Recommendations from NICE guidance or NHS policy (for example, national service frameworks (NSFs)) can be quoted verbatim.

Table 6.1 Translating evidence into recommendations: challenges and possible solutions

Challenge	Possible solution
Literature search found no evidence that answers the question or Poor quality evidence	PHIAC/the PDG considers current best practice. This process should be robust and issues may be resolved through group/committee discussions
Only evidence of similar type and quality is available and it conflicts	PHIAC/the PDG should identify studies that are most applicable to the target population and base the recommendations on them
Evidence not directly applicable to the target population (for example, it covers a different age group)	PHIAC/the PDG may extrapolate the recommendations from this evidence (for example, from high quality evidence drawn from a largely similar but different population group). PHIAC/the PDG needs to make its approach explicit, stating the basis used for extrapolating from the data and the assumptions made.

Formal consensus methods may be used to agree the final recommendations. Whatever approach is used, the proceedings are recorded and a clear statement is made about how areas of disagreement have been handled.

6.2 Testing the recommendations: fieldwork

Fieldwork is carried out to test the draft recommendations against the experience of practitioners.

The fieldwork method is described in chapter 7. A summary of the collated and analysed data and key pointers arising from the material – including the implications for local and national public health policy and practice – is presented as a fieldwork report. This makes extensive use of quotations to provide the reader with the ‘voice of the practitioner’.

6.2.1 PHIAC or the PDG meeting following fieldwork

PHIAC or the PDG meets to review the evidence in the light of fieldwork data, revise the recommendations, if necessary, and finalise the guidance. It uses the following documents:

- the fieldwork report
- a summary of how the fieldwork data impacts on the draft recommendations.

6.3 *Wording the recommendations*

Great care should be taken when wording the recommendations, as they will be the focus of attention for most readers. NICE asks PHIAC and the PDGs to ensure the recommendations meet the following criteria.

- Recommendations contain sufficient information so that they can stand alone and be understood without reference to supporting material. Terminology should be avoided where possible – where not, it needs to be clearly defined and unambiguous.
- They focus on what needs to be done (they should not contain background information). When writing recommendations, answer the reader's question: 'what does this mean for me?' Each recommendation should specify four elements: target user (who?), intervention/action (what?), target population (with whom?), context/circumstances (where/when?)
- Reference to products (for example, pedometers) and services (for example, slimming clubs) is in general terms, not by company name.
- Tables are used only if they substantially improve clarity (note, tables can cause difficulties when the document is posted on the website).
- The use of the word 'subjects' is avoided – 'people' or 'patients' is used instead.
- Generally the advisory 'should' is used, rather than the imperative 'must'. (However, occasionally 'must' will be required – for example, when the recommendation links to enforceable legislation, such as Health and Safety regulations.)

- They should provide a clear link to the supporting evidence.

6.4 *Prioritising recommendations for implementation*

Public health guidance often contains multiple recommendations. To help users decide which order to implement them in, key priorities for implementation are also identified.

The number of prioritised recommendations may vary, depending on the topic area, but there are likely to be several for each target user group. Typically, intervention guidance will have around 10 recommendations and programme guidance will have up to 25.

PHIAC or the PDG makes the final decision on key priorities for implementation – there is no blueprint for this process. However, the following criteria are taken into account:

- impact on improving health and/or reducing inequalities in health
- degree of change in practice (if possible, key recommendations already being implemented are avoided)
- cost effective use of resources
- balance of risks and benefits
- values.

Information from various sources, including national audit data and the experience of professionals or lay members involved in PHIAC or the PDG, is used to identify implementation priorities. The choice is also informed by a broad range of stakeholders, as part of the formal consultation process.

There should be a clear record detailing the criteria used for selecting the key priorities for implementation and the process used for agreeing them.

6.5 Formulating research recommendations

PHIAC or the PDG is likely to identify areas where good evidence is lacking and where future research is needed. This section provides a framework for formulating and selecting research recommendations.

6.5.1 Principles for formulating research recommendations

Research recommendations can cover questions about effectiveness, implementation, acceptability, feasibility and costs. They may also call for a full systematic review on a specific topic. Each research recommendation is formulated as one question, or a set of closely related questions. It considers the importance of issues relating to gender, ethnicity and people with special needs.

6.5.2 Selecting research recommendations

PHIAC or the PDG selects a maximum of five research recommendations for public health intervention guidance and 10 for public health programme guidance. These are the most important research recommendations, as defined by the criteria in table 6.5. A research recommendation has two components:

- a well formulated, answerable question (see table 6.4)
- a statement about the importance of the recommended research (see table 6.5).

The selected recommendations are added to a database on the NICE website, and are prioritised by NICE. For more details, see the NICE website (www.nice.org.uk).

They comprise a question with explanatory text of not more than 150 words. The following example is a research question from the guidelines for brief interventions and referral for smoking cessation developed by the CPHE.

Question:

Which brief interventions work best for increasing smoking cessation among lower socio-economic and vulnerable groups?

Explanatory paragraph:

Smoking remains the leading cause of preventable morbidity and premature mortality in England, causing an estimated annual average of 86,500 deaths between 1998 and 2002. There is a clear social class gradient in smoking and it accounts for over half of the difference in risk of premature death between social classes. Smoking prevalence remains worryingly high in some groups.

Vulnerable groups within society are the groups most likely to bear the burden of ill health with the fewest resources with which to cope.

6.5.3 Other research recommendations

Other important research recommendations lying outside the public health intervention guidance scope are communicated to research and development funders such as: the National Coordinating Centre for Health Technology Assessment, the NHS Service Delivery and Organisation Programme, the Medical Research Council (MRC), the Economic and Social Research Council (ESRC), or the DH Policy Research Programme (PRP).

Table 6.4 Checklist for formulating research questions using the PICO model

PICO is a widely used mnemonic summarising the four major components of every clinical or research question: patient (population), intervention, comparison and outcome.

Population and/or problem	<p>What is the primary problem, disease or condition you are interested in? What are the most important characteristics of the population to be studied?</p> <p><i>Consider:</i></p> <ul style="list-style-type: none"> • determinants of health, health status • gender, age, ethnic group, specific exclusions • setting
Intervention or indicator	<p>Which main intervention are you considering? What determinants of risk are important?</p> <p><i>Consider:</i></p> <ul style="list-style-type: none"> • type, frequency, duration (for intervention or exposure)
Comparison or control	<p>What is the main alternative(s) or control to compare with the intervention?</p> <p><i>Consider:</i></p> <ul style="list-style-type: none"> • all the parameters mentioned above under population and intervention, where applicable
Outcome	<p>What will the researcher need to measure, improve, influence or accomplish? What intervention outcomes should be measured?</p> <p><i>Consider:</i></p> <ul style="list-style-type: none"> • outcomes to be measured (for example, mortality, morbidity, quality of life, intermediate outcomes, health promotion) • method of measurement (type, frequency or timing) • the need for blinding of target populations, provider or outcome assessor.

Table 6.5 Why is this question important?

Draft a paragraph explaining the need for research, using the following headings. PHIAAC or the PDG can then use these headings to select the five most important research recommendations.

1. Relevance to NICE	How would the research change future NICE guidance?
2. Importance to the population	What would be the impact of any new or amended guidance? (For example, on quality of life, morbidity or disease prevalence, severity or mortality)
3. Relevance to the NHS and the public sector	What would be the impact of any new or amended guidance – on the NHS and the public sector? (For example, financial advantages, effect on staff, impact on strategic planning or service delivery)
4. National priorities	Is the question relevant to a national public health priority area (such as a NSF or 'Choosing health')? Specify the relevant document
5. Lack of current evidence	How much research has been carried out in this area? What are the problems, if any, with previous research? Provide details of any previous systematic review
6. Feasibility	Can it be carried out in a realistic timescale and at an acceptable cost?
7. Other comments	Mention any other important issues, such as potential funders, or the outcome of previous attempts to address this issue. However, remember that this is not a research protocol.

NOTE: Continuous feedback by users is essential to improve the above list of questions. How user-friendly, acceptable, reliable and valid is it? How successful is it in generating clear and precise recommendations for feasible studies?

6.6 Further reading

Glasziou P, Del Mar C, Salisbury J (2003) *Evidence-based medicine workbook*. London: BMJ Books.

Kelly MP, Chambers J, Huntley J et al. (2004) *Method 1 for the production of effective action briefings and related materials*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=507904

Michie S and Johnston M (2004) Changing clinical behaviour by making guidelines specific. *British Medical Journal* 328: 343-345.

National Institute for Clinical Excellence (2004) *Research and development strategy*. www.nice.org.uk/page.aspx?o=114221 [accessed 3 March 2005]

Sackett DL, Straus SE, Richardson WS (2000) *Evidence-based medicine: how to practice and teach EBM*. Second edition. Edinburgh: Churchill Livingstone.

Schunemann HJ, Best D, Vist G et al. for the GRADE Working Group (2003) Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *Canadian Medical Association Journal* 169:677–80.

Scottish Intercollegiate Guidelines Network (2002) *SIGN 50. A guideline developer's handbook*. Edinburgh: Scottish Intercollegiate Guidelines Network.

Weightman A, Ellis S, Cullum A, et al. (2005) *Grading evidence and recommendations for public health interventions: developing and piloting a framework*. London: Health Development Agency.

7 Testing the recommendations: fieldwork

7.1 Introduction

Scientific evidence gathered from the earlier phases of guidance development determines the draft recommendations (see chapter 6). The fieldwork phase aims to use practitioner experience and views to enhance these recommendations. Practitioner knowledge, gathered from the fieldwork, provides the basis for understanding whether and how particular interventions will work.

The method used is derived from:

- two pilot ‘evidence into practice’ studies undertaken by the Health Development Agency (HDA) during 2002/3 (Kelly et al. 2004)
- the work of the HDA’s public health collaborating centres in developing ‘Effective action briefings’ during 2004/5.

7.2 Guiding principles

Recommendations initially drafted by PHIAC or the PDG are based on the scientific plausibility of an intervention or programme (principally, based on judgements presented as evidence of effectiveness). A further analysis of the way in which they might be implemented in routine practice – as distinct from controlled experimental and/or evaluative conditions – is needed to judge their ‘likelihood of success’.

This type of information is very rarely found in published scientific papers or reviews. Fieldwork allows practitioner knowledge to inform the recommendations in the guidance.

The following principles underpin this phase.

- Successful implementation of guidance depends on evidence-based recommendations informed by practical experience.
- Practitioners know what is feasible – and the context in which public health guidance and interventions are delivered and experienced.

- Good quality fieldwork elicits practitioner knowledge in a transparent, reliable and systematic way.
- Draft recommendations are presented as indicative of what might work, not as prescriptions about what will work.
- Evidence and recommendations presented to fieldwork participants is appropriate to their needs and accessible.

7.3 *Initiating fieldwork meetings*

The process for initiating fieldwork meetings is summarised below. Information about preliminary work in this area can be found in Kelly et al. (2004).

- At least 6 day-long fieldwork meetings are convened for programme guidance and four meetings for intervention guidance. These take place across the regions.
- Independent professional facilitators are selected to lead each meeting. They do not necessarily require a background in the topic under consideration. However, they do need a good knowledge of the public health community and the methods used to translate research evidence into practice. They also need an understanding of the working methods of the public sector and some knowledge of health promotion.
- A purposive sample of up to 35 fieldworkers – professionals (not researchers) with a remit for the relevant public health topic – are invited to each meeting.
- If the topic area means that researchers need to be included, a separate meeting is convened for them, using the same processes.

7.4 *Fieldwork participant criteria*

Participants are chosen to ensure they represent a broad range of disciplinary groups within the statutory, non-statutory and voluntary sectors, including those working with disadvantaged or minority groups. They include, for example, commissioners, outreach workers, police and other emergency services, probation services, GPs, health visitors and educational welfare officers.

They should also:

- actively promote relevant public health messages
- have relevant experience and knowledge of the topic
- operate mainly in a regional and local capacity.

They can also be professionals working indirectly to promote the desired health gain.

7.5 *The fieldwork meetings*

7.5.1 Facilitation

Independent facilitators attend each workshop meeting, along with a member of the CPHE Project Team (who presents the existing evidence statements and draft recommendations and answers any technical questions).

7.5.2 Recording meetings

The plenary discussions and group work are either tape recorded and subsequently transcribed, or captured by a stenographer. All participants consent in writing to the recording – and to its subsequent use in discussions and group work.

During group work, the facilitator guides a scribe (someone recording the points made using a flipchart) to ensure an accurate record of points raised against each question (mediating factors, barriers and solutions). They also categorise each point according to whether it relates to action at the macro, micro or social level. (For example, whether it relates to policy, PCT targets or cultural needs, respectively.) Previous experience has shown that stenography is the best way of recording and transcribing this information.

7.5.3 Structure

Presentations are kept succinct to make best use of time and strike the right balance between passive and active participation. Time is required to allow for a full discussion.

Meetings are based on the following format.

7.5.3.1 Session 1: describes the purpose of the meeting

The lead facilitator explains the aims and objectives for the day, and the values underpinning NICE's fieldwork process (see section 7.2). The facilitator describes the guidance development process, introduces the draft recommendations and describes how participants will help refine them. They should make explicit reference to inequalities – and to the importance of judging the impact of interventions on different segments of the population.

The initial presentations should take around 20 minutes.

7.5.3.2 Session 2: explores participants' working environments

Participants consider the draft recommendations and comment on the context in which they operate: social, political and economic factors relevant to participants' work – and the communities that they serve – may be raised here. This session lasts for up to an hour.

7.5.3.3 Session 3: group work – appraisal of the draft recommendations and evidence statements

Participants are divided into four or five multidisciplinary groups, each working with a facilitator. Each group appraises the draft recommendations and their supporting evidence statements. Discussion focuses on the following question:

'Given that the evidence suggests that a particular kind of intervention/activity has worked in the following circumstances, and that this should form the basis of a recommendation, what would need to be done to make it work in your local situation?' A follow up prompt is: 'If this would not work, why not – and what would?'

Groups are also asked to address the implications for health inequalities and, generally, for their own practice/profession. To conclude, discussion considers barriers to – and facilitators for – change, including potential local drivers for change.

Other issues that may be raised include:

- political drivers and imperatives for activity planning
- decision and influence
- partnerships
- budgets
- stakeholders
- consultation
- commissioning
- shared data and IS
- performance management
- examples of local good practice.

7.5.3.4 Session 4: feedback

The Lead Facilitator provides plenary feedback on the day in relation to implementation barriers, opportunities and solutions. Completed case study templates are collected, and additional templates distributed if needed (to capture any additional information raised or prompted in the course of the day). These are collected at close of session, or returned to the team by a specified date. Forms are clearly marked with instructions for completion and return.

7.5.3.5 Evaluation

At the end of the day, evaluation sheets are completed by all participants. After the meeting the facilitators and members of the CPHE Project Team meet together to share notes and transcripts that the team leader will use as a basis for the fieldwork report of the meeting. All original notes are retained. This fieldwork report is circulated to all participants to check for accuracy. Fieldwork reports from all the meetings form the basis for the fieldwork technical report considered by PHIAC or the PDG.

7.6 *Fieldwork analysis*

7.6.1 Data presentation

Proceedings from the fieldwork meetings are transcribed in full (electronically or by hand). Transcripts are available and a summary of the outcomes of the day is distributed to participants when the fieldwork report is disseminated.

7.6.2 Analysis

Data analysis begins as soon as possible following transcription, preferably in time for inclusion at the next fieldwork meeting. Analysis may be performed using qualitative research software, or by hand, but the method is fully reported in the fieldwork technical report.

The fieldwork data is broken down into common and consistent themes using a content analysis approach. Usually, one researcher prepares an initial analysis which is then verified by 'blind' coding and sorting of a sample of the transcript by a second researcher. For an example of this kind of analysis, refer to part three (chapters 7–13) of Silverman D (2000) – see 7.10 further reading for full reference.

Once the analysis is complete, participant quotes are selected to illustrate each theme.

A summary of the qualitative fieldwork data is prepared to inform the fieldwork technical report.

7.7 *Fieldwork report*

The fieldwork report sent to PHIAC or the PDG is a summary of analysed fieldwork data and key points arising from it collated from all the fieldwork meetings.

7.7.1 Style and transcription notation

The fieldwork report describes the fieldwork meetings and questions posted to participants. It also summarises emerging themes from the data, which are illustrated with verbatim quotes from participants.

- Quotes should not be edited, other than to clarify where text is not clear: when an extra word is needed to make sense of a quote, it should be put in square brackets [] to indicate a word has been inserted.
- Short quotes should be inserted into the text and should be clearly marked with single quotation marks (' '). Longer quotes should be presented as inset paragraphs with single quotation marks.
- The use of ... can be used to indicate words have been missed from the direct quote (to present a more coherent quote for the reader), but the omission must not alter the meaning of the original quotation.
- Quotes should be anonymous (for example, they should refer to a transcript number and line number within the transcript).

7.8 Finalising recommendations

PHIAC or the PDG uses the fieldwork report to refine and prioritise the draft recommendations. Essentially, the fieldwork data is used as another source of evidence on implementation. The fieldwork report also helps PHIAC or the PDG make the recommendations behaviourally specific.

Please refer to chapter 6 for further details about formulating and prioritising recommendations.

7.9 Further reading

Kelly MP, Chambers J, Huntley J et al. (2004) *Method 1 for the production of effective action briefings and related materials*. London: Health Development Agency. www.nice.org.uk/page.aspx?o=507904

Silverman D (2004) (Editor) *Doing qualitative research: A practical handbook*. London: Sage Publications.

Appendix A

Methodology checklists

A.1 Methodology checklist: systematic reviews and meta-analyses

Study identification <i>Include author, title, reference, year of publication</i>			
Guideline topic		Key question no:	
Checklist completed by:			
SECTION 1: INTERNAL VALIDITY			
In a well-conducted systematic review:		In this study this criterion is: <i>(Circle one option for each question)</i>	
1.1	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	A description of the methodology used is included.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	The literature search is sufficiently rigorous to identify all the relevant studies.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	Study quality is assessed and taken into account.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	There are enough similarities between the studies selected to make combining them reasonable.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY		
2.1	How well was the study done to minimise bias? <i>Code ++, + or –</i>	
2.2	If coded as + or – what is the likely direction in which bias might affect the study results?	
SECTION 3: DESCRIPTION OF THE STUDY <i>Please print answers clearly</i>		
3.1	What types of study are included in the review? (<i>Highlight all that apply</i>)	RCT CCT Cohort Case-control Other
3.2	How does this review help to answer your key question? <i>Summarise the main conclusion of the review and how it relates to the relevant key question. Comment on any particular strengths or weaknesses of the review as a source of evidence for a guideline produced for the NHS in Scotland.</i>	

A.1 Notes on the use of methodology checklist: systematic reviews and meta-analyses

Section 1 identifies the study and asks a series of questions aimed at establishing the internal validity of the study under review – that is, making sure that it has been carried out carefully, and that the outcomes are likely to be attributable to the intervention being investigated. Each question covers an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the review.

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed
(that is, not mentioned, or indicates that this aspect of study design was ignored)
- Not reported
(that is, mentioned, but insufficient detail to allow assessment to be made)
- Not applicable

A.1.1 The study addresses an appropriate and clearly focused question

Unless a clear and well-defined question is specified in the report of the review, it will be difficult to assess how well it has met its objectives or how relevant it is to the question you are trying to answer on the basis of the conclusions.

A.1.2 A description of the methodology used is included

One of the key distinctions between a systematic review and a general review is the systematic methodology used. A systematic review should include a detailed description of the methods used to identify and evaluate individual studies. If this description is not present, it is not possible to make a thorough evaluation of the quality of the review, and **it should be rejected as a source of level 1 evidence** (though it may be useable as level 4 evidence, if no better evidence can be found).

A.1.3 The literature search is sufficiently rigorous to identify all the relevant studies

A systematic review based on a limited literature search – for example, one limited to Medline only – is likely to be heavily biased. A well-conducted review should as a minimum look at Embase and Medline, and from the late 1990s onward, the Cochrane Library. Any indication that hand searching of key journals, or follow up of reference lists of included studies were carried out in addition to electronic database searches can normally be taken as evidence of a well-conducted review.

A.1.4 Study quality is assessed and taken into account

A well-conducted systematic review should have used clear criteria to assess whether individual studies had been well conducted before deciding whether to include or exclude them. If there is no indication of such an assessment, **the review should be rejected as a source of level 1 evidence**. If details of the assessment are poor, or the methods are considered to be inadequate, the quality of the review should be downgraded. In either case, it may be worthwhile obtaining and evaluating the individual studies as part of the review you are conducting for this guideline.

A.1.5 There are enough similarities between the studies selected to make combining them reasonable

Studies covered by a systematic review should be selected using clear inclusion criteria (see question B.1.4 above). These criteria should include, either implicitly or explicitly, the question of whether the selected studies can legitimately be compared. It should be clearly ascertained, for example, that the populations covered by the studies are comparable, that the methods used in the investigations are the same, that the outcome measures are comparable and the variability in effect sizes between studies is not greater than would be expected by chance alone.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

++	All or most of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
+	Some of the criteria have been fulfilled. Those criteria that have not been fulfilled or not adequately described are thought unlikely to alter the conclusions.
-	Few or no criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

The code allocated here, coupled with the study type, will decide the **level of evidence** that this study provides.

The aim of the other two questions in this section is to summarise your view of the quality of this study and its applicability to the patient group targeted by the guideline you are working on.

Section 3 asks you to summarise key points about the study that will be used when you come to formulate recommendations at a later stage of the process.

A.2 Methodology checklist: randomised controlled trials

Study identification <i>Include author, title, reference, year of publication</i>			
Guideline topic:		Key question no:	
Checklist completed by:			
SECTION 1: INTERNAL VALIDITY			
In a well-conducted RCT study:		In this study this criterion is: (Circle one option for each question)	
1.1	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	The assignment of subjects to intervention groups is randomised.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	An adequate concealment method is used.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	Subjects and investigators are kept 'blind' about intervention allocation.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	The intervention and control groups are similar at the start of the trial.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	The only difference between groups is the intervention under investigation.	Well covered Adequately addressed	Not addressed Not reported Not applicable

		Poorly addressed	
1.7	All relevant outcomes are measured in a standard, valid and reliable way.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.8	What percentage of the individuals or clusters recruited into each intervention arm of the study dropped out before the study was completed?		
1.9	All the subjects are analysed in the groups to which they were randomly allocated (often referred to as intention-to-treat analysis).	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.10	Where the study is carried out at more than one site, results are comparable for all sites.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
SECTION 2: OVERALL ASSESSMENT OF THE STUDY			
2.1	How well was the study done to minimise bias? <i>Code ++, + or –</i>		
2.2	If coded as + or – what is the likely direction in which bias might affect the study results?		
2.3	Taking into account clinical considerations, your evaluation of the methodology used, and the statistical power of the study, are you certain that the overall effect is due to the study intervention?		

2.4	Are the results of this study directly applicable to the patient group targeted by this guideline?	
-----	--	--

SECTION 3: DESCRIPTION OF THE STUDY (The following information is required to complete evidence tables facilitating cross-study comparisons. Please complete all sections for which information is available) PLEASE PRINT CLEARLY		
3.1	<p>How many individuals are included in this study?</p> <p><i>Please indicate number in each arm of the study, at the time the study began.</i></p>	
3.2	<p>What are the main characteristics of the population?</p> <p><i>(Include all relevant characteristics – for example, age, sex, ethnic origin, comorbidity, disease status, community/hospital based)</i></p>	
3.3	<p>What intervention (treatment, procedure) is being investigated in this study?</p> <p><i>List all interventions covered by the study.</i></p>	
3.4	<p>What comparisons are made in the study?</p> <p><i>Are comparisons made between interventions, or between intervention and placebo/no intervention?</i></p>	
3.5	<p>How long are individuals followed up in the study?</p> <p><i>Length of time individuals are followed from beginning participation in the study. Note specified end point used to decide end of follow-up (for</i></p>	

	<i>example, death, complete cure). Note if follow-up period is shorter than originally planned.</i>	
3.6	<p>What outcome measure(s) are used in the study?</p> <p><i>List all outcomes that are used to assess effectiveness of the interventions used.</i></p>	
3.7	<p>What size of effect is identified in the study?</p> <p><i>List all measures of effect in the units used in the study – for example, absolute or relative risk, number needed to treat. Include p values and any confidence intervals that are provided.</i></p>	
3.8	<p>How was this study funded?</p> <p><i>List all sources of funding quoted in the article, whether Government, voluntary sector or industry.</i></p>	
3.9	<p>Does this study help to answer your key question?</p> <p><i>Summarise the main conclusions of the study and indicate how it relates to the key question.</i></p>	

A.2 Notes on the use of methodology checklist: randomised controlled trials

Section 1 identifies the study and asks a series of questions aimed at establishing the internal validity of the study under review – that is, making sure that it has been carried out carefully, and that the outcomes are likely to be attributable to the intervention being investigated. Each question covers an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study.

- Well covered.
- Adequately addressed.
- Poorly addressed.
- Not addressed (that is, not mentioned, or indicates that this aspect of study design was ignored).
- Not reported (that is, mentioned, but insufficient detail to allow assessment to be made).
- Not applicable.

A.2.1 The study addresses an appropriate and clearly focused question

Unless a clear and well-defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer on the basis of its conclusions.

A.2.2 The assignment of subjects to intervention groups is randomised

Random allocation of individuals to receive one or other of the interventions under investigation, or to receive either intervention or placebo, is fundamental to this type of study. **If there is no indication of randomisation, the study should be rejected.** If the description of randomisation is poor, or the process used is not truly random (for example, allocation by date, alternating between one group and another) or can otherwise be seen as flawed, the study should be given a lower quality rating.

A.2.3 An adequate concealment method is used

Research has shown that where allocation concealment is inadequate, investigators can overestimate the effect of interventions by up to 40%. Centralised allocation, computerised allocation systems or the use of coded identical containers would all be regarded as adequate methods of concealment, and may be taken as indicators of a well-conducted study. If the method of concealment used is regarded as poor, or relatively easy to subvert, the study must be given a lower quality rating, and can be rejected if the concealment method is seen as inadequate.

A.2.4 Subjects and investigators are kept 'blind' about intervention allocation

Blinding can be carried out up to three levels. In single-blind studies, individuals are unaware of which intervention they are receiving; in double-blind studies the doctor and the individual are unaware of which intervention the individual is receiving; in triple-blind studies individuals, healthcare providers and those conducting the analysis are unaware of which individuals received which intervention. The higher the level of blinding, the lower the risk of bias in the study.

A.2.5 The intervention and control groups are similar at the start of the trial

Individuals selected for inclusion in a trial should be as similar as possible, in order to eliminate any possible bias. The study should report any significant differences in the composition of the study groups in relation to gender mix, age, stage of disease (if appropriate), social background, ethnic origin or comorbid conditions. These factors may be covered by inclusion and exclusion criteria, rather than being reported directly. Failure to address this question, or the use of inappropriate groups, should lead to the study being downgraded.

A.2.6 The only difference between groups is the intervention under investigation

If some individuals received additional treatment, even if of a minor nature or consisting of advice and counselling rather than a physical intervention, this treatment is a potential confounding factor that may invalidate the results. **If groups were *not* treated equally, the study should be rejected unless no other evidence is available.** If the study *is* used as evidence it should be treated with caution, and given a low quality rating.

A.2.7 All relevant outcomes measured in a standard, valid and reliable way

If some significant clinical outcomes have been ignored, or not adequately taken into account, the study should be downgraded. It should also be downgraded if the measures used are regarded as being doubtful in any way, or applied inconsistently.

A.2.8 What percentage of the individuals or clusters recruited into each intervention arm of the study dropped out before the study was completed?

The number of individuals that drop out of a study should give concern if the number is very high. Conventionally, a 20% drop-out rate is regarded as acceptable, but this may vary. Some regard should be paid to *why* individuals dropped out, as well as how many. It should be noted that the drop-out rate

may be expected to be higher in studies conducted over a long period of time. A higher drop-out rate will normally lead to downgrading, rather than rejection of a study.

A.2.9 All the subjects are analysed in the groups to which they were randomly allocated (often referred to as intention-to-treat analysis)

In practice, it is rarely the case that all individuals allocated to the intervention group receive the intervention throughout the trial, or that all those in the comparison group do not. Individuals may refuse treatment, or contra-indications arise that lead them to be switched to the other group. If the comparability of groups through randomisation is to be maintained, however, individual outcomes *must* be analysed according to the group to which they were originally allocated, irrespective of the intervention they actually received. (This is known as *intention-to-treat* analysis.) If it is clear that analysis was not on an intention-to-treat basis, the quality of the study should be downgraded.

A.2.10 Where the study is carried out at more than one site, results are comparable for all sites

In multi-site studies, confidence in the results should be increased if it can be shown that similar results were obtained at the different participating centres. **Section 2** relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

++	All or most of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
+	Some of the criteria have been fulfilled. Those criteria that have not been fulfilled or not adequately described are thought unlikely to alter the conclusions.
-	Few or no criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

The code allocated here, coupled with the study type, will decide the **level of evidence** that this study provides.

The aim of the other two questions in this section is to summarise your view of the quality of this study and its applicability to the patient group targeted by the guideline you are working on.

Section 3 asks you to summarise key points about the study that will be added to an evidence table at the next stage of the process.

A.3 Methodology checklist: case–control studies

Study identification <i>Include author, title, reference, year of publication</i>			
Guideline topic		Key question no:	
Checklist completed by:			
SECTION 1: INTERNAL VALIDITY			
In a well conducted case–control study:		In this study the criterion is: (<i>Circle one option for each question</i>)	
1.1	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
SELECTION OF SUBJECTS			
1.2	The cases and controls are taken from comparable populations.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	The same exclusion criteria are used for both cases and controls.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	What percentage of each group (cases and controls) participated in the study?	Cases: Controls:	
1.5	Comparison is made between participants and non-participants to establish their similarities or differences.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	Cases are clearly defined and differentiated from controls.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	Is it clearly established that controls are non-cases?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

ASSESSMENT			
1.8	Measures have been taken to prevent knowledge of primary exposure influencing case ascertainment.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.9	Exposure status is measured in a standard, valid and reliable way.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
CONFOUNDING			
1.10	The main potential confounders are identified and taken into account in the design and analysis.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
STATISTICAL ANALYSIS			
1.11	Have confidence intervals been provided?		
SECTION 2: OVERALL ASSESSMENT OF THE STUDY			
2.1	How well was the study done to minimise the risk of bias or confounding? <i>Code ++, + or –</i>		
2.2	Taking into account clinical considerations, your evaluation of the methodology used and the statistical power of the study, are you certain that the overall effect is due to the exposure being investigated?		
2.3	Are the results of this study directly applicable to the patient group targeted by this guideline?		

SECTION 3: DESCRIPTION OF THE STUDY (Note: The following information is required for evidence tables to facilitate cross-study comparisons. Please complete all sections for which information is available)

PLEASE PRINT CLEARLY

3.1	<p>How many individuals participated in the study?</p> <p><i>List the number of cases and controls separately.</i></p>	
3.2	<p>What are the main characteristics of the study population?</p> <p><i>Include all characteristics used to identify both cases and controls – for example, age, sex, social class, disease status.</i></p>	
3.3	<p>What environmental or prognostic factor is being investigated in this study?</p>	
3.4	<p>What comparisons are made in the study?</p> <p><i>Normally only one factor will be compared, but in some cases the extent of exposure may be stratified – for example, non-smokers v. light, moderate or heavy smokers. Note all comparisons here.</i></p>	
3.5	<p>For how long are individuals followed up in the study?</p> <p><i>Length of time participant histories are tracked in the study.</i></p>	
3.6	<p>What outcome measure is used in the study?</p> <p><i>List all outcomes that are used to assess</i></p>	

	<i>the impact of the chosen environmental or prognostic factor.</i>	
3.7	<p>What size of effect is identified in the study?</p> <p><i>Effect size should be expressed as an odds ratio. If any other measures are included, note them as well. Include p values and any confidence intervals that are provided.</i></p>	
3.8	<p>How was this study funded?</p> <p><i>List all sources of funding quoted in the article, whether Government, voluntary sector or industry.</i></p>	
3.9	<p>Does this study help to answer your key question?</p> <p><i>Summarise the main conclusions of the study and indicate how it relates to the key question.</i></p>	

A.3 Notes on the use of methodology checklist: case-control studies

The studies covered by this checklist are designed to answer questions of the type ‘What are the factors that caused this event?’, and involve comparison of individuals with an outcome with other individuals from the same population who do not have the outcome. These studies start after the outcome of an event, and can be used to assess multiple causes of a single event. They are generally used to assess the causes of a new problem, but they may also be useful for the evaluation of population-based interventions such as screening.

Section 1 identifies the study and asks a series of questions aimed at establishing the internal validity of the study under review – that is, making sure that it has been carried out carefully, and that the any link between events and outcomes is clearly established. Each question covers an aspect of methodology that has been shown to make a significant difference to the conclusions of a study.

Case-control studies need to be very carefully designed, and the complexity of their design is often not appreciated by investigators, leading to many poor quality studies being conducted. The questions in this checklist are designed to identify the main features that should be present in a well-designed study. There are few criteria that should, alone and unsupported, lead to rejection of a study. However, a study that fails to address or report on more than one or two of the questions addressed below should almost certainly be rejected.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study.

- Well covered.
- Adequately addressed.
- Poorly addressed.
- Not addressed
(that is, not mentioned, or indicates that this aspect of study design was ignored).
- Not reported
(that is, mentioned, but insufficient detail to allow assessment to be made).
- Not applicable.

A.3.1 The study addresses an appropriate and clearly focused question

Unless a clear and well-defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer on the basis of its conclusions.

A.3.2 The cases and controls are taken from comparable populations

Study participants may be selected from the target population (all individuals to which the results of the study could be applied), the source population (a defined subset of the target population from which participants are selected)

or from a pool of eligible subjects (a clearly defined and counted group selected from the source population. **If the study does not include clear definitions of the source population it should be rejected.**

A.3.3 The same exclusion criteria are used for both cases and controls

All selection and exclusion criteria should be applied equally to cases and controls. Failure to do so may introduce a significant degree of bias into the results of the study.

A.3.4 What percentage of each group (cases and controls) participated in the study?

Differences between the eligible population and the participants are important, as they may influence the validity of the study. A *participation rate* can be calculated by dividing the number of study participants by the number of eligible subjects. It is more useful if calculated separately for cases and controls. If the participation rate is low, or there is a large difference between the two groups, the study results may well be invalid due to differences between participants and non-participants. In these circumstances, the study should be downgraded, and rejected if the differences are very large.

A.3.5 Comparison is made between participants and non-participants to establish their similarities or differences

Even if participation rates are comparable and acceptable, it is still possible that the participants selected to act as cases or controls may differ from other members of the source population in some significant way. A well-conducted case–control study will look at samples of the non-participants among the source population to ensure that the participants are a truly representative sample.

A.3.6 Cases are clearly defined and differentiated from controls

The method of selection of cases is of critical importance to the validity of the study. Investigators have to be certain that cases are truly cases, but must balance this with the need to ensure that the cases admitted into the study are representative of the eligible population. **The issues involved in case selection are complex, and should ideally be evaluated by someone with a good understanding of the design of case–control studies.** If the study does not comment on how cases were selected, it is probably safest to reject it as a source of evidence.

A.3.7 Is it clearly established that controls are non-cases?

Just as it is important to be sure that cases are true cases, it is important to be sure that controls do not have the outcome under investigation. Control subjects should be chosen so that information on exposure status can be obtained or assessed in a similar way to that used for the selection of cases. If the methods of control selection are not described, the study should be rejected. **If different methods of selection are used for cases and controls**

the study should be evaluated by someone with a good understanding of the design of case-control studies.

A.3.8 Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment

If there is a possibility that case ascertainment can be influenced by knowledge of exposure status, assessment of any association is likely to be biased. A well-conducted study should take this into account in the design of the study.

A.3.9 Exposure status is measured in a standard, valid and reliable way

The inclusion of evidence from other sources or previous studies that demonstrate the validity and reliability of the assessment methods used, or that the measurement method used is a recognised procedure, should further increase confidence in study quality.

A.3.10 The main potential confounders are identified and taken into account in the design and analysis

Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors is one of the principal reasons why observational studies are not more highly rated as a source of evidence. The report of the study should indicate which potential confounders have been considered, and how they have been assessed or allowed for in the analysis. Clinical judgement should be applied to consider whether all likely confounders have been considered. If the measures used to address confounding are considered inadequate, the study should be downgraded or rejected, depending on how serious the risk of confounding is considered to be. **A study that does not address the possibility of confounding should be rejected.**

A.3.11 Have confidence intervals been provided?

Confidence limits are the preferred method for indicating the precision of statistical results, and can be used to differentiate between an inconclusive study and a study that shows no effect. Studies that report a single value with no assessment of precision should be treated with caution.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

++	All or most of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought <i>very unlikely</i> to alter.
+	Some of the criteria have been fulfilled. Those criteria that have not been fulfilled or not adequately described

	are thought unlikely to alter the conclusions.
–	Few or no criteria fulfilled The conclusions of the study are <i>thought likely or very likely</i> to alter.

The code allocated here, coupled with the study type, will decide the **level of evidence** that this study provides.

The aim of the other two questions in this section is to summarise your view of the quality of this study and its applicability to the patient group targeted by the guideline you are working on.

Section 3 asks you to summarise key points about the study that will be added to an evidence table at the next stage of the process.

A.4 Methodology checklist: Qualitative studies

Study identification <i>Include author, title, reference, year of publication</i>	
Guideline topic:	Key question no:
Checklist completed by:	

Criterion	Is the criterion clearly addressed?	
Epistemology		
<p>1. Is a qualitative approach appropriate?</p> <ul style="list-style-type: none"> Does the research seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the question 	<input type="checkbox"/> Appropriate <input type="checkbox"/> Inappropriate <input type="checkbox"/> Not sure	Comments:
<p>2. Is the study clear in what it seeks to do?</p> <ul style="list-style-type: none"> Is the purpose of the research discussed – aims/objectives/research question Is there adequate reference to the literature Are underpinning values/assumptions/theory discussed 	<input type="checkbox"/> Clear <input type="checkbox"/> Unclear	Comments:

Study Design		
<p>3. How defensible is the research design?</p> <ul style="list-style-type: none"> • Is the design appropriate to the question • Are there clear accounts of the criteria used for sampling, data collection, data analysis • Is the selection of cases/sampling strategy theoretically justified • Is a rationale given for the choice of method 	<input type="checkbox"/> Defensible <input type="checkbox"/> Indefensible <input type="checkbox"/> Not sure	<p>Comments:</p>

Data collection		
<p>4. How well was the data collection carried out?</p> <ul style="list-style-type: none"> • Were the data collected in a way which addressed the research question • Was the data collection and record keeping systematic 	<input type="checkbox"/> Appropriately <input type="checkbox"/> Inappropriately <input type="checkbox"/> Not sure	<p>Comments:</p>

Validity		
<p>5. Is the role of the researcher clearly described?</p> <ul style="list-style-type: none"> • Has the relationship between the researcher and the participants been adequately considered • Is there evidence about how the research was explained and presented to the participants 	<input type="checkbox"/> Clear <input type="checkbox"/> Unclear <input type="checkbox"/> Not sure	<p>Comments:</p>
<p>6. Is the context clearly described?</p> <ul style="list-style-type: none"> • Are the characteristics of the participants and settings clearly defined • Were observations made in a sufficient variety of circumstances • Was context bias considered 	<input type="checkbox"/> Clear <input type="checkbox"/> Unclear <input type="checkbox"/> Not sure	<p>Comments:</p>
<p>7. Were the methods reliable?</p> <ul style="list-style-type: none"> • Was data collected by more than one method • Is there triangulation, or justification for not triangulating • Do the methods investigate what they claim to 	<input type="checkbox"/> Reliable <input type="checkbox"/> Unreliable <input type="checkbox"/> Not sure	<p>Comments:</p>

Analysis		
<p>8. Is the data analysis sufficiently rigorous?</p> <ul style="list-style-type: none"> • Is the procedure explicit – i.e. is it clear how the data was processed to arrive at the results • How systematic is the analysis, is the procedure reliable/dependable • Is it clear how the themes and concepts were derived from the data 	<input type="checkbox"/> Rigorous <input type="checkbox"/> Not rigorous <input type="checkbox"/> Not sure	<p>Comments:</p>
<p>9. Is the data rich?</p> <ul style="list-style-type: none"> • How well are the contexts of the data preserved • Has the diversity of perspective and content been explored • How well has the detail and depth been preserved • Are responses compared and contrasted across groups/sites 	<input type="checkbox"/> Rich <input type="checkbox"/> Poor <input type="checkbox"/> Not sure	<p>Comments:</p>
<p>10. Is the analysis reliable?</p> <ul style="list-style-type: none"> • Did more than one researcher theme and code transcripts • If so, how were differences resolved • Did participants feed back on the data if possible and relevant • Were negative/ discrepant results addressed or ignored 	<input type="checkbox"/> Reliable <input type="checkbox"/> Unreliable <input type="checkbox"/> Not sure	<p>Comments:</p>
<p>11. Are the findings credible?</p> <ul style="list-style-type: none"> • Is there a clear statement of the findings 	<input type="checkbox"/> Credible <input type="checkbox"/> Not credible	<p>Comments:</p>

<ul style="list-style-type: none"> • Are the findings internally coherent • Are elements from the original data included • Can the data sources be traced • Is the reporting clear and coherent 	<input type="checkbox"/> Not sure	
12. Are the findings relevant?	<input type="checkbox"/> Relevant <input type="checkbox"/> Irrelevant <input type="checkbox"/> Not sure	Comments:
13. Conclusions <ul style="list-style-type: none"> • How clear are the links between data, interpretation and conclusions • Are the conclusions plausible and coherent • Have alternative explanations been explored and discounted • Does this enhance understanding of the research topic • Are the implications clearly defined • Is there adequate discussion of limitations 	<input type="checkbox"/> Adequate <input type="checkbox"/> Inadequate <input type="checkbox"/> Not sure	Comments:

Ethics		
<p>14. How clear and coherent is the reporting of ethics?</p> <ul style="list-style-type: none"> • Have ethical issues been taken into consideration • Are they adequately discussed e.g. do they address consent and anonymity • Have the consequences of the research been considered i.e. raising expectations, changing behaviour etc • Was the study approved by an ethics committee 	<input type="checkbox"/> Appropriate <input type="checkbox"/> Inappropriate <input type="checkbox"/> Not sure	Comments:
Overall Assessment		
<p>Is this study relevant?</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No	Comments:
<p>How well was the study conducted? (see table below)</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> -	

++	All or most of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
+	Some of the criteria have been fulfilled. Those criteria that have not been fulfilled or not adequately described are thought unlikely to alter the conclusions.
-	Few or no criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

**A.5 Methodology checklist: Cohort studies (adapted from
Tooth et al. 2005*)**

Study identification <i>Include author, title, reference, year of publication</i>		
Guideline topic:	Key question no:	
Checklist completed by:		
1. Are the objectives or hypotheses of the study stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
2. Is the target population defined?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
3. Is the sampling frame defined?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
4. Is the study population defined?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
5. Are the study setting (venues) and/or geographic location stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
6. Are the dates between which the study was conducted stated or implicit?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
7. Are eligibility criteria stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
8. Are issues of 'selection in' to the study mentioned? **	Well covered Adequately	Not addressed

	addressed Poorly addressed	Not reported Not applicable
9. Is the numbers of participants justified?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
10. Are the numbers meeting and not meeting the eligibility criteria stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
11. For those not eligible, are the reasons why stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
12. Are the numbers of people who did/did not consent to participate stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
13. Are the reasons that people refused to consent stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
14. Were consenters compared with nonconsenters?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
15. Was the number of participants at the beginning of the study stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
16. Were the methods of data collection stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
17. Was the reliability (repeatability) of measurement methods mentioned?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

18. Was the validity (against a “gold standard”) of measurement methods mentioned?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
19. Were any confounders mentioned?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
20. Was the number of participants at each stage/wave specified?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
21. Were reasons for loss to follow-up quantified?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
22. Was the missingness of data items at each wave mentioned?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
23. Was the type of analyses conducted stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
24. Were “longitudinal” analysis methods stated?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
25. Were absolute effect sizes reported?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
26. Were relative effect sizes reported?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
27. Was loss to follow-up taken into account in the analysis?	Well covered Adequately	Not addressed Not reported

	addressed Poorly addressed	Not applicable
28. Were confounders accounted for in analyses?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
29. Were missing data accounted for in the analyses?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
30. Was the impact of biases assessed qualitatively?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
31. Was the impact of biases estimated quantitatively?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
32. Did authors related results back to a target population?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
33. Was there any other discussion of generalisability?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
34. Overall Assessment of Study. How well was the study done to minimise the risk of bias or confounding, and to establish a causal relationship between exposure and effect? Code ++, + or -		

*Criteria taken from: Tooth L, Ware R, Bain C et al. (2005) Quality of reporting of observational longitudinal research. *American Journal of Epidemiology* vol 161(3):280-288.

**Represents selection bias at the beginning of study. Other biases (i.e. loss to follow-up, missing data items) are dealt with by other checklist criteria

A.5 Notes on the use of methodology checklist: cohort studies

Criteria taken from: Tooth L et al. (2005) Quality of reporting of observational longitudinal research. *American Journal of Epidemiology* 161(3):280-288.

1. Self-explanatory
2. The group of persons toward whom inferences are directed.
Sometimes the population from which a study group is drawn
3. The list of units from which the study population will be drawn. Ideally, the sampling frame would be identical to the target population, but it is not always possible
4. The group selected for investigation
5. Comment required about location of research. Could include name of centre, town, or district
6. Self-explanatory
7. The words 'eligibility criteria' or equivalent are needed, unless the entire population is the study population
8. Any aspect of recruitment or setting that results in the selective choice of participants (e.g. gender or health status influenced recruitment)
9. Justification of number of subjects needed to detect anticipated effects.
Evidence that power calculations were considered and/or conducted
10. Quantitative statement of numbers
11. Broad mention of the major reasons
12. Quantitative statement of numbers
13. Broad mention of the major reasons
14. Quantitative comparison of the different groups
15. Total number of participants (after screening for eligibility and consent) included in the first stage of data collection
16. Descriptions of tools (e.g. Surveys, physical examinations) and processes (e.g. Face-to-face, telephone)
17. Evidence of reproducibility of the tools used
18. Evidence that the validity was examined against, or discussed in relation to, a gold standard
19. Confounders were defined as a variable that can cause or prevent the outcome of interest, is not an intermediate variable, and is associated with the factors under investigation
20. Quantitative statement of numbers at each follow-up point
21. Broad mention and quantification of the major reasons
22. Differences in numbers of data points (indicating missing data items) explained
23. Specific statistical methods mentioned by name
24. Longitudinal analyses were defined as those assessing change in outcome over two or more time points and that take into account the fact that the observations are likely to be correlated
25. Absolute effect was defined as the outcome of an exposure expressed, for example, as the difference between rates, proportions, or means, as opposed to the ratios of these measures
26. Relative effects were defined as a ratio of rates, proportions, or other measure of an effect
27. Specific mention of adjusting for, or stratifying by, loss to follow-up

28. Specific mention of adjusting for, or stratifying by, confounders
29. Specific mention of adjusting for, or stratifying by, or imputation of missing data items
30. Specific mention of bias affecting results, but magnitude not quantified
31. Specific mention of numerical magnitude of bias
32. A study is generalisable if it can produce unbiased inferences regarding a target population (beyond the subjects in the study).
Discussion could include that generalisability is not possible
33. Discussion of generalisability beyond the target population

A.6 Methodology checklist: Controlled Before and After studies (adapted from 'EPOC' by Cardiff University PHCC)

Study identification <i>Include author, title, reference, year of publication</i>	
Guideline topic:	Key question no:
Checklist completed by:	

This checklist to be used for studies where there is involvement of intervention and control groups other than by random process, and inclusion of baseline period of assessment of main outcomes. There are two minimum criteria for inclusion of CBAs in EPOC reviews:

a) Contemporaneous data collection

- Score DONE pre and post intervention periods for study and control sites are the same.
- Score NOT CLEAR if it is not clear in the paper, e.g. dates of collection are not mentioned in the text. (N.B. the paper should be discussed with the contact editor for the review before data extraction is undertaken).
- Score NOT DONE if data collection was not conducted contemporaneously during pre and post intervention periods for study and control sites.

b) Appropriate choice of control site

- Studies using second site as controls:
- Score DONE if study and control sites are comparable with respect to dominant reimbursement system, level of care, setting of care and academic status.
- Score NOT CLEAR if not clear from paper whether study and control sites are comparable. (N.B. the paper should be discussed with the contact editor for the review before data extraction is undertaken).
- Score NOT DONE if study and control sites are not comparable.

Quality criteria for controlled before and after (CBA) designs

Seven standard criteria are used for CBAs included in EPOC reviews:

a) **Baseline measurement**

- Score DONE if performance or patient outcomes were measured prior to the intervention, and no substantial differences were present across study groups (e.g. where multiple pre intervention measures describe similar trends in intervention and control groups);
- Score NOT CLEAR if baseline measures are not reported, or if it is unclear whether baseline measures are substantially different across study groups;
- Score NOT DONE if there are differences at baseline in main outcome measures likely to undermine the post intervention differences (e.g. are differences between the groups before the intervention similar to those found post intervention).

b) **Characteristics for studies using second site as control**

- Score DONE if characteristics of study and control providers are reported and similar;
- Score NOT CLEAR if it is not clear in the paper e.g. characteristics are mentioned in the text but no data are presented;
- Score NOT DONE if there is no report of characteristics either in the text or a table OR if baseline characteristics are reported and there are differences between study and control providers.

c) **Blinded assessment of primary outcome(s)* (protection against detection bias)**

- Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly OR the outcome variables are objective e.g. length of hospital stay, drug levels as assessed by a standardised test;
- Score NOT CLEAR if not specified in the paper;
- Score NOT DONE if the outcomes were not assessed blindly.

** Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately and label each outcome variable clearly.*

d) Protection against contamination

- Studies using second site as control
- Score DONE if allocation was by community, institution, or practice and is unlikely that the control group received the intervention;
- Score NOT CLEAR if providers were allocated within a clinic or practice and communication between experimental and group providers was likely to occur;
- Score NOT DONE if it is likely that the control group received the intervention (e.g. cross-over studies or if individuals rather than providers were randomised).

e) Reliable primary outcome measure(s)

- Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8 OR the outcome is obtained from some automated system e.g. length of hospital stay, drug levels as assessed by a standardised test;
- Score NOT CLEAR if reliability is not reported for outcome measures that are obtained by chart extraction or collected by an individual;
- Score NOT DONE if agreement is less than 90% or kappa is less than 0.8.

** In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately and label each outcome variable clearly.*

f) Follow-up of professionals (protection against exclusion bias)

- Score DONE if outcome measures obtained 80-100% subjects allocated to groups. (Do not assume 100% follow-up unless stated explicitly.);
- Score NOT CLEAR if not specified in the paper;
- Score NOT DONE if outcome measures obtained for less than 80% of individuals allocated to groups.

g) Follow-up of individuals

- Score DONE if outcome measures obtained 80-100% of individuals allocated to groups or for individuals who entered the study. (Do not assume 100% follow-up unless stated explicitly.);
- Score NOT CLEAR if not specified in the paper;

- Score NOT DONE if outcome measures obtained for less than 80% of individuals allocated to groups or for less than 80% of individuals who entered the study.

OVERALL ASSESSMENT OF STUDY	
How well was the study conducted? Code ++, + or –	
Are the results of the study directly applicable to the groups targeted by the guideline?	

A.7 Methodology checklist: Interrupted time series (EPOC version)

Study identification <i>Include author, title, reference, year of publication</i>	
Guideline topic:	Key question no:
Checklist completed by:	

This checklist to be used for ITS studies where there is a change in trend attributable to the intervention. There are two minimum criteria for inclusion of ITS designs in EPOC reviews:

- a) **Clearly defined point in time when the intervention occurred.**
 - Score DONE if reported that intervention occurred at a clearly defined point in time.
 - Score NOT CLEAR if not reported in the paper (will be treated as NOT DONE if information cannot be obtained from the authors).
 - Score NOT DONE if reported that intervention did not occur at a clearly defined point in time.
- b) **At least three data points before and three after the intervention.**
 - Score DONE if 3 or more data points before and 3 or more data points recorded after the intervention.
 - Score NOT CLEAR if not specified in paper e.g. number of discrete data points not mentioned in text or tables (will be treated as NOT DONE if information cannot be obtained from the authors).
 - Score NOT DONE if less than 3 data points recorded before and 3 data points recorded after intervention.

Quality criteria for interrupted time series (ITSs)

The following seven standard criteria should be used to assess the methodology quality of ITS designs included in EPOC reviews. Each criterion is scored DONE, NOT CLEAR or NOT DONE. The results of the quality assessment for each study are reported in the Table of Included Studies in RevMan. Examples can be obtained from the EPOC review group coordinator.

- a) Protection against secular changes**
The intervention is independent of other changes.
- Score DONE if the intervention occurred independently of other changes over time;
 - Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
 - Score NOT DONE if reported that intervention was not independent of other changes in time.
- b) Data were analysed appropriately**
- Score DONE if ARIMA models were used **OR** time series regression models were used to analyse the data and serial correlation was adjusted/tested for;
 - Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
 - Score NOT DONE if **it is clear that neither** of the conditions above not met.
- c) Reason for the number of points pre and post intervention given**
- Score DONE if rationale for the number of points stated (eg monthly data for 12 months post-intervention was used because the anticipated effect was expected to decay) **OR** sample size calculation performed;
 - Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
 - Score NOT DONE if **it is clear that neither** of the conditions above met.
- d) Shape of the intervention effect was specified**
- Score DONE if a rational explanation for the shape of intervention effect was given by the author(s);
 - Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
 - Score NOT DONE if **it is clear that** the condition above is not met
- e) Protection against detection bias**
Intervention unlikely to affect data collection
- Score DONE if reported that intervention itself was unlikely to affect data collection (for example, sources and methods of data collection were the same before and after the intervention);
 - Score NOT CLEAR if not reported (will be treated as NOT DONE if information cannot be obtained from the authors);
 - Score NOT DONE if the intervention itself was likely to affect data collection (for example, any change in source or method of data collection reported).
- Blinded assessment of primary outcome(s)*
- Score DONE if the authors state explicitly that the primary outcome variables were assessed blindly **OR** the outcome variables are

objective e.g. length of hospital stay, drug levels as assessed by a standardised test;

- Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
- Score NOT DONE if the outcomes were not assessed blindly.

** Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately and label each outcome variable clearly.*

c) Completeness of data set

- Score DONE if data set covers 80-100% of total number of participants or episodes of care in the study;
- Score NOT CLEAR if not specified (will be treated as NOT DONE if information cannot be obtained from the authors);
- Score NOT DONE if data set covers less than 80% of the total number of participants or episodes of care in the study.

d) Reliable primary outcome measure(s)*

- Score DONE if two or more raters with at least 90% agreement or kappa greater than or equal to 0.8 OR the outcome is obtained from some automated system e.g. length of hospital stay, drug levels as assessed by a standardised test;
- Score NOT CLEAR if reliability is not reported for outcome measures that are obtained by chart extraction or collected by an individual (will be treated as NOT DONE if information cannot be obtained from the authors);
- Score NOT DONE if agreement is less than 90% or kappa is less than 0.8.

** In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately.*

e) Consumer involvement

Were consumers (i.e. potential patients) involved at any point of the design, conduct or interpretation of the study? (E.g., consumers involved in clinical practice guideline development, or their views collected.)

- Score DONE if specified in paper, and give details;
- Score NOT CLEAR if not reported;
- Score NOT DONE if consumers explicitly not involved.

OVERALL ASSESSMENT OF STUDY	
How well was the study conducted? Code ++, + or –	

Are the results of the study directly applicable to the groups targeted by the guideline?	
---	--

A.8 Methodology checklist: economic evaluations

The criteria used in this checklist are extracted from Drummond MF et al. (1997) Critical assessment of economic evaluation. In: 'Methods for the economic evaluation of health care programmes'. 2nd edition. Oxford: Oxford Medical Publications. Explanatory notes on these criteria are available from the publication.

Study identification <i>Include author, title, reference, year of publication</i>		
Checklist completed by:		
Evaluation criterion		Comments
1.	Was a well-defined question posed in answerable form?	
1.1	Did the study examine both costs and effects of the service(s) or programme(s)?	
1.2	Did the study involve a comparison of alternatives?	
1.3	Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context?	
2.	Was a comprehensive description of the competing alternatives given (that is, can you tell who? did what? to whom? where? and how often?)?	

2.1	Were any important alternatives omitted?	
2.2	Was (Should) a do-nothing alternative (be) considered?	
3.	Was the effectiveness of the programmes or services established?	
3.1	Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?	
3.2	Was effectiveness established through an overview of clinical studies?	
3.3	Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	
4.	Were all the important and relevant costs and consequences for each alternative identified?	
4.1	Was the range wide enough for the research question at hand?	
4.2	Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of individuals and third-	

	party payers.)	
4.3	Were capital costs, as well as operating costs, included?	
5.	Were costs and consequences measured accurately in appropriate physical units (for example, hours of nursing time, number of physician visits, lost work-days, gained life-years)?	
5.1	Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?	
5.2	Were there any special circumstances (for example, joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?	
6.	Were costs and consequences valued credibly?	
6.1	Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers' views and health	

	professionals' judgements.)	
6.2	Were market values employed for changes involving resources gained or depleted?	
6.3	Where market values were absent (for example, volunteer labour), or did not reflect actual values (for example, clinic space donated at reduced rate), were adjustments made to approximate market values?	
6.4	Was the valuation of consequences appropriate for the question posed (that is, has the appropriate type or types of analysis – cost-effectiveness, cost-benefit, cost-utility – been selected)?	
7.	Were costs and consequences adjusted for differential timing?	
7.1	Were costs and consequences which occur in the future 'discounted' to their present values?	
7.2	Was any justification given for the discount rate used?	
8.	Was an incremental analysis of costs and consequences of alternatives performed?	

8.1	Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits or utilities generated?	
9.	Was allowance made for uncertainty in the estimates of costs and consequences?	
9.1	If data on costs or consequences were stochastic, were appropriate statistical analyses performed?	
9.2	Were study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)?	
10.	Did the presentation and discussion of study results include all issues of concern to users?	
10.1	Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (for example, cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion?	
10.2	Were the results compared with those of others who have	

	investigated the same question? If so, were allowances made for potential differences in study methodology?	
10.3	Did the study discuss the generalisability of the results to other settings and patient/client groups?	
10.4	Did the study allude to, or take account of, other important factors in the choice or decision under consideration (for example, distribution of costs and consequences, or relevant ethical issues)?	
10.5	Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?	
OVERALL ASSESSMENT OF THE STUDY		
How well was the study conducted? <i>Code ++, + or –</i>		
Are the results of this study directly applicable to the patient group targeted by this guideline?		

Appendix B

Guidance scope template

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

PUBLIC HEALTH PROGRAMME/INTERVENTION GUIDANCE

SCOPE/DRAFT SCOPE

[This should be 8 pages maximum for an intervention and 12 pages maximum for a programme. It needs to be written in plain English and should be snappy and to the point: our stakeholders are busy people.]

Note: references should be included as footnotes, in alphabetical order. (If there are more than three authors or editors, use 'et al.' (not in italics). Names of authors or editors should be separated by commas and given in the form 'Surname Initials', with no punctuation between initials or between surname and initial. (For example, 'Smith S, Jones TD', not 'Smith, S., Jones, T.D.'))]

1 Guidance title

[Full title of guidance]

1.1 Short title

[Short title of guidance – NB: must be derived from the full title]

2 Background

(a) The National Institute for Health and Clinical Excellence ('NICE' or 'the Institute') has been asked by the DH to develop guidance on a public health programme/intervention aimed at [preventing/promoting - add text according to topic].

(b) NICE public health programme/intervention guidance supports implementation of the preventive aspects of NSFs where a framework has been published. The statements in each NSF reflect the evidence that was used at the time the Framework was prepared. The public health guidance

published by the Institute after an NSF has been issued will have the effect of updating the Framework. Add if necessary: Specifically, in this case, the guidance will support the following NSFs and other government policy documents:

(c) This guidance will provide recommendations for good practice, based on the best available evidence of effectiveness, including cost effectiveness. It is aimed at professionals with public health as part of their remit working within the NHS, local authorities and the wider public, private, voluntary and community sectors.

3 The need for guidance

[Single para of unnumbered text; style = NICE normal – for more than one para, label a), b), c) etc. Provide 6 points maximum (for programmes and interventions). Each point should comprise no more than 2-4 sentences each (of no more than 30 words per sentence). These points should cover:

- statistics related to the prevalence of disease/risky behaviour
- the consequences in terms of ill health
- costs to the NHS/society
- who is most affected within the general population.

Other factors to cover if particularly relevant and/or the maximum limit won't be exceeded include:

- barriers to prevention
- how England compares with other countries.]

4 The guidance

a) Public health guidance will be developed according to NICE processes and methods. For details see Section 5.

b) This document is the scope. It defines exactly what this guidance will (and will not) examine, and what the guidance developers will consider. The scope is based on a referral from the Department of Health (see Appendix A).

4.1 Populations

4.1.1 Groups that will be covered

[input as appropriate]

4.1.2 Groups that will not be covered

[input as appropriate]

4.2 Areas [interventions and any other activities]

4.2.1 Areas that will be covered

The interventions/activities to be considered by this guidance are: for each intervention/activity include aim, content, delivery mode, target audience, setting, duration, outcome measures.[Single para of unnumbered text; style = NICE normal – for more than one para, label a), b), c) etc]

4.2.2 Areas that will not be covered

[Input as appropriate]

4.3 Comparators

[This section is optional for programmes but COMPULSORY for interventions (it refers to the fact that a new approach/intervention will – or might be – compared with current practice). Use the following text:]

Interventions/approaches will be examined, where possible, against relevant comparators.

4.4 Outcomes

[Briefly outline the outcome measures to be considered as evidence of the effectiveness of an intervention or approach.]

4.5 Key questions

The following questions will be addressed: [List questions in a bullet point list, bearing in mind that no sentence should be more than 15-20 words long.]

4.6 Target audiences and settings

The guidance will be aimed at professionals working in the NHS, in other public sector organisations, the private sector and in the voluntary and community sectors who have either a direct or indirect role in and/or responsibility for [input as appropriate]

4.7 Status of this document

[Draft] This is the draft scope, released for consultation on date, to be discussed at a stakeholder meeting on date. Following consultation, the final version of the scope will be available at the NICE website in month.

[Final] This is the final scope, incorporating comments from a xxx week consultation which included a stakeholder meeting on date.

5 Further information

The public health guidance development process is described in detail in title of process and methods manuals/prior to sign-off The Operating Model for the Centre of Public Health Excellence, available at:
www.nice.org.uk/page.aspx?o=248187

6 NICE related guidance

Appendix A Referral from the Department of Health

The Department of Health asked the Institute to insert [DH wording].

Appendix C

NICE review format

Title of review

Executive Summary

Narrative summary of the evidence reviewed and how/whether it answers the research questions.

Includes all of the evidence statements and the linked references.

A list of all included references is incorporated into the Exec Summary.

Contents/structure for main report

1. Background

2. Methodology

- 2.1 Literature Search
- 2.2 Selection of Studies for Inclusion
- 2.3 Quality Appraisal
- 2.4 Study categorisation
- 2.5 Assessing applicability
- 2.6 Synthesis

3. Summary of Findings

- 3.1 Overall summary of studies identified (including numbers, types, quality, applicability)

Address each key question (and sub question) set out in the Scope.

- 3.2 Key question (1) (for example, effectiveness of intervention x)
 - 3.2.1 Narrative summary (including about applicability to UK)
 - 3.2.2 Evidence statement(s)
 - 3.2.3 sub questions, narrative summaries and evidence statements

- 3.3 Key question (3) (for example, effectiveness of intervention y)
 - 3.3.1 Narrative summary (including about applicability to UK)
 - 3.3.2 Evidence statement(s)
 - 3.3.3 sub questions, narrative summaries and evidence statements
- 3.4 Key question (3) (for example, about acceptability/implementability)
 - 3.4.1 Narrative summary
 - 3.4.2 Evidence statement(s)
 - 3.4.3 sub questions, narrative summaries and evidence statements

4. Evidence tables

- A. Efficacy studies (separate tables for each key question/ intervention/ outcome)
- B. Other evidence (for example, about implementation, barriers)

5. Meta-analyses (if applicable)

[based on Evidence table A which includes effect size]

Section 3. Summary of Findings

3.1 Key question (1) – Intervention(s), population(s), setting(s), outcome(s)

See Chapter 4.4 for further details.

For each key question, intervention and outcome, there should be an **evidence statement** about:

- the **strength** of evidence (reflecting appropriateness, quality and quantity of evidence)
- **applicability** to the question/target population

For example:

- a body of **1++ evidence of efficacy** with consistent findings about the impact of intervention X on outcome Y
- a body of ++ evidence **directly applicable** to the target audience in terms of ethnicity, age, gender, etc

Ideally, the summary and the evidence statement itself will also include an assessment of the implementability of the intervention and its typical effect size.

3.1.1 Narrative summary

Evidence of efficacy

[Results across a group of related studies are assessed narratively, using text and tables; quantitative methods (for example, meta-analyses) are also performed where possible and useful. See chapter 4.4 for further details. In addition, strive to make a statement about a typical size of effect (see D below)]

For example, *evidence from eleven randomised controlled trials (RCTs) and one controlled non randomised trial suggest that xxx programmes can result in short-term xxx outcome.*

Increase in xxx relative to control were noted in seven of the RCTs ([insert refs]) and one CCT ([insert ref]). There were a wide range of results xxxxx. For instance, xxx xxx found that xxx etc. A non randomised UK trial with a weak study design ([ref]) found that those in the advice group increased intermediate outcome xxxxxx.

A recent UK-based individual RCT with five month follow up ([ref]) found no significant difference in xxx. Another UK-based RCT ([ref]) looking at xxx found that, although there was no difference in xxxxxxx

One RCT among male blue collar workers in xxx ([ref]) resulted in an increase in xxx in the intervention compared to the control group despite xxx.

Effect size: Narrative summary and/or meta-analysis of effect size/impact:

- population outcome
- inequalities impact

Other effects (including harms): Narrative summary...

Overall summary of strength (quality and quantity) of evidence of efficacy:

- ***‘A body of level 1+ evidence of efficacy...’***

Applicability (of evidence from efficacy studies) to UK population/setting

[Narrative assessing applicability of the evidence from efficacy studies to each of the populations/settings identified in the scope]

- ***‘...directly applicable to the populations/settings...’***

[NB It may be useful to use one of following 4 conclusions to describe applicability of the body of efficacy data:

1. Likely to be applicable across broad range...
2. Likely applicable across broad range..., assuming appropriately adapted
3. Applicable only to populations or settings included in the studies, and broader applicability is uncertain
4. Applicable only to settings or populations included in the studies]

3.1.2 Evidence statement(s)

A body of level 1+ evidence of efficacy directly applicable to [the populations/settings in question] showing typical effect size/impact.

3.1.3 sub-questions

Narrative summaries and evidence statements.

3.1.4 Other evidence

Implementability of intervention (corroborative evidence)

[Narrative summary of evidence, from the efficacy studies and elsewhere, which allows an assessment of whether the intervention would be feasible, acceptable, etc. It would include an assessment of **barriers**. This could draw on published studies and grey literature, as appropriate, to answer the research questions. Evidence from published studies should be presented in a separate evidence table from the efficacy studies, although it may refer to some of the studies that have been included in the efficacy evidence table]

[Crucially, evidence to support implementation will also come from the fieldwork data, presented to PHIAC or the PDG in a technical report.

NOTE: PHIAC/ the PDG draft **recommendations**, including their priority, will be based on:

- The **evidence statement**, based on the strength of evidence of efficacy (quality and quantity) and its applicability to the populations/settings in question
- The strength of evidence about implementation
- Effect size including impact on inequalities
- Cost effectiveness (NHS and public sector)
- Other effects, including potential harm

4. Evidence tables

Evidence tables should be completed for each key question and outcome. They should be sub-divided and/or annotated to demonstrate where outcomes are self-reported versus those that are objectively validated.

A. Efficacy/effectiveness studies

List study types in the following order: Systematic review(s); randomised trial(s); controlled non randomised trials (CCTs), controlled before & after (CBA), interrupted time series (ITS); other study(ies)

First author and date	Study design	Research Type (ie category 1,2,3,4)	Research Quality (see methodology checklist)	Study Population	Research Question & Design (include power calculation if available)	Length of follow-up	Main results (include effect size(s) and CIs for each outcome if available; report whether significant)	Applicability to the UK populations and settings	Confounders (potential sources of bias – see methodology checklist and relate to research quality assessment)/ Comments
Evidence of Efficacy (Internal Validity)									
	RCT Individual	1	+						
	CCT	2	++						

APPENDIX A – Included studies

APPENDIX B – Excluded studies (including reasons for exclusion)

APPENDIX C – Data extraction forms: included studies

Appendix D

Data extraction form

Data Extraction Form	Ref ID [this will be the reference manager ID number assigned after the literature search]
Authors/ Title/ Source [Insert reference manager citation from literature search]	
Project:	
Data extracted by:	Date of extraction:
Describe the study:	
Systematic review (including at least one RCT)	<input type="checkbox"/>
Systematic review of experimental studies	<input type="checkbox"/>
Systematic review of observational studies	<input type="checkbox"/>
Randomised controlled trial: Individual	<input type="checkbox"/>
Randomised controlled trial: Cluster	<input type="checkbox"/>
Controlled non-randomised trial	<input type="checkbox"/>

Controlled before-and-after		<input type="checkbox"/>
Interrupted time series	<input type="checkbox"/>	
Before and after study	<input type="checkbox"/>	
Cross sectional (survey)	<input type="checkbox"/>	
Audit/Evaluation	<input type="checkbox"/>	
Economic analysis	<input type="checkbox"/>	
Case study	<input type="checkbox"/>	
Local practice report		<input type="checkbox"/>
Qualitative study	<input type="checkbox"/>	
Focus group(s)	<input type="checkbox"/>	
Brief interview	<input type="checkbox"/>	
Extended interview	<input type="checkbox"/>	
Semi-structured interview	<input type="checkbox"/>	
Document Analysis	<input type="checkbox"/>	
Observation (Passive/Participant)		<input type="checkbox"/>
Other (please state)	<input type="checkbox"/>	
What was the research question?		
Review parameters (if applicable):		
Describe the search method:		
Databases/sources searched:		
Years searched:		

Study selection criteria:

Inclusion:

Exclusion:

Number of studies and participants included:

What data was extracted?

How was the data synthesised?

Was there heterogeneity across studies?

Describe the method of analysis (meta-analysis/narrative synthesis etc):

Other study parameters:

Setting:

Geographical (City/country):

Social (school/workplace etc):

Date of study (to/from):

Resources (people/money/organisations etc):

Participants:

Number of participants/organisations etc enrolled:

Socio-economic data (if presented):

Were intervention groups balanced at baseline?:

Comments:

Unit of allocation/recruitment:

Individual

Sex (%):

Age (range or mean):

Group

Describe:

Organisation/institution

Describe:

Community/environment

Describe:

Policy/socio-political

Describe:

Method of recruitment/enrolment and response rate:

Method of allocation to intervention:

Was allocation concealed?

Yes

No

Not clear

Selection criteria:

Inclusion:

Exclusion:

Intervention:

Description of the Intervention:

Description of the comparator(s):

Method/mode of delivery (for example, peer education):

Providers/deliverers of the intervention:

Length, duration and intensity of the intervention:

Time to follow-up (average/median):

How many (n, %) participants completed the intervention?

For non-completers, were the reasons for non-completion described?

Outcomes³:

Health promotion outcomes:

Health literacy

Social action and influence

Healthy policy changes

Other

Describe:

Were baseline measurements of outcomes assessed? Yes No

Intermediate outcomes:

Lifestyle changes

³ Adapted from Nutbeam's model (1998).

Social action and influence

Healthy policy changes

Other

Describe:

Were baseline measurements of outcomes assessed? Yes No

Health and Social outcome measure(s):

Mortality

Morbidity

Quality of life

Other

Describe:

Were baseline measurements of outcomes assessed? Yes No

Were the outcome measure(s) validated? Yes No Not clear

If yes, how?

Analyses:

Data collection methods used:

Describe methods used (intention to treat, descriptive statistics, qualitative analysis etc):

Unit of analysis:

Individual Group

Organisation/institution

Community/environment

Policy/socio-political

Other (describe)

Power

Was a power calculation presented?

Yes

No

If yes, describe:

Was the study powered to detect an effect if one exists?

Yes

No

Not clear

Any other process details:

Results:

Briefly describe the results for each of the main outcomes, paying particular attention to issues relating to health inequalities and cost effectiveness:

Are there any key criticisms of the conclusions drawn by the author's?

Does the paper address or offer any evidence of effect in the following groups? If so, please ensure that evidence is presented in results above.

Children and young people	Yes <input type="checkbox"/>	No <input type="checkbox"/>	Not clear <input type="checkbox"/>
Older people	Yes <input type="checkbox"/>	No <input type="checkbox"/>	Not clear <input type="checkbox"/>
Gender	Yes <input type="checkbox"/>	No <input type="checkbox"/>	Not clear <input type="checkbox"/>
Black and minority ethnic groups	Yes <input type="checkbox"/>	No <input type="checkbox"/>	Not clear <input type="checkbox"/>
Lower socio-economic status	Yes <input type="checkbox"/>	No <input type="checkbox"/>	Not clear <input type="checkbox"/>

Other (please specify):

Does the paper demonstrate any evidence of harms or adverse effects associated with the intervention?

Do the authors identify any strengths and/or weaknesses of the evidence presented?

In your opinion, are the results generalisable to the UK?

Yes No Not clear

Why:

Do the authors identify any evidence gaps or make any recommendations for further research?

Is there any data on cost-effectiveness presented?
Are there policy implications of the work?
Are there effective practice implications of the work?
Pass to other reviewer for second opinion?
Comment:

Appendix E

Glossary

CBA	controlled before and after
CPHE	Centre for Public Health Excellence
DH	Department of Health
HDA	Health Development Agency
IS	Information Services
ITS	interrupted time series
NCC	National Collaborating Centre
NHS	National Health Service
NICE	The National Institute for Health and Clinical Excellence
NSF	national service frameworks
PDF	portable document format
PHCC	Public Health Collaborating Centre
PHIAC	Public Health Interventions Advisory Committee
PDG	Programme Development Group
QALY	Quality-adjusted life year
RCT	randomised controlled trial