



*National Institute for
Health and Clinical Excellence*

Issue date: April 2009

**Methods for the
development of NICE
public health guidance
(second edition)**

About this document

This document describes the methods used in the development of NICE public health guidance.

The document replaces 'Methods for the development of NICE public health guidance' (published March 2006).

The document is available from the NICE website (www.nice.org.uk). A related document on processes is also available, 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public (second edition)' (NICE 2009).

Nothing in this document shall restrict any disclosure of information by NICE 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

ISBN: 1-84629-986-1

© Copyright National Institute for Health and Clinical Excellence, 2006 (updated 2009). 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 allowed without the express written permission of the Institute.

Published by the National Institute for Health and Clinical Excellence
First issued March 2006 (updated April 2009)

Contents

1	Introduction	7
1.1	NICE guidance	7
1.2	Equality and social value judgements	7
1.3	Who is this manual for?	8
1.4	Framework for public health guidance	9
1.5	Types of knowledge and evidence	13
1.6	Stakeholders	15
1.7	Quality assurance principles	15
1.8	References and further reading	16
<hr/>		
2	Topic selection and scoping the guidance	19
2.1	Introduction	19
2.2	Briefing paper for the Public Health Topic Selection Consideration Panel	20
2.3	Purpose of the scope	26
2.4	Drafting the scope	27
2.5	Consulting on the scope	30
2.6	Further reading	30
<hr/>		
3	Determining the evidence for review and consideration	32
3.1	Introduction	32
3.2	Types of evidence	32
3.3	Types of review and types of research question	34
3.4	Planning reviews	44
3.5	Colloquial evidence	46
3.6	Equality and diversity	47
3.7	References and further reading	48
<hr/>		

4	Identifying the evidence	49
4.1	Introduction	49
4.2	Phase one: constructing the search protocol	51
4.3	Phase two: developing the search strategy	52
4.4	Phase three: gathering the evidence, conducting searches and documenting the process	53
4.5	Equality and diversity	57
4.6	References	57
5	Reviewing the scientific evidence	58
5.1	Introduction	58
5.2	Selecting relevant evidence	59
5.3	Assessing the quality of the evidence	60
5.4	Extracting, synthesising and presenting the evidence	63
5.5	Deriving evidence statements	70
5.6	Assessing applicability	74
5.7	Published guidance	75
5.8	Equality and diversity	76
5.9	References and further reading	77
6	Incorporating health economics	80
6.1	Introduction	80
6.2	Reviewing economic evaluations	81
6.3	Prioritising areas for further economic analysis	84
6.4	Economic evidence and guidance recommendations	90
6.5	References and further reading	92
7	Developing recommendations	93
7.1	Introduction	93
7.2	Considering the evidence and other issues	94
7.3	Format and wording of recommendations	100
7.4	Revisions following fieldwork and stakeholder consultation	105
7.5	Considerations	106

7.6	Equality and diversity	107
7.7	Prioritising recommendations for implementation	108
7.8	Formulating research recommendations	108
7.9	Further reading	112
<hr/>		
8	Testing the recommendations: fieldwork	113
8.1	Introduction	113
8.2	Aim of fieldwork	113
8.3	Guiding principles	114
8.4	Fieldwork commissioning	115
8.5	Approaches to fieldwork	115
8.6	Ethical approval	117
8.7	Quality standards for fieldwork methods	117
8.8	Fieldwork analysis	124
8.9	Fieldwork report	124
8.10	Using fieldwork findings to finalise recommendations	125
8.11	Equality and diversity	125
8.12	References and further reading	126
<hr/>		
	Appendix A Conceptual framework for the work of the Centre for Public Health Excellence (CPHE)	127
	Appendix B Electronic resources	151
	Appendix C Example of audit information to accompany search strategies	153
	Appendix D Glossary of study designs	154
	Appendix E Algorithm for classifying quantitative (experimental and observational) study designs	164
	Appendix F Quality appraisal checklist – quantitative intervention studies	165
	Appendix G Quality appraisal checklist – quantitative studies reporting correlations and associations	186

Appendix H Quality appraisal checklist – qualitative studies	193
Appendix I Quality appraisal checklist – economic evaluations	204
Appendix J Process for using review-level material in exceptional circumstances	218
Appendix K Evidence tables	224
Appendix L NICE review format	230

1 Introduction

The National Institute for Health and Clinical Excellence (NICE) is the independent organisation responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health.

1.1 NICE guidance

NICE produces guidance in four areas:

- public health – guidance for the promotion and protection of good health and the prevention of disease
- health technologies – guidance on the use of new and existing health technologies (including drugs, medical devices, diagnostic techniques and surgical procedures)
- interventional procedures – guidance on the efficacy and safety of surgical, endoscopic and endovascular procedures and allied techniques
- clinical practice – guidance on the appropriate treatment and care of people with specific diseases and conditions.

All types of NICE guidance are developed using the best available evidence and involving stakeholders in a transparent and collaborative manner. Stakeholders include national organisations that represent the public, patients and carers, practitioners, community interests and companies that have an interest in the guidance in development.

1.2 Equality and social value judgements

NICE is committed to promoting equality, eliminating unlawful discrimination and considering the implications of its guidance for human rights. It aims to comply fully with all legal obligations to:

- promote race and disability equality and equality of opportunity between men and women
- eliminate unlawful discrimination on grounds of race, disability, age, sex, sexual orientation and religion or belief in the way it carries out its functions and in its employment policies and practices.

NICE's equality scheme sets out how it is meeting these obligations on equality and discrimination and what it still needs to do¹. In line with NICE's equalities scheme and action plan, this manual includes explicit consideration of how the development of guidance will consider equalities issues at the scoping, development, and validation stages, and how an audit trail of this activity will be maintained.

All NICE guidance takes account of the principles set out in 'Social value judgements: principles for the development of NICE guidance. Second edition' (2008)².

1.3 Who is this manual for?

This manual describes the methods used by the Centre for Public Health Excellence (CPHE) in NICE to develop public health guidance. It is aimed at: staff within CPHE, review teams working in NICE's public health collaborating centres or based in other agencies, other contractors engaged in the guidance production process (for example, those conducting fieldwork), and members of the Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Groups (PDGs) that develop public health guidance. However, it is also likely to be useful and of interest to stakeholders.

This second edition has been prepared by staff within CPHE, drawing on experience of the first 4 years of public health guidance development at NICE. It is based on international criteria of quality for guidance development, as articulated by the AGREE instrument³.

The structure of this manual follows the **methods** for development of NICE public health guidance from inception to publication. An overview of the **processes** for public health guidance development is described in 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public (second edition)' (NICE 2009).

The manuals define the philosophical and methodological principles which govern the production of guidance. They also set out the operational framework which collaborating centres, contractors, advisory committees and the NICE CPHE team will follow. The manuals are prescriptive to ensure the:

- aims and objectives of guidance production are as clear as possible
- procedures used are as systematic, transparent, auditable and accountable as possible.

¹ The equality scheme and action plan for implementation (approved by NICE's Board in March 2007) are available at www.nice.org.uk/aboutnice/howwework/NICEEqualityScheme.jsp

² www.nice.org.uk/aboutnice/howwework/socialvaluejudgements/socialvaluejudgements.jsp

³ www.agreetrust.org

1.4 Framework for public health guidance

NICE's methods and processes for developing public health guidance are based on a clear set of values and principles. The range of activities and topics covered is inclusive. It is based on a conceptual framework (Kelly et al. 2009) for public health (see appendix A). This comprises four vectors – population, environment, society and organisations – linked to human behaviour. They explain the patterns of potentially preventable diseases. Social differences in the population are linked to patterns of mortality and morbidity. These differences manifest themselves in a number of key areas including:

- conditions and diseases such as cardiovascular disease, cancer, obesity, diabetes, vaccine-preventable infections
- behaviours such as smoking, drug or alcohol misuse, sexual activity, physical activity
- other factors affecting health such as environment, work, housing and transport
- accidents and injuries
- child and maternal health, mental health, oral health.

In any of these areas, public health activities may be direct (for example, providing contraceptive services 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 ensuring the health of pregnant and breastfeeding women) and other issues associated with the wider determinants of health are all covered.

In summary, NICE public health guidance considers a variety of approaches, from health promotion and public education campaigns, through uptake of immunisation and screening to activities such as community development.

1.4.1 Determinants of health

NICE public health guidance recognises the wide spectrum of determinants of health. The determinants work through the four vectors mentioned above – population, environment, society and organisations (see appendix A). Human behaviour and the vectors interact via causal pathways to determine the health of individuals and populations. Patterns of illness can occur in whole populations or subpopulations. Both illnesses and the resulting patterns have causes. The task of developing guidance involves describing how health is affected and finding preventive mechanisms, both primary and secondary. The conceptual framework is used to define the scope of a topic and to articulate the mechanisms of cause and intervention.

Public health guidance may be aimed at population, community, organisational, group, family or individual level, as appropriate.

1.4.2 Equality and equity

1.4.2.1 Health difference and health inequality

Health difference is a simple difference in health state or status, however measured, between individuals or groups such as social classes, socioeconomic groups, men and women, ethnic groups or geographical communities. Health inequalities are not exclusively biological in origin but are also the consequence of human activity. Where inequalities arise as a consequence of human actions, they can be changed if the causes are changed.

1.4.2.2 Health inequity

A health inequity is an unnecessary, avoidable, unfair and unjust difference in someone's health or healthcare. 'Health inequity' should not be used interchangeably with the term 'health inequality' because the differences in health or healthcare that people experience are not necessarily unfair or unjust. Health inequity is concerned with social justice, values or politics, while inequalities in health are a matter of fact.

Health inequities, like health inequalities, can be eradicated or reduced because they are products of human action. However, addressing them can have considerable political implications because of the value judgement involved: not all people will judge the same health difference to be unfair.

1.4.2.3 Equality legislation

NICE public health guidance is developed within the legal framework relating to equalities. Addressing health inequalities has been a fundamental element of public health practice since its inception as a branch of medicine in the nineteenth century. Considering evidence on inequalities has been part of the work of CPHE since its foundation, and recommendations and research recommendations have been developed to attempt to redress these issues.

The Equality Act (2006) imposes a duty on public bodies to consider the effect of their activities in relation to equality. NICE considers issues of equality in its guidance production processes, so that unlawful discrimination does not arise from NICE guidance and so that opportunities for promoting equality can be identified. The legislation considers six areas of equality:

- age
- gender

- race
- disability
- sexual orientation
- ethnicity.

Public health guidance also takes account of socioeconomic differences.

Two legal concepts are important when considering equality: relevance and proportionality. Relevance assesses how much an issue affects equality and proportionality assesses an appropriate outcome. The weight given to equality in a function should be proportionate to its relevance for that function.

The outcome of an intervention can affect health equity in different ways:

- An intervention might improve the health of people in different groups to the same degree, so that any differences in health between those groups will remain after the intervention.
- An intervention may be more effective in one group than in another. If it is more effective in the more disadvantaged group, the net outcome will be a reduction in inequity. If it is more effective in the less disadvantaged group, the net outcome will be an increase in health inequity.
- An intervention can improve the population's overall health while increasing inequities. For example, historically health education campaigns to prevent smoking have tended to be more effective among those who are better off.

The ideal outcome is to benefit health in all groups, while reducing health inequities. However this may in practice be difficult to achieve. It may be justified in certain circumstances to increase overall population health while relative differences between groups remain the same (or get slightly worse). It is also possible that the information needed to achieve the most equitable outcomes is not available. These issues will be considered by the public health advisory bodies – the Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Groups (PDGs) – when developing recommendations. They will also be highlighted in the considerations section of final issued guidance.

Note that conforming to equality legislation will not necessarily mean that all health inequality and equity issues will be dealt with. Hence, PHIAC/the PDG's examination of these issues has to go beyond the legal definition as stated under the Equality Act (2006).

1.4.2.4 Health gradient

In whatever way health or disease is measured, there tends to be a gradient on which the most socially and economically advantaged groups have better health and lower rates of illness and death than disadvantaged groups. In advanced western societies such as the UK, the shape of the gradient tends to be relatively smooth with mortality and morbidity increasing steadily as social disadvantage increases. Over time, the gradient as a whole tends to shift upwards because overall the health of most groups is improving. However, the degree and rate of improvement tend to be greater in higher social groupings, meaning that relative differences, and therefore the degree of inequities and inequalities, also tend to increase.

1.4.2.5 Universal and targeted interventions

Universal interventions, which produce overall health improvement, can exacerbate relative health differences, since the well-off tend to make more use of (and derive greater benefit from) available services. This tendency has implications for public health interventions. Greater efforts are required to help people at the lower end of the social gradient improve their health faster than the rest of the population to combat health inequity. In other words, special efforts need to be made with disadvantaged groups. The problem is that targeted interventions tend to be both more difficult to implement and less cost effective than universal programmes.

Targeted interventions can reduce health inequalities. However, producing interventions to target relative need is difficult, because the social differences that constitute the gradient are not well described in the epidemiological, sociological or intervention literature. Lack of detail makes it very difficult to accurately focus interventions on particular sections of the population.

Efforts at health improvement should not be aimed solely at the most disadvantaged or socially excluded groups: this approach might lead to negligible benefits for the whole population. This is important because the graded characteristic of health differences means that there are still considerable levels of ill health and premature death in the midrange of the gradient. So public health guidance should focus on all levels of the gradient but include elements targeted at the most disadvantaged groups.

Factors that lead to general health improvement – improvements in the environment, good sanitation and clean water, better nutrition, high levels of immunisation, good housing – do not necessarily reduce health inequity. This is because the determinants of good health are not necessarily the same as the determinants of inequities in health (Graham and Kelly 2004). It is necessary, therefore, to distinguish between the causes of health improvement and the causes of health inequities.

In developing guidance, the decision to focus on universal, targeted or hybrid approaches will be made on a case-by-case basis. In some areas of public health work, universal approaches designed to produce overall health improvement will be suitable; in other cases a focus on the most disadvantaged will be suitable, and in others a hybrid approach along the gradient will be needed. Decisions about whether to take a targeted, universal or hybrid approach to equity will be made at the point when the topic is scoped.

1.4.2.6 Evidence for addressing inequalities

Evidence used as the basis for producing guidance should be assessed to see if it meets equality and diversity criteria as follows:

1. Do the evidence questions reflect the scope (which would have been equality-assessed, see section 3.3)?
2. Is the search strategy sensitive to evidence about social differences in the population (see sections 4.3.1 and 4.5)?
3. Is there evidence about inequalities from stakeholders (see section 4.4.4)?
4. Are evidence-review criteria inclusive (see section 5.8)?
5. Has the relevant data on equalities and diversity been appropriately extracted and presented in the review evidence statements (see section 5.8)?
6. What is the state of the evidence base in relation to health inequalities (see section 5.8)?

These questions should be dealt with at appropriate points throughout the guidance process, including: development of search strategies, the assembly of evidence, data extraction and the review and presentation of evidence.

1.5 Types of knowledge and evidence

NICE public health guidance draws on evidence and knowledge from across a spectrum of sources, using different methodologies and approaches. Sources include:

- organisations
- practitioners
- the policy community, gained from the wider policy context
- research, gathered systematically with a planned design
- service users and carers.

The guidance may have many different audiences including different practitioner groups, decision and policy makers and commissioners of services. This includes those working in local authorities and other public sector organisations, and within the community, voluntary and private sectors. It includes employers as well as the general public. It may address a wide range of topics and issues, and usually does more than just make recommendations about effective and cost-effective interventions. As well as 'what works', it will often need to address: when, why, how and for whom it does (and does not) work (Pawson 2006).

NICE public health guidance needs, therefore, to be informed by a wide variety of evidence and other forms of information (Lomas et al. 2005) (see chapter 3). This will include knowledge gathered using explicit, systemic and replicable scientific and social scientific methods. It will also include models, theories, expert testimony, mapping/practice reviews, consultation and practice.

NICE public health guidance is developed using methods and processes that can incorporate these different types of knowledge and evidence at various stages.

1.5.1 Best available scientific evidence

Scientific evidence is 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, political science and health education and promotion. Evidence is selected and appraised according to well-defined criteria. It is summarised according to general principles developed by NICE and using methods that are appropriate for public health and social scientific evidence.

The core issues are:

- What is the most appropriate type of evidence to answer the question (see below and chapter 3 for further information)?
- How can the most relevant evidence (published and unpublished) be identified (see chapter 4)?
- How can the quality and applicability of evidence be assessed (see chapter 5)?
- How can evidence from different kinds of research be synthesised and, in particular, how can quantitative and qualitative data be combined (see chapter 5)?

The randomised controlled trial (RCT) is normally the most appropriate source of evidence for judging the efficacy of interventions. However, such evidence is not always available or appropriate: in some cases it is impossible for researchers to conduct RCTs since to do so would be unethical. Further, given the complexity of causal chains

in public health, the external validity of RCT findings often has to be enhanced by observational studies. For evaluating large-scale interventions, observational studies may be the only feasible option (Victora et al. 2004).

In public health, social scientific as well as clinical and epidemiological evidence is used to examine outcomes, context, process and implementation (as well as barriers to and facilitators for interventions). There is little academic consensus about how best to synthesise these different approaches and there is still less agreement about how to use these disciplines to develop guidance.

The following types of evidence are of particular importance:

- social scientific data as adjuncts to studies of clinical or public health effectiveness
- social scientific trials and experimental and quasi-experimental designs
- social scientific and other empirical information about context and process
- social scientific theories and models.

For each of the above, different methods of appraisal and synthesis will be appropriate.

1.6 Stakeholders

Stakeholders are central to the development of NICE public health guidance. Guidance is subject to scrutiny and validation by stakeholders throughout the development process to ensure the resulting recommendations are realistic and appropriate.

1.7 Quality assurance principles

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

- Guidance development processes are governed by clear, published statements of methods and process, including a standard timeline, which are developed and updated at regular, predetermined 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, predetermined intervals.
- Guidance publications are authorised for publication, on behalf of the Board, by the Guidance Executive.

- The Senior Management Team and the Board receive bimonthly 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 potential risks associated with the development of a piece of guidance 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 all work, including review in every case by a professional editor. (This does not include supporting documents such as fieldwork reports and evidence reviews compiled by collaborating centres or external contractors.)
- Dissemination of publications takes place after analysis of the extent and needs of the audience for each piece of guidance.
- Training is provided for staff to enable them to apply these principles in their daily practice.

1.8 References and further reading

Campbell M, Fitzpatrick R, Haines A et al. (2000) Framework for design and evaluation of complex interventions to improve health. *British Medical Journal* 321: 694–6

Davidson K, Goldstein M, Kaplan RM et al. (2003) Evidence-based behavioral medicine: what it is and how do we achieve it? *Annals of Behavioral Medicine* 26: 161–71

Dixon-Woods M, Agarwal S, Young B et al. (2004) Integrative approaches to qualitative and quantitative evidence. London: Health Development Agency

Flay BR (1986) Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine* 15: 451–74

Graham H, Kelly MP (2004) Health inequalities: concepts, frameworks and policy. London: Health Development Agency. Available from www.nice.org.uk/page.aspx?o=502453

Kelly MP, Morgan A, Bonnefoy J et al. (2007) The social determinants of health: developing an evidence base for political action. Final report to the World Health Organization Commission on the Social Determinants of Health, from the Measurement and Evidence Knowledge Network [online]. Available from www.who.int/social_determinants/resources/mekn_report_10oct07.pdf

Kelly MP, Stewart E, Morgan A et al. (2009) A conceptual framework for public health: NICE's emerging approach. *Public Health* 123 e14–20. Available from www.sciencedirect.com/science?_ob=MIimg&_imagekey=B73H6-4V5NT10-3-3&_cdi=11546&_user=10&_orig=browse&_coverDate=01%2F31%2F2009&_sk=998769998&view=c&wchp=dGLzVlz-zSkWz&md5=390c0efa1befbc3c9df71f10fc6e4b2a&ie=/sdarticle.pdf

Lomas J, Culyer T, McCutcheon C et al. (2005) Conceptualizing and combining evidence for health system guidance: final report. Ottawa: Canadian Health Services Research Foundation

Michie S, Berentson-Shaw J, Pilling S et al. (2007) Turning evidence into recommendations: protocol for a study of guideline development group. *Implementation Science* 2: 29

National Institute for Health and Clinical Excellence (2007) Behaviour change at population, community and individual levels. NICE public health guidance 6. London: National Institute for Health and Clinical Excellence. Available from www.nice.org.uk/PH6

National Institute for Health and Clinical Excellence (2008) Social value judgements: principles for the development of NICE guidance. Second edition. London: National Institute for Health and Clinical Excellence

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

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: 804–8

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: 886–92

Pawson R (2006) Evidence based policy: a realist perspective. London: Sage

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

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

Weiss CH (1995) Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In: Connell JP, Kubisch A, Schorr LB et al. editors. *New approaches to evaluating community initiatives: concepts, methods and context*. Washington DC: Aspen Institute

2 Topic selection and scoping the guidance

2.1 Introduction

Topics for public health guidance are selected by ministers and then referred to NICE. Prior to ministerial referral, there are a number of steps.

Ideas for topics come from a variety of sources including stakeholders, NICE advisory committees and members of the public. There is a facility for making suggestions on the NICE website (at www.nice.org.uk/topicselection). In addition, the Centre for Public Health Excellence (CPHE) runs workshops to gather potential ideas from stakeholders, including relevant government departments. Since April 2009, these discussions take place using the conceptual framework outlined in appendix A.

The framework is used to:

- help define the key issues involved in a broad topic area
- specify where more focused and clearly defined topics fit into the overall strategic map of public health topics developed so far by CPHE.

Topics are mapped onto the vectors of causation outlined in the framework (population, environment, society and organisation) and cross-classified according to the potential level of intervention (population, community, organisation, family/domestic or individual).

Using this organising framework, suggestions enter the topic selection process and are assessed against selection criteria which have been jointly agreed with the Department of Health (DH) following public consultation. These criteria can be found at www.nice.org.uk/niceMedia/pdf/DH_selection_criteria_July06.pdf. Suggestions which meet the selection criteria are prioritised by the Public Health Topic Selection Consideration Panel. This panel, which meets several times a year, is made up of external public health experts and lay members. It is attended by representatives from relevant government departments and other NHS bodies. An internal consistency check is carried out after the panel review to ensure topics do not overlap and are properly integrated with other public health or clinical guidance. The topics are then considered by the DH's Referral Oversight Group. Following that meeting the minister makes a final decision and, if the topic is chosen, makes a final referral to NICE.

After the final referral, NICE defines exactly what the guidance will and will not examine and determines whether it should be developed using the programme or intervention

process (see 2.2.6). This stage is referred to as 'scoping' and the document containing this information is referred to as the scope.

The remainder of this chapter describes how the briefing paper is prepared for the Public Health Topic Selection Consideration Panel and how the scope of the guidance is developed.

2.2 Briefing paper for the Public Health Topic Selection Consideration Panel

2.2.1 Using the conceptual framework to construct a logic model

The CPHE, assisted by NICE information services, will prepare the briefing paper based on the original topic suggestion. (External experts or the public health collaborating centres may also be involved.) The first step is to use the NICE public health conceptual framework to construct a topic-specific conceptual framework (see below and appendix A). This framework may incorporate the assumed relationships between action and outcomes and, if possible, any relevant programme theory or theory of change (Pawson 2006; Weiss 1995).

Figure 2.1 Conceptual framework for public health guidance

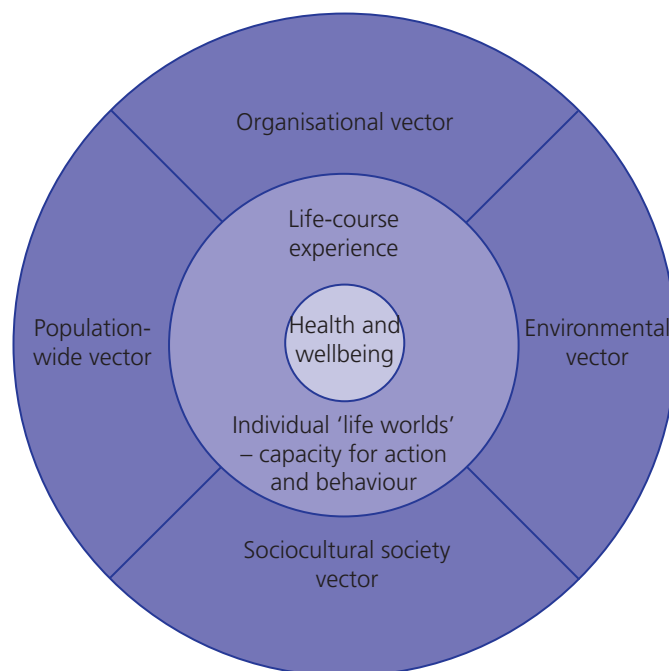


Figure 2.2 A conceptual framework for promoting mental wellbeing at work

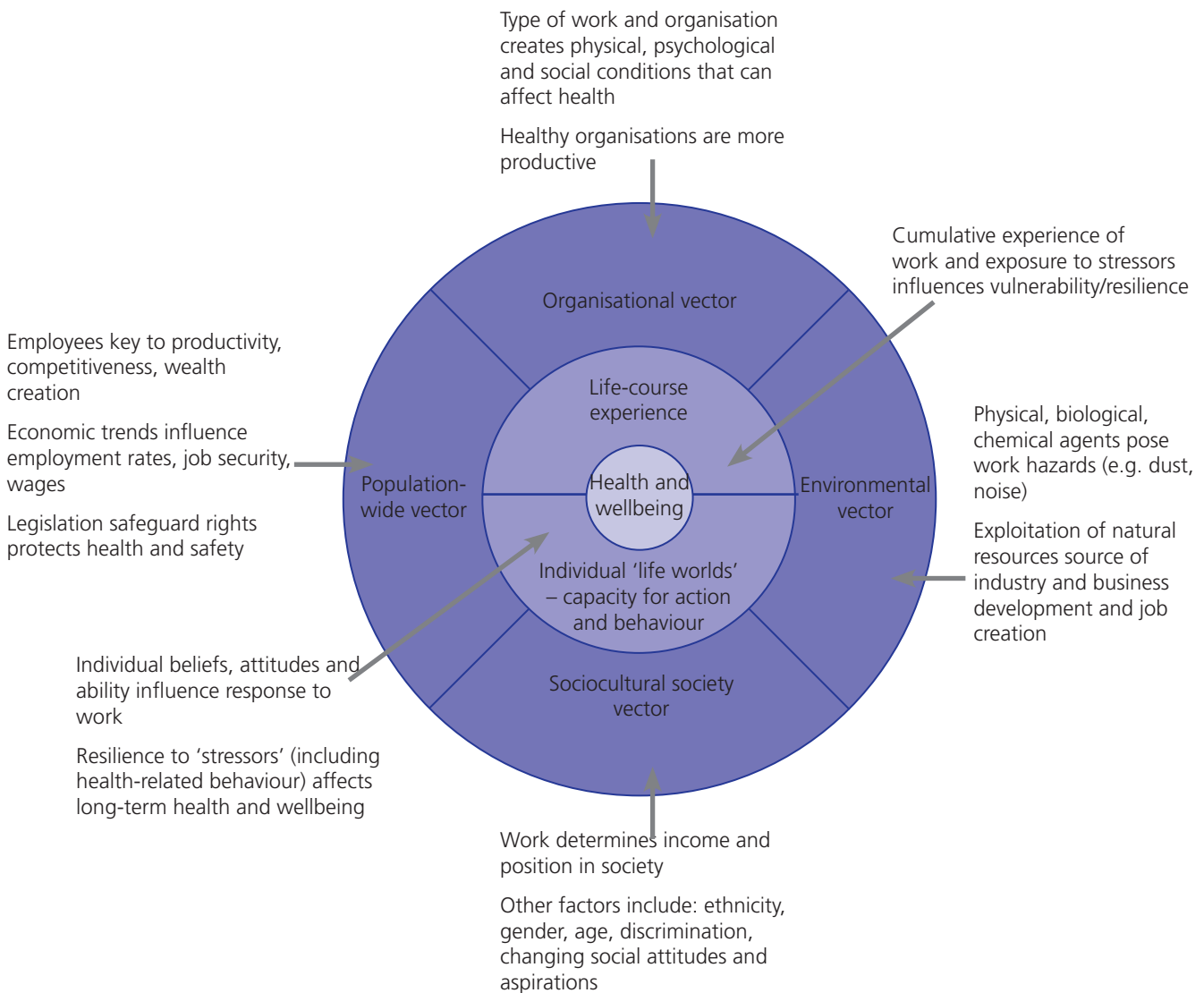
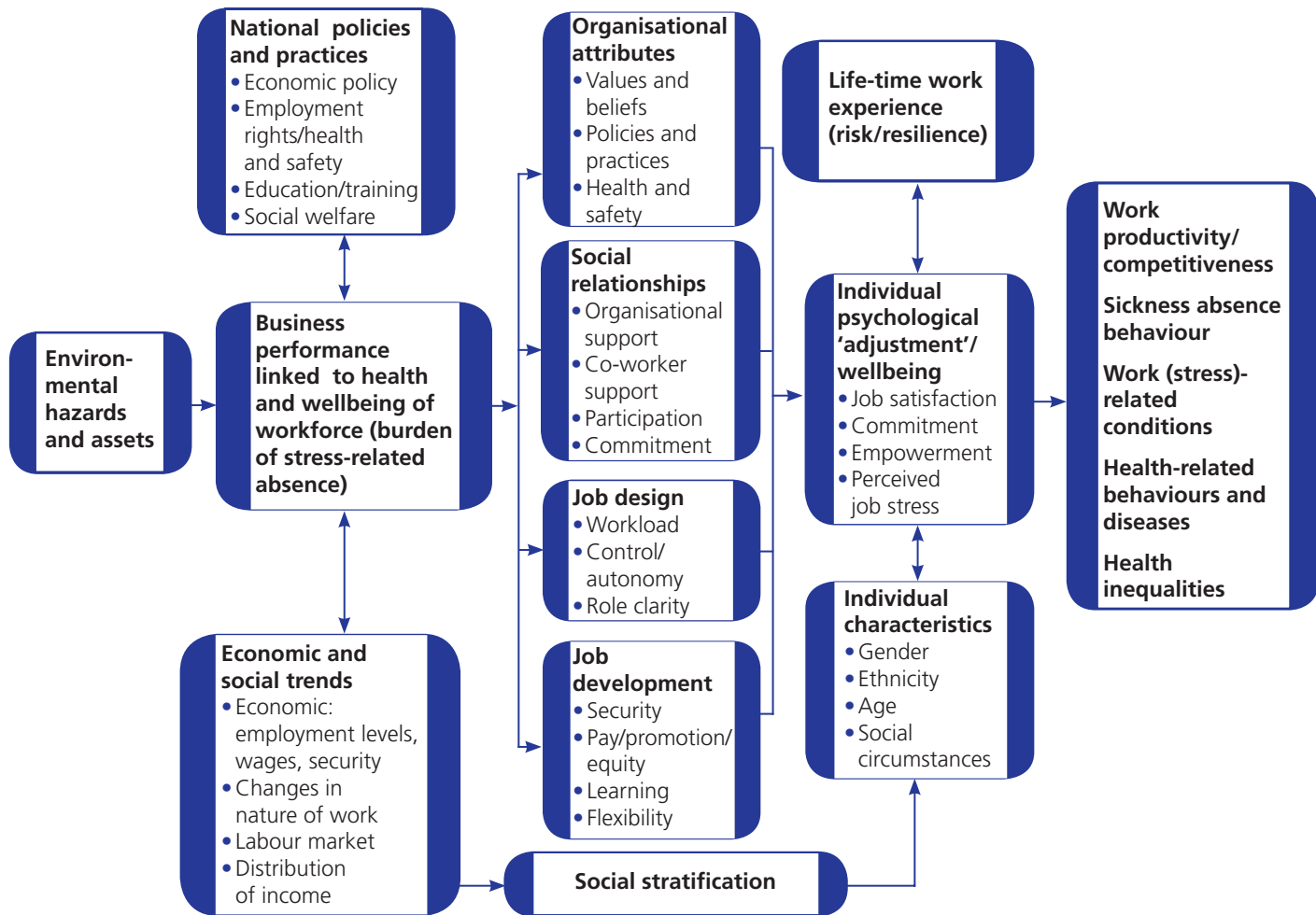


Figure 2.3 A logic model for promoting mental wellbeing at work



2.2.2 Preliminary information exercise

The topic-specific conceptual framework may be used to construct a logic model. This should incorporate the assumed relationships between action and outcomes and, if possible, any relevant programme theory or theory of change (Pawson 2006; Weiss 1995). Before the NICE information services team starts the preliminary search for information, it consults the logic model to determine which literature (including which type of literature) will be most relevant to the topic.

It then carries out the preliminary information exercise. In some cases, this will be related to a broad topic area, in others it will be more specific. It should bring together different sources of information to provide an overview of the topic (including key issues). This is not a comprehensive search but, rather, aims to provide a background for development of the subsequent scope.

Databases and websites may be searched to identify key epidemiological reviews, policy documents and existing reviews of effectiveness. It is particularly important to identify any related NICE guidance (existing or in development), other related guidance published in the UK (and abroad, if relevant) and relevant literature from the former Health Development Agency database. Key sources of relevant health inequalities information are also important.

Useful generic resources might include:

- standard bibliographic research databases – see below (these are useful for assessing the extent of the effectiveness literature and for epidemiological assessments of the extent of the problem and its determinants)
- relevant texts and monographs
- government and other policy websites
- guideline gateways in the UK and overseas
- websites of UK and overseas agencies that produce guidance or recommendations
- practice databases
- routine statistics (for example, monitoring and surveillance data).

The databases and websites searched might include:

- 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)
- Clinical Knowledge
- Clinical Trials.gov
- Cochrane Central Register of Controlled Trials
- Cochrane Library
- Current Contents
- Current Controlled Trials
- EMBASE

- King's Fund
- MEDLINE
- National Guidelines Clearinghouse (US) website
- National Library for Health (NLR)
- National Library for Health – National Library of Guidelines
- National Research Register (NRR)
- NHS Evidence
- NICE website
- PsycINFO
- SIGLE (grey literature)
- Sociological Abstracts
- Trials registers for periods not covered by the highest level of evidence available
- UK Clinical Research Network (UKCRN)
- Zetoc (electronic tables of contents).

The precise details of the search will be governed by the logic model and the topic.

2.2.3 Preparing the briefing paper

The briefing paper aims to:

- demonstrate how the suggested topic fulfils the selection criteria
- map the topic onto the vectors of causation and levels of delivery (see appendix A)
- outline the vectors and behaviours involved in the causal pathway
- outline the systems and processes involved in delivery
- establish the epidemiological importance of the topic
- identify links to policy
- describe any intervention involved as precisely as possible

- identify the links between the intervention and its health and other outcomes using a logic model, a programme theory or a theory of change, with reference to the vectors, behaviours and organisational structures
- identify any equity issues
- discuss whether a universal, targeted or hybrid approach is likely to be appropriate
- discuss the resource implications (for the NHS or other sectors) of taking or not taking action
- highlight issues that may affect implementation
- outline the approach for assessing effectiveness, cost effectiveness and equity
- develop the topic-specific logic model based on the overarching conceptual framework.

Briefing papers submitted to the Public Health Topic Selection Consideration Panel should not specify which guidance development process should be used.

2.2.4 Topic Selection Consideration Panel

The Public Health Topic Selection Consideration Panel reviews the briefing paper to decide whether the topic is suitable for inclusion in the NICE work programme, and if so, what level of priority it should be given. As part of this review, it considers the equity assessment carried out by the NICE topic selection team. The equity assessment is based on information gathered from standard epidemiology resources, supplemented by data from any background papers provided by those proposing the topic. The Panel assesses:

- the extent of inequalities in health in relation to the topic under consideration
- whether or not the potential guidance would address these issues.

2.2.5 Referral Oversight Group

If it is deemed appropriate, the topic is sent to the DH's Referral Oversight Group and then on to the minister for a final decision on referral.

2.2.6 Choosing the development approach

NICE public health guidance is developed using one of two processes: the intervention or programme development process (see 'The NICE public health guidance development

process: an overview for stakeholders including public health practitioners, policy makers and the public [second edition]' [NICE 2009]).

Generally, guidance developed using the public health intervention process focuses on local, clearly circumscribed actions, interventions or technologies. Typically these are delivered by frontline staff and target certain populations, communities or individuals.

Guidance developed using the public health programme process is broader. It is usually concerned with a general public health problem involving multi-agency and multi-faceted policies, services, systems and interventions. The activities covered may be topic-, setting- or population-based and may involve changes to organisational infrastructures.

Public health guidance based on the intervention process is developed by NICE's independent Public Health Interventions Advisory Committee (PHIAC). Independent programme development groups (PDGs) are set up by NICE to develop guidance based on the public health programme process.

After the ministerial referral, the CPHE senior team will determine whether the topic is best dealt with methodologically as a single programme or intervention, a group of interventions, or as a group of interventions and a programme. This decision should be made in light of discussions held by the Public Health Topic Selection Consideration Panel and the Referral Oversight Group, and in consultation with relevant advisory committees and the rest of the CPHE team. The individual pieces of work are then added to the CPHE's forward programme.

The next stage is to scope the individual pieces of work (see 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public [second edition]' for further details).

2.3 Purpose of the scope

The logic model constructed for the briefing paper is the starting point for development of the scope.

The scope for guidance being produced using the public health intervention process should specify **precisely** which intervention(s) are covered, by describing them in some detail. There should only be one or two. The scope for guidance being produced using the public health programme process should specify what **types** of interventions/strategies/activities are covered and may include some specific examples.

In addition, the scope should:

- Provide a clear definition of the intervention(s) or programme to be addressed and, where appropriate, the relationship between them. This definition should cover the nature, content and the way each public health action is delivered.

- Describe the assumed mechanism/mediator/link between the action(s) and the outcome(s). In the case of guidance based on the intervention process, outcomes will normally be specified in terms of health or disease; an intermediate outcome could be a behaviour leading to the disease or to health improvement. For guidance based on the programme process, outcomes may be more general, relating to strategy, organisational structure or service delivery.
- Identify the causal pathway with reference to the vectors, health behaviours and means of delivery (appendix A).
- Identify the level of delivery (individual, family, community or population-level).
- Identify the settings where the intervention(s) takes place.
- Define what the guidance will include and exclude.
- Identify the population(s) to be included and excluded.
- Briefly describe the relevant epidemiology.
- Identify the approach to equity.
- Set the policy context.
- Include the key, overarching questions that the guidance will address.
- Outline the issues which PHIAC/the PDG is likely to consider.
- Set clear parameters to ensure the guidance can be developed within the allocated time period.
- Identify the economic approach to be taken, including any additional perspectives that will be taken into account (apart from the NHS), such as that of employers or the private sector.
- Consider the extent to which evidence may be generally applicable.
- Specify the outcomes and any comparators that will be used to judge effectiveness.
- Indicate what kinds of lay evidence will be appropriate.
- Identify links to other NICE guidance.

2.4 Drafting the scope

The following sections describe the procedures involved in developing the draft scope. A further literature search may be needed. If so, it must be based on the logic model. In addition, a list of 'considerations' should be developed (see 2.4.2).

2.4.1 Developing key questions

The CPHE project team should refine the outline logic model constructed for the briefing paper (see figure 2.3) to draft a set of key questions that the guidance will aim to address. (These key questions, in turn, will be the starting point for developing research questions for the reviews.) The number of questions will depend on the topic, whether the intervention or programme development process will be used, and the breadth of the scope. However, it is important to ensure the total number:

- is manageable for PHIAC/the PDG
- is feasible within the allocated budget
- provides a sufficient focus for the guidance.

Key questions should be clear, realistic and focused. They should cover all areas of the guidance remit and should not introduce new aspects. They should be concerned with effectiveness, cost effectiveness, the relationship between interventions and outcomes, feasibility and acceptability and inequalities (also see section 2.4.2). They may cover the determinants of health, including risk and protective factors (and the relationships between them). Most important, the key questions should clearly relate to the interventions being considered and the outcomes that will be used to assess their effectiveness and cost effectiveness. Therefore, in addition to the logic model, the CPHE project team should use the population, intervention, comparison and outcome (PICO) framework to specify interventions, as follows.

- 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)?

2.4.2 Considerations

The list of considerations should set out the factors that PHAC/the PDG are likely to consider when developing recommendations. They may become sub-questions for the reviews and may indicate where expert testimony will be required. They may include the various confounding factors that could influence the outcome, effectiveness and cost effectiveness of each type of activity, along with any other relevant information. The programme theory or theories of change approach and the logic model adopted may be helpful when compiling this list.

Typically, some of the following issues may be considered:

- What factors and 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 and cost effectiveness? (For example, how valid is self-report versus biologically validated measures?)
- How does the content of an intervention or programme influence effectiveness and cost effectiveness?
- How does the way the intervention or programme is carried out influence effectiveness and cost effectiveness?
- Does effectiveness and cost effectiveness depend on the job title or position of the intervener or other factors such as their age, gender or ethnicity? What are the significant features of an effective deliverer or leader?
- Does the site or setting influence effectiveness and cost effectiveness?
- Does the intensity, length or frequency of the intervention influence its effectiveness and cost effectiveness or duration of effect?
- How does effectiveness and cost 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?
- Are there any factors that prevent – or support – effective implementation?
- How acceptable is the intervention to the target population?

- Is it effective and cost effective?
- Are there any adverse or unintended outcomes?
- Are there any trade-offs between equity and efficiency?
- What differences are there in terms of availability and accessibility for different groups?

The considerations, key questions and other parameters (for example, populations, activities and outcomes) outlined in the scope should be used to develop specific research questions for the reviews (see chapter 3). The programme theory or theories of change approach that has been adopted may also help.

Equity issues should have been considered at the briefing paper stage and should be considered again as the scope is being written. Any issues identified should be used to find interested groups to invite to become stakeholders (these issues may also need to be raised with them).

In certain circumstances, prior to finalising the draft scope, it may be appropriate to hold a meeting with policy and other colleagues to help position the work relative to government policy. The meeting could include national clinical directors in the DH or key officials in other government departments.

2.5 Consulting on the scope

Once the scope has been drafted, it is posted on the NICE website and stakeholders are notified. They can register their interest, attend a public consultation meeting and provide feedback during the consultation period.

After the public consultation, the scope should be reassessed for its potential impact on health inequities and any necessary amendments noted in the stakeholder response table. A final scope is prepared and published, along with stakeholder comments and NICE's response to those comments.

2.6 Further reading

Kelly MP, Stewart E, Morgan A et al. (2009) A conceptual framework for public health: NICE's emerging approach. *Public Health* 123: 14–20. Available from www.sciencedirect.com/science?_ob=MIimg&_imagekey=B73H6-4V5NT10-3-3&_cdi=11546&_user=10&_orig=browse&_coverDate=01%2F31%2F2009&_sk=998769998&view=c&wchp=dGLzVlz-zSkWz&md5=390c0efa1befbc3c9df71f10fc6e4b2a&ie=/sdarticle.pdf

Pawson R (2006) Evidence based policy: a realist perspective. London: Sage

US Centers for Disease Control and Prevention (2005) The guide to community preventive services. What works to promote health? Atlanta: US Centers for Disease Control and Prevention. Available from: www.thecommunityguide.org

Weiss CH (1995) Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In Connell JP, Kubisch A, Schorr LB et al. editors. New approaches to evaluating community initiatives: concepts, methods and context. Washington DC: Aspen Institute

3 Determining the evidence for review and consideration

3.1 Introduction

NICE public health guidance is informed by a variety of types of evidence (Lomas et al. 2005). This chapter describes some of the types of NICE review that can be commissioned to provide good quality, scientific evidence to answer different types of research question. It also outlines how to develop specific research questions for these reviews. It concludes by describing how other types of evidence may be used to help develop the guidance.

3.2 Types of evidence

The Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Group (PDG) need both scientific and other types of evidence about what works generally, why it works, and what might work (and how) in specific circumstances. They may also need to consider evidence about the impact on, or user experience of, different types of interventions, and to place this alongside other issues related to context, ethics and theory, in making their recommendations (Tannahill 2008). They need evidence from multiple sources, extracted for different purposes and through different methods.

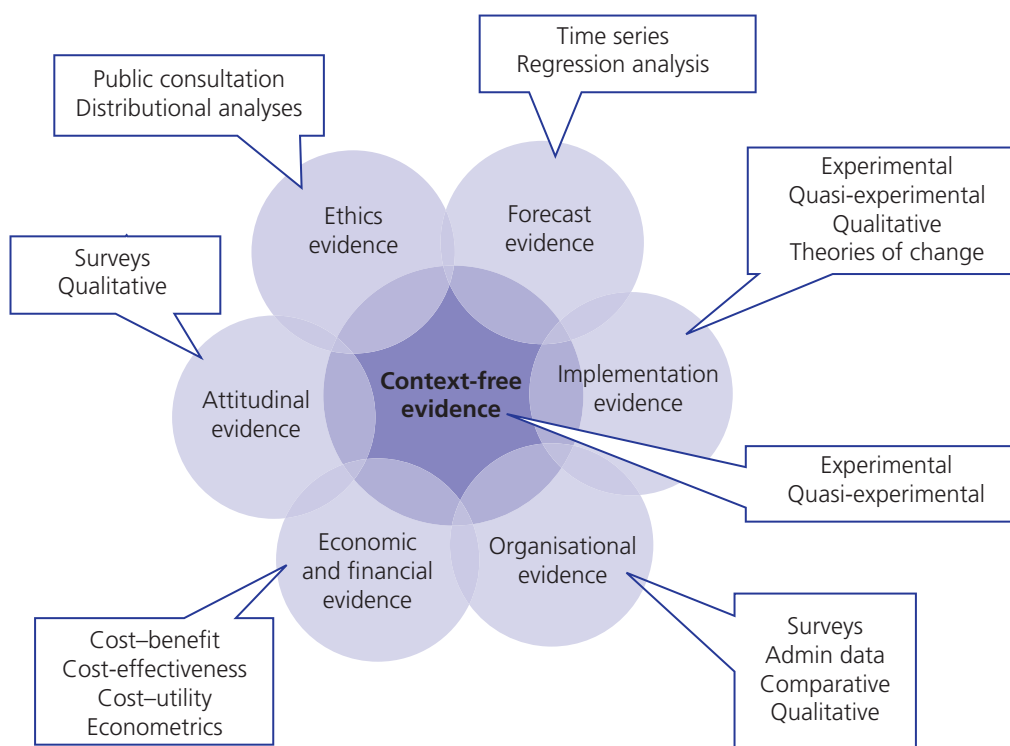
3.2.1 Scientific evidence

Scientific evidence is explicit (codified and propositional), systemic (uses transparent and explicit methods for codifying), and replicable (using the same methods with the same samples will lead to the same results). It can be context-free or context-sensitive (Lomas et al. 2005). Context-sensitive evidence can complement context-free evidence (see figure 3.1):

- Context-free scientific evidence explains universal truths about what might generally work, in ideal circumstances. It is knowledge produced from a scientific framework which assumes that objects of the natural world are real, objective, measurable and exist independent of the observer and context (Woolgar 1988). It can be derived from systematic reviews or meta-analyses of quantitative research of intervention effectiveness, primary empirical research or theoretical models or propositions.

- Context-sensitive scientific evidence is concerned with what works and how, why or whether it might work in specific circumstances. It includes information on attitudes, implementation, organisational capacity, forecasting, economics and ethics. It is mainly derived from social scientific and behavioural research methods including: quantitative and qualitative research studies, surveys, theories, cost-effectiveness analyses and mapping reviews. Sometimes, it is even derived from systematic reviews and the other research techniques described above. Context-sensitive evidence can be used to complement context-free evidence, providing the basis for more specific and practical recommendations (see figure 3.1). It may be used to:
 - supplement evidence on effectiveness (for example, to look at differential effectiveness according to occupation, educational attainment and income)
 - construct logic models (see chapter 2) and causal pathways (for example, to explain what factors predict teenage parenthood)
 - provide information about the characteristics of the population (including its social and physical circumstances) and about the process of implementation
 - describe psychological processes and behaviour change.

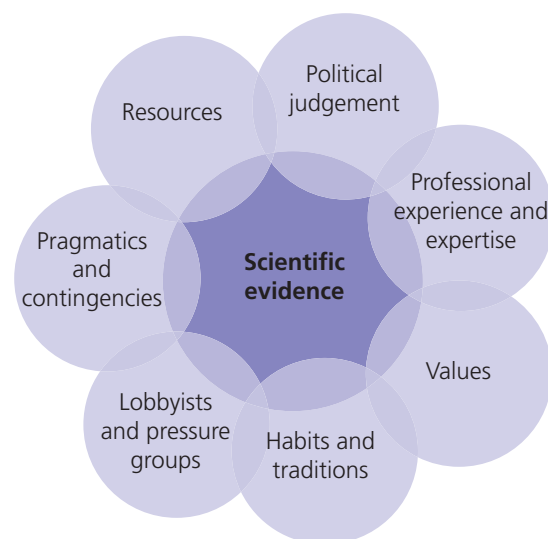
Figure 3.1 Scientific evidence: context-sensitive complements context-free
 (Source: Lomas et al. 2005 adapted from Davies 2005)



3.2.2 Colloquial evidence

‘Colloquial evidence’ can complement the scientific evidence or provide missing information on context (see figure 3.2). It may come from expert testimony, from members of PHIAC or the PDG, or from stakeholder comments (see section 3.5). It includes evidence about values (values and political judgement), practical considerations (resources, professional experience/expertise and habits/traditions) and the interests of specific groups (views of lobbyists and pressure groups).

Figure 3.2 Colloquial evidence informs scientific evidence
(Source: Lomas et al. 2005 adapted from Davies 2005)



3.3 Types of review and types of research question

Every piece of NICE guidance is informed by several high-quality reviews (one to two for guidance produced using the intervention process and four to five for guidance developed using the programme process). These reviews explicitly address research questions based on the key questions in the scope. Rather than relying on the standard hierarchy of evidence (with randomised controlled trials [RCTs] at the top), a wide range of study designs and methodologies should be used to answer these questions (Petticrew and Roberts 2002). This is because:

- Public health covers a broad range of (often multiple) interactions between the different effects of an intervention (even when the intervention is relatively simple). Public health methods have to reflect this complexity.

- In the traditional evidence hierarchy, RCTs are usually accorded the highest status. However, it is difficult to run RCTs to test many public health issues; they tend to produce an over-inflated effect estimate and are usually limited to answering questions on the efficacy of interventions.

Table 3.1 shows the type of evidence that should be used to address different types of research question and the type of standard NICE review it will involve. It builds on Petticrew and Roberts' general typology of public health evidence for social interventions with children (Petticrew and Roberts 2002).

3.3.1 Types of review research question

The key questions in the scope are based on the Department of Health's (DH) original referral, the topic area and the views of practitioners, decision makers and other stakeholders. All public health guidance aims to recommend the most effective ways to promote health, prevent disease and reduce health inequalities. However, the scope may include several other questions and potential considerations that reflect the nature of the specific issue being tackled and its context (see sections 2.4.1 and 2.4.2).

The Centre for Public Health Excellence (CPHE) project team and the review team should use the key questions and considerations from the final scope, along with its other parameters (populations, activities and outcomes) to develop more specific research questions for the reviews. In addition to questions of effectiveness and cost effectiveness, there will often be questions about the epidemiology of a problem or issue, the acceptability and accessibility of interventions, and client or practitioner experiences.

The type of research questions will determine the number and type of reviews and the type of evidence they include (for example, intervention studies and qualitative data). Whatever method is used, the process for developing questions is the same.

The CPHE project team and the review team should draft clear, focused review questions. The exact structure of each question will depend on what is being asked, but it is likely to cover one of the following:

- extent and nature of the public health problem
- factors, causal mechanisms and the role of the various vectors (see appendix A)
- interventions that work in ideal circumstances and might work in specific circumstances (the extent to which something works, how and why)
- a relevant programme theory or theory of change
- views and experiences of the target population (people who may be affected by the recommendation), including how acceptable and accessible they find the intervention

Table 3.1 Types of evidence and standard⁴ NICE reviews to address different research questions

Research question	Type of evidence					
	Systematic review of effectiveness (and cost effectiveness)	Experimental study	Observational study	Qualitative study	Practice/ case report	Economic/ cost-effectiveness study
Extent of public health problem/issue			NICE epidemiological review			
Factors/ determinants/ associations	NICE review of reviews		NICE correlates review			
Intervention effectiveness/ cost effectiveness	NICE review of reviews	NICE effectiveness review (can include observational and qualitative studies as well as experimental studies)				NICE cost-effectiveness review
Views and experiences of practitioners			NICE correlates review	NICE qualitative review	NICE mapping report	
Views and experiences of target population			NICE correlates review	NICE qualitative review		

⁴ Note that it is possible to undertake 'hybrid' reviews of more than one type of evidence and/or of more than one type of research question

- practitioners' views, experiences and working methods (including any barriers to and factors supporting implementation of the intervention)
- cost effectiveness
- potential for an intervention to do harm.

The different types of review that may be carried out to answer these questions are set out below – along with how to develop appropriate research questions in each case.

Every piece of public health guidance will be informed by at least one effectiveness review. The decision on whether or not to use other, additional types of review will depend on the topic area and the type, depth and breadth of relevant evidence available. Sometimes, a review may draw on a combination of different sources of evidence or types of data (for example, combining mapping information and qualitative data). The nature of the area, the technical expertise and topic knowledge of the CPHE and collaborating centre teams, and feedback from stakeholders at the scoping stage may all influence this decision.

The remainder of this chapter gives a general description of when different types of review may be considered and the process of developing appropriate research questions for them. It does not rule out the possibility of using 'hybrid' reviews (for example, combining mapping information and qualitative data), but simply illustrates a range of potential approaches.

3.3.2 NICE effectiveness reviews

The CPHE project team and the review team should explicitly consider what type of study will provide the best evidence to answer a question on effectiveness. Traditional systematic reviews of effectiveness answer specific questions about the effectiveness and efficacy of a particular intervention. Usually, they draw on RCTs as this type of study is most likely to give an unbiased estimate of the effects. However, for many public health topics it can be difficult or unethical to assign populations to control and intervention groups (for example, for interventions aiming to change policy). In such cases, a non-randomised controlled trial (see appendix D) might be a more appropriate way of establishing cause and effect.

3.3.2.1 Review research questions

At least one review question, possibly more (depending on the populations and outcomes of interest), will be needed for each intervention considered. The CPHE project team must ensure these questions are clear and focused.

As PHIAC/the PDG need to take into account factors that could affect the outcomes and effectiveness of an intervention, it may be helpful to list outcomes and other key criteria.

Effectiveness reviews could also consider context and implementation issues, as far as the available evidence permits – although these may also be dealt with in a separate evidence review.

The PICO framework is a helpful, structured approach to formatting intervention questions and should certainly be used as a starting point. Other approaches or questions may also be used, where helpful.

Box 3.1 PICO guide to review questions on intervention effectiveness

- **Population:** Which populations are we interested in? How can they best be described? Do any subgroups need to be considered?
- **Intervention:** Which intervention, activity or approach should be used?
- **Comparison:** What is/are the main alternative(s) to the intervention being considered?
- **Outcome:** What outcomes should be considered? (Examples include increases in knowledge, skills or availability of services; changes in behaviour; mortality rates; morbidity rates; changes to quality of life; and adverse effects.)

Once the review questions have been framed, the review team should identify key words as potential search terms.

Examples of research questions

- What types of mass-media intervention help prevent children and young people from taking up smoking? Are the interventions delaying rather than preventing the onset of smoking?
- Which of the harm-reduction services offered by needle and syringe programmes (including advice and information on safer injecting, onsite vaccination services, and testing for hepatitis B/C and HIV) effectively reduce blood-borne viruses and other infections among people who inject drugs?
- What types of intervention and programme effectively increase physical activity levels among children under 8 – particularly those who are not active enough to meet the national recommendations for their age – or help to improve their core physical skills?
- Does brief advice from GPs increase adult patients' physical activity levels?

- Do school-based interventions effectively change young people's attitudes to alcohol use?

3.3.3 NICE epidemiological reviews

NICE epidemiological reviews describe the public health problem.

The topic referral and scope development stages may draw on observational studies to gauge the nature of the public health problem (see chapter 2). Most topics, however, will benefit from further epidemiological reviews to determine the focus of the guidance. For example, such a review of accidents would provide information on the most common ones, as well as morbidity and mortality statistics. This would help focus the research questions. The resulting recommendations would be based on a logical appraisal of the problem and its causes, and hypotheses about the kinds of interventions that could address them.

The CPHE project team will commission an epidemiological review if:

- the exact nature of the public health problem is unclear
- those who could benefit from the guidance cannot be clearly defined
- the cause of the public health problem is unclear.

Examples of research questions

- What are the patterns of physical activity among children from different populations and of different ages in England?
- Which populations of children are least physically active and at which developmental stage are all children least physically active?
- What effect does physical activity have on children's health and other outcomes in the short and long term?

3.3.4 NICE correlates reviews

NICE correlates reviews describe relationships between epidemiological factors and outcomes.

If an epidemiological review has been carried out, information will have been gathered from observational studies on the nature of the public health problem. However, further analysis of this information – in the form of a correlates review – may be needed to establish the epidemiological factors associated with any positive or negative behaviours or outcomes.

Examples of research questions

- What factors are associated with children/adolescents' physical activity and how strong are those associations?
- What are the barriers to and facilitators for participation in physical activity?
- How do the barriers and facilitators differ for the least active subpopulations and age groups?

3.3.5 NICE qualitative reviews

Where there is a mature body of relevant evidence, or where the evidence is important enough to merit a substantive piece of work, the CPHE project team may commission a review of qualitative research.

Qualitative studies may be the primary source of evidence to answer questions on:

- client or practitioner experiences (including information on what works, for whom and under what circumstances)
- patient and practitioner views
- opportunities for and barriers to improvement (including issues of access or acceptability)
- variations in delivery and implementation for different groups, populations or settings
- barriers to and facilitators that aid implementation
- social context and the social construction and representation of health and illness
- background on context, from the point of view of an observer (and not necessarily that of a user or practitioner)
- an explanation of the associations between interventions and outcomes.

Qualitative information can also be used to enhance logic models (see chapter 2).

Of course, some evidence from quantitative research may be able to answer these types of questions: this can be picked up by the effectiveness (or other) review(s). Alternatively, the review team may choose to include supplementary quantitative evidence within a qualitative review, when appropriate.

Examples of research questions

The CPHE project team, in conjunction with the review team, should fine-tune questions posed by the scope to produce research questions for each qualitative review. These

questions should be relevant, specific (about factors such as the populations or settings involved), actionable and realistic in relation to the area covered by the review.

The PICO framework (see box 3.1) may help ensure the views and experiences of the target population are integrated into questions about intervention effectiveness. Specific aspects of these views and experiences, the social context or the social construction of an issue or area can be tackled in separate review questions. (However, if the review does not involve an intervention, the PICO framework may not be appropriate).

Questions that are too broad and lack focus should be avoided (for example, 'What is the experience of looked after children?'). Examples of good questions include:

- How do different groups of practitioners, clients or stakeholders perceive the issue (for example, broken down by profession, age, sex or ethnicity)?
- What social and cultural beliefs, attitudes or practices might affect this issue?
- How do different groups perceive the intervention or available options? What are their preferences?
- What approaches are used in practice? How effective are they, according to different groups of practitioners, clients or stakeholders?
- What is a desired, appropriate or acceptable outcome as far as the client groups are concerned? What outcomes are important to them? What do professional, client or stakeholder groups perceive to be the barriers to, and opportunities for, change in this area?
- What status or importance do professional or public groups attach to NICE guidance?
- What do the target populations think about current or proposed practice?
- Why do people make the choices they do or behave in the way that they do?
- How is a public health issue represented in the media and popular culture?
- How does an issue such as access or acceptability affect the delivery and effectiveness of an approach?

Once the research questions have been agreed between the review team and the CPHE project team, the review team should develop and implement a search and appraisal process (see chapters 4 and 5). As with other types of review, the review team should agree a set of inclusion and exclusion criteria with the CPHE project team. This should be consistent with the scope and research questions.

Examples of research questions for qualitative evidence that supplements a NICE effectiveness review

The process for developing research questions for an effectiveness review which will include qualitative evidence is similar to that described in section 3.3.2. A qualitative component should be included if there are issues that could be better assessed using qualitative evidence. For example, to assess **why** an intervention does or does not work.

Research questions should be developed by the CPHE project team and the review team, using the scope and the main review questions. Again, care should be taken to ensure the questions are relevant, specific, realistic and actionable. They might include:

- How is the intervention perceived by different client or practitioner groups?
- How acceptable is the intervention?
- How accessible is the intervention or service to different client groups? What factors affect its accessibility?
- Does the mode or structure of delivery (including characteristics of relevant health or other professionals, the setting and language) affect user perceptions?

3.3.6 NICE cost-effectiveness reviews

Cost-effectiveness reviews aim to identify, appraise and present data from economic studies (see chapter 6). Cost-effectiveness data may be considered as part of each effectiveness review undertaken for a piece of guidance, or in just one or two key areas (depending on the topic and the available data). Collation of this data is vital for any subsequent modelling. It may be integrated within the effectiveness reviews or presented as a separate report.

Research questions

Cost-effectiveness review questions will mirror those used for the effectiveness reviews, but with a focus on cost effectiveness.

3.3.7 NICE mapping reports

Where evidence from published formal research is limited, a mapping report could provide a snapshot of current practice. It might address, for example, current professional practice or multidisciplinary theoretical accounts. It could draw on published or grey literature – or both. (Grey literature is unpublished information or locally published information from practice, for example, audit reports.)

A mapping report can help with the development of other reviews, supplement the knowledge gained from other reviews, help inform recommendations and help guide

the implementation process. It does not provide a quality assessment of the material. There are two types:

- **A map of the literature** describes the type (research method, design or theoretical basis/approach) and quantity of published material. Selected databases of important literature are sampled and a typology developed. These reports are useful where there is a lot of material or when several keywords are used for indexing complex topics. They can help to identify or clarify areas where a review of effectiveness might be useful and to determine the scope and breadth of the guidance.
- **A map of current policy and practice** is useful when the available evidence is largely non-UK based. It can help deal with applicability issues and identify how to overcome barriers to effective implementation. This type of mapping study can encompass published information such as practice surveys (based on context-sensitive scientific evidence). Alternatively, it can draw on practitioner views and reports (colloquial evidence) to highlight the conditions needed for interventions to work in practice (detail that is often absent from published material). Mapping policy and practice can help to ensure the context within which the recommendations are made is up to date. It is important because some of the most innovative action is never written up, due to a lack of expertise or resources.

3.3.8 NICE reviews of reviews

Generally, review-level material (for example, systematic reviews, literature reviews and meta-analyses) will **only be used** as an additional source of potentially relevant primary studies. In such cases, the review team should screen reference lists of reviews to identify additional primary studies that might not have been captured in the electronic or other searches. Generally, review-level material will not be quality-assessed or data-extracted for inclusion in the evidence reviews because it:

- rarely covers inclusion/exclusion criteria relevant to the guidance topic's referral and parameters (for example, comparable research questions, relevant outcomes, settings, population groups or time periods)
- often aggregates different outcome or study types, some of which may be part of the topic's exclusion criteria and would need to be disaggregated
- often covers both clinical and public health-related data which is either difficult or impossible to disaggregate
- rarely provides enough data to develop recommendations: obtaining detail on the implementation of specific interventions is particularly problematic and, as a result, it is often necessary to refer to the primary studies.

Review-level material can be useful when developing the scope and when defining research questions, outcomes and outcome measures for the evidence reviews. The discussion section can help identify some of the limitations or difficulties associated with a topic area. It can also provide a critical account of the state of the evidence base for a specific topic (see chapter 2).

Review-level material should only be quality-assessed, data-extracted and integrated with primary studies within the evidence reviews in the following, exceptional circumstances:

- If the parameters (for example, research question, inclusion/exclusion criteria) of a recent, good quality review (see screening checklist in appendix J) are sufficiently similar to the guidance topic's research parameters to be able to answer one or more specific research question. In such cases, a search should be undertaken for primary studies published after the search date covered by the identified review.
- If a series of systematic reviews and/or meta-analyses for a specific well-defined topic have led to a definitive decision or consensus on a specific research question and investment in further research has stopped.
- If a topic referral draws heavily on published theories (for example, on how to support attitude and behaviour change).
- If the evidence base for the specific topic is so large that resource limitations make it impossible to cover all available primary studies.

The review team should discuss and agree with the CPHE project team when 'exceptional circumstances' dictate that review-level evidence should be included in a review. See appendix J for further details.

3.4 Planning reviews

The review team will prepare a protocol that outlines the background, objectives and methods for each review. It should be agreed with and signed off by the CPHE project team. The protocol provides an explicit public record of how the review is to be carried out, making it possible for it to be repeated by others at a later date. The protocol should also make it clear how inequality issues will be considered, where appropriate.

3.4.1 Structure of review protocol

The review protocol should describe any deviations from the methods described in this manual and should include the components outlined in table 3.2. It should not duplicate the methodology stated in this manual.

Table 3.2 Components of the review protocol

Component	Description
Guidance title	Full and short title of guidance
Review team	Provide names, affiliations and number of whole-time equivalent members of the review team. Where lead/key staff are working on other NICE reviews, please state
Clarification of the scope	
Groups that will be covered	Based on the scope, additional detail as required
Groups that will not be covered	As above
Activities/interventions that will be covered	As above
Activities/interventions that will not be covered	As above
Other aspects of the scope	Include other relevant information. Outline schematically, where relevant, how the scope relates to NICE's public health conceptual framework (see chapter 2 and appendix A)
Overview of project	Provide a brief overview of project aims and objectives. If a logic model or conceptual framework is to be used, provide a brief outline including how it will inform the search process
Review questions	Make these clear, focused and realistic, adopting the PICO approach (if relevant). Specify all the outcomes of interest relevant to the review populations, subpopulations, settings etc. Include any subquestions
Methods	Include details of: inclusion criteria, data extraction, quality assessment, data synthesis and any deviations from the NICE method or process manuals. Include the search protocol – any limitations, such as country, language, dates and study design should be stated, along with an initial list of the search terms. If appropriate these details can be listed by review. Separate search protocol(s) to be developed as part of the first phase of the review should be appended to that review protocol
Economic analysis modelling	Describe all components
Timetable	Provide a table showing agreed dates for: PDG/PHIAC meetings, the search protocol/searches, delivery of draft and final reviews, NICE team comments and for progress meetings. Where the focus of some reviews is still to be determined, insert the date(s) for agreeing the protocol of each one; once agreed, append to this document
Deliverables	Describe what products will be delivered for/presented to each meeting/deadline for reports
Additional information	Any other relevant information
Date and version of protocol	

3.4.2 Process for developing the review protocol

As a first step, the CPHE project team and the review team should develop detailed research questions for the review, using the key questions and considerations from the final scope, along with the other parameters of the guidance (populations, activities and outcomes) – see section 3.3.1.

The types of question will dictate which methodology will be used and the types of evidence that will be considered. All review protocols should be included as appendices in the review. Any changes made to a protocol in the course of guidance development should be described in the appendix.

3.5 Colloquial evidence

The different types of review outlined in this chapter mainly focus on gathering and assessing scientific evidence (mapping reports are an exception – see section 3.3.7). However, ‘colloquial evidence’ – about values, practice, political judgement, operational considerations and interests – is also key to developing public health guidance. It takes the following forms.

3.5.1 Expert testimony

An expert witness may be invited to give testimony when:

- reviews have uncovered significant gaps in the evidence (or the CPHE project team is aware from the outset that the formal evidence is likely to be limited)
- the available evidence conflicts significantly
- PHIAC/the PDG wishes to seek the views and experiences of specific groups of practitioners or clients/service users.

Experts may be identified via stakeholders, via PHIAC/the PDG, or in the course of carrying out the reviews (for example, key authors or researchers). The testimony will usually take the form of a short, focused presentation to the committee, followed by discussion. Witnesses will normally be asked to submit a short paper (with references to any relevant published work) based on their presentation. This is treated as evidence and subject to consultation, along with any reviews.

3.5.2 Consultation with stakeholders

Stakeholder views are routinely sought during the development of public health guidance: at the scoping stage and during consultation on the draft guidance. See

sections 4.3 and 4.5 of 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public (second edition)' (NICE 2009).

The CPHE project team always considers and responds to comments made during these consultation exercises and PHIAC/the PDG takes them into account when developing the final guidance. All stakeholder comments – and CPHE project team responses – are available on the NICE website.

3.5.3 Fieldwork

The CPHE project team carries out fieldwork with key practitioner groups on the draft recommendations (see chapter 8). Fieldwork findings are taken into account by PHIAC/the PDG when developing the final guidance. The findings are also published on the NICE website.

3.5.4 Committee discussion and expertise

PHIAC and the PDGs comprise public health practitioners, clinicians (both specialists and generalists), local authority officers, teachers, social care professionals, representatives of the public, patients and/or carers, academics and technical experts. See sections 5.2 and 6.2 of 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public (second edition)' (NICE 2009). They bring considerable expertise and knowledge to bear during the process of developing guidance. Minutes of committee meetings are published on the NICE website.

3.5.5 Mapping reports of local practice

Mapping reports can help to clarify local practice, research or theory. For more information, see section 3.3.7.

3.6 Equality and diversity

Specific issues in relation to groups identified in the Equality Act (or groups who are particularly disadvantaged with respect to the topic under consideration) should be addressed. These issues should be identified during the topic selection and scope development phases. They should also be considered during development of the review questions.

3.7 References and further reading

- 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 and Community Health* 58: 794–800
- Kelly MP, Swann C, Morgan A et al. (2002) *Methodological problems in constructing the evidence base in public health*. London: Health Development Agency
- Lomas J, Culyer T, McCutcheon C et al. (2005) *Conceptualizing and combining evidence for health system guidance: final report*. Ottawa: Canadian Health Services Research Foundation
- 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: 804–8
- 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: 886–92
- Petticrew M (2003) Why certain systematic reviews reach uncertain conclusions. *British Medical Journal* 326: 756–8
- Petticrew M, Roberts H (2002) Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57: 527–9
- Popay J, Rogers A, Williams G (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* 8: 341–51
- Popay J, editor (2006) *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
- Tannahill A (2008) Beyond evidence – to ethics: a decision making framework for health promotion, public health and health improvement. *Health Promotion International* 23 (4): 380–90
- Victora C, Habicht J, Bryce J (2004) Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health* 94: 400–5
- Woolgar S (1988) *Science: the very idea*. London: Routledge

4 Identifying the evidence

4.1 Introduction

This chapter describes how evidence is identified for reviews of public health information. It outlines the stages of developing a robust, transparent search protocol and planning a search strategy. It also identifies a variety of sources to search and additional methods of finding relevant information.

The Centre for Public Health Excellence (CPHE) encourages the use of search methods that balance precision and sensitivity. The aim is to identify the best available evidence to address a particular public health question, without producing an unmanageable volume of results. This involves a forensic search which includes:

- creating precise search questions and identifying the study types required to answer those questions
- matching key databases to the questions being asked (and not necessarily trawling all available databases just because they exist)
- adopting a pragmatic and flexible approach which allows a continual review of how best to find evidence and where
- having an understanding of the existing evidence base.

For the technicalities involved in developing a search strategy read 'Unit six: finding the evidence in systematic reviews of health promotion and public health interventions' (available from www.ph.cochrane.org/en/authors.html). For technical advice on developing a search strategy for systematic reviews, read section 6.4 of the 'Cochrane handbook for systematic reviews of interventions' (Lefebvre et al. 2008).

4.1.1 Issues to consider when searching for public health literature

Public health encompasses a wide range of disciplines including health, economics, psychology and the social sciences. Literature searches for public health evidence are not always straightforward:

- Searches are often long and complex and can present a technical challenge because of the databases being searched.
- Search strategies have their limitations. Public health information resources do not use a standard indexing vocabulary or thesaurus and the thesauruses used by

clinical databases only cover a limited number of public health concepts. The use of natural language varies and studies, outcomes, measures and populations are not described in a consistent way.

- The broad multi-disciplinary nature of public health means that searches are carried out across a wide range of databases – currently, there are no dedicated national databases that bring this information together⁵.
- There is a lack of particular types of evidence, such as controlled trials. This limits the methodological coverage of systematic reviews if the review process follows the imperatives of evidence-based clinical medicine.

The CPHE project team will support innovative and flexible approaches to searching, as it is not possible to know in advance where the best available evidence or interventions of interest are likely to be located. The aim is to retrieve a relevant and manageable set of results. The logic model developed when the initial briefing paper was drawn up (or when the scope was developed) should inform the approach taken.

All search processes should be transparent, clearly documented and reproducible. They should also be agreed with the CPHE project team.

The search process itself should be as comprehensive as possible, bearing in mind time and resource limitations. The type of search strategy used will depend on the logic model and the scope questions (see chapter 2). It will also depend on the type of review required: effectiveness, cost effectiveness, correlates and qualitative reviews all require different types of evidence.

4.1.2 Procedure

The search stage consists of three phases:

- constructing the search protocol
- developing the search strategy
- gathering the evidence, conducting searches and documenting the process.

About 1 week should be allocated for each phase. The following sections describe the stages and processes for each phase. The principles set out below will be the same, whether the protocol identifies the need for an exhaustive or a more flexible search.

⁵ Previous examples of dedicated resources include Healthpromis and EB2000 (managed by the former Health Development Agency) – these are no longer maintained.

4.2 Phase one: constructing the search protocol

4.2.1 Introduction

The review team should develop a search protocol from the review protocol and this should be agreed with the CPHE project team. The search protocol sets out how evidence will be identified and provides a basis for the review team to develop a detailed search strategy. Development of the search protocol is likely to be an iterative process involving the CPHE project team, the review team and information specialists within both organisations. The search protocol is normally added as an appendix to each review protocol (see chapter 3).

The search protocol should clearly state the:

- search approach taken (for example, systematic/exhaustive, emergent)
- rationale for the chosen approach
- search question(s) and key concepts
- electronic sources that will be searched (core, additional and economic databases plus any websites)
- plans for any additional searches (for example, citation or hand-searching) (see section 4.4.2)
- main study types that will be identified (for example, primary, review-level)
- inclusion and exclusion criteria
- restrictions (if any) on searches.

4.2.2 Electronic resources

Identifying public health evidence involves searching a wide range of electronic resources, in addition to the traditional scientific and clinical databases. The list of information sources should be individually tailored for each review to ensure they are relevant to the guidance topic. This list should be agreed between the review team and the CPHE project team.

Searches should include a mix of: core databases, subject-specific databases and other resources, depending on the subject of the research question and the level of evidence sought.

4.2.2.1 Databases

The databases searched must be relevant to the public health topic in terms of their coverage and content. Where there are a large number of possibilities, it would be expedient to prioritise those most likely to produce relevant evidence. (For example, MEDLINE is unlikely to be a useful source of information for a review of personal, health and social education in schools.) Databases cover different time periods, index different journals, use different subject headings and provide different amounts of bibliographic information. Consequently, each has its strengths and weaknesses. A list of core and topic-specific sources can be found in appendix B.

4.2.2.2 Websites

Websites can be a useful source of grey literature for public health reviews, particularly as a search of traditional, peer-reviewed literature may not produce much information. A list of core websites for consideration can be found in appendix B. These can be supplemented by information from topic-specific websites. Careful selection of websites is required to ensure that the type of evidence available is likely to be relevant: finding relevant data is more important than doing an exhaustive search.

4.2.3 Restrictions

NICE public health reviews are usually restricted to English language studies.

The start date for searches is determined by the nature of the evidence base and the time available to process data (the rationale needs to be documented). The NICE process allows stakeholders to identify whether any restrictions would exclude important data.

Study-type limits or filters should be used with caution, due to the broad nature of public health evidence and the fact that the:

- majority of sociological and social science databases do not provide adequate indexing by study design
- quality of indexing for – and the vocabulary used in – study methodologies and designs varies extensively and, in some instances, is poor.

4.3 Phase two: developing the search strategy

During this phase, the review team ‘translates’ the concepts from the search protocol, including all the synonyms that will be used (thesaurus terms and free-text/keywords) into a search strategy. Development of the search strategy is an iterative process between the review team and the CPHE project team. The latter will agree its final form.

Support and advice are available from the appointed NICE information services team lead.

4.3.1 Health inequalities and searches

It is essential that the list of synonyms used is exhaustive and that they describe a homogenous group or setting. They should not describe characteristics of subgroups, as this increases the risk of omitting a group (being inequitable or discriminatory). This detail should only be included by default. For example, if the population group is 'older people' a search for 'older people' should pick up subpopulations such as 'disabled older people' or 'black and minority ethnic older people'. Similarly, if the setting is 'communities and religious places', the search terms should cover relevant faith settings (such as 'church', 'temple' and 'mosque') not characteristics of the settings, such as whether the churches are gothic or neo-classical.

4.3.2 Sensitivity versus precision

The strategy needs to balance sensitivity (ability to identify relevant information) and precision (ability to exclude irrelevant documents). However, the need for an exhaustive search (involving additional resources) also needs to be balanced against a more modest search which may miss some studies. The balance will depend on the nature of the review questions and the available evidence.

4.4 Phase three: gathering the evidence, conducting searches and documenting the process

During this phase, the review team translates the search strategy (as necessary) for use with various databases. The results should be downloaded into 'Reference manager' or 'Endnote' (or other reference management software). Items which cannot be downloaded into bibliographic software can be recorded in a Word document or spreadsheet.

4.4.1 Economic searches

The search for economic evidence should identify papers that are most relevant to public sector practice and hence likely to inform PHIAC/the PDG decision-making. A systematic search for economic evaluations should be undertaken in NHS EED (economic evaluation database) and Econ Lit (as appropriate). Other databases and websites can be searched using appropriate economic filters, where these are available. Economic evidence can also be identified when sifting effectiveness or qualitative search results.

It is advisable to simplify the agreed search strategy for the economic searches because a complex search may exclude relevant studies. For example, instead of searching for population group **and** setting **and** intervention **and** the problem, it might be more reliable to just search for the public health problem. If this produces too many results, then additional concepts can be added.

4.4.2 Additional searches

If the main searches have not retrieved enough relevant material and the search needs to be widened, the review team may carry out additional types of searches. These could include: 'snowballing' to find citations, a search of the grey literature, journal hand-searches or making contact with experts and stakeholders.

4.4.2.1 Citations using 'snowballing'

A search can be usefully extended by looking for articles which cite other, more specific articles containing additional relevant references. However, it depends on whether the database software can perform this search; even if it is possible, such a search will only retrieve cited articles from journals indexed in the same database.

4.4.2.2 Grey literature

A search of the 'grey literature' can help identify material that will not be picked up by mainstream sources (such as the MEDLINE database).

This type of search can be difficult and time consuming: both a database and an Internet search may be necessary. It is essential to be clear about the type of material needed. In particular, it is useful to distinguish between data that might supplement the effectiveness literature (for example, ongoing evaluative research) and information that could aid implementation.

4.4.2.3 Hand-searching

Hand-searching involves a manual search through the contents tables of selected journal titles for relevant articles. There is no requirement to do this and it can be time consuming. However, it is worth doing if the reviewers are aware of any relevant journal titles that are not included in the bibliographic databases being searched. Hand-searching can also be worthwhile if the database searches have failed to retrieve much relevant evidence (though it should be limited to a few relevant, specialised journals). Bibliographic details of any studies identified should be added manually to the database of references which have been downloaded.

Hand-searching is a useful way of identifying experts who may be able to identify other relevant research (see section below).

4.4.2.4 Contacting experts

Ongoing research may be needed to inform PHIAc/the PDG of important studies likely to be published or completed during the development – or soon after publication – of the public health guidance. Some types of research, notably intervention trials, are often documented in databases of ongoing research. However, these are not always up-to-date and it is advisable to ask experts in the area.

Experts can be identified and contacted via research networks, relevant journal abstracts or via relevant reference lists. PHIAc/PDG members may also be able to help; try to give them advance notice, taking into account the time involved and their availability.

Any additional evidence received should be entered into the bibliographic database. The number of articles identified by this means must be specified in the methods section of the review.

4.4.2.5 Using review-level material to identify primary studies

Review-level material (for example, systematic reviews, literature reviews and meta-analyses) may provide an additional source of primary studies. Relevant reviews can be identified using an appropriate checklist. The reference lists in the reviews can be used to identify potentially relevant primary studies.

The Centre for Reviews and Dissemination (CRD), Cochrane and Campbell databases are useful sources of robust, quality reviews.

4.4.3 Documenting the search process

The search process needs to be transparent and replicable. For these reasons, as well as to aid quality assurance, it is important to document it. The review team should save the following files once the searches are complete:

- Word document containing the search strategies for each resource searched. Each strategy should include audit information, as shown in appendix C. The document should clearly state the title of the review and the review type (that is, effectiveness/qualitative/economic).
- Final de-duplicated 'Reference manager' database.
- Word document of other results (for those records which cannot be downloaded into 'Reference manager', for example, website results).

4.4.4 Submission of evidence from stakeholders

For some review questions, members of PHIAC/the PDG and the CPHE project team may have good reason to believe that information exists, even though it has not been found using the standard searches. For example, they may be aware of:

- ongoing research (if an intervention is relatively new)
- studies that have only been published as abstracts
- data on adverse effects relevant to the interventions being studied
- economic models
- studies of public or professional views and experiences.

In these situations, the CPHE project team may ask stakeholders for evidence (these calls for evidence may be made at any point during development of the guidance). The team will specify the question being addressed, along with details of the type of evidence being sought. For example, for questions of effectiveness, details of the intervention, participants, comparisons, outcome and study design may be needed.

Stakeholders are usually given 4 weeks to respond.

4.4.5 Confidential information

In addition to published studies, stakeholders may submit unpublished data or studies in response to a call for evidence. They will be asked to complete a checklist that identifies the location of any confidential information contained in their submission. Box 4.1 summarises what may be considered confidential by NICE.

Box 4.1 Information on what may be considered confidential

- Information that may influence share price values ('commercial in confidence') or is deemed intellectual property (that is, awaiting publication) is deemed confidential
- The relevant part of a sentence, a particular result from a table or a section of code is deemed confidential (that is, information deemed confidential should be kept to an absolute minimum)
- NICE will not agree to a whole study being designated confidential. At a minimum, a structured abstract of the study or economic model will have to be made available for public disclosure during consultation on the guidance
- Results derived from calculations using confidential data are not considered confidential unless releasing those results would enable back-calculation to the original data

Stakeholders should also indicate on the submission itself which part contains the confidential information. For example, they could use a highlighter pen on a paper copy or the highlighter function in an electronic version. These markings should be maintained throughout the development process. When the draft and final versions of the guidance are prepared for publication, the CPHE project team should ensure that these sections are replaced by a note stating that confidential information has been removed, so that the public is aware of exactly where confidential data have been used. These criteria also apply to 'academic in confidence' data.

Following the principles outlined in box 4.1, the amount of confidential information should be kept to an absolute minimum and, as a minimum, a summary of this information should be publicly available by the time of consultation on the guidance. NICE needs to be able to justify the recommendations made in its public health guidance on the basis of the evidence considered by PHIAC/the PDG. To this end, NICE will work with the data owners to find an agreed balance between confidentiality and transparency (for example, see www.nice.org.uk/229411).

4.4.6 Documenting evidence from stakeholders

The CPHE project team should enter information received from stakeholders into a table or bibliographic database (see section 4.4.4). Details should be cross-checked against evidence identified through the database searches. It should be assessed in the same way as evidence identified from published studies (see chapter 5).

4.5 Equality and diversity

All searches should be inclusive, capturing evidence related to all groups identified in the Equality Act (or to groups that are particularly disadvantaged with respect to the topic under consideration). Search strategies should only be narrowed to specific groups if these have been specified during the topic or scoping development phases.

4.6 References

Lefebvre C, Manheimer E, Glanville J (2008) Searching for studies In: Higgins JPT, Green S editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley. Available from www.cochrane-handbook.org

5 Reviewing the scientific evidence

5.1 Introduction

This chapter describes how 'evidence reviews' are produced by NICE: typically there are two for guidance developed using the intervention process and five if the programme process has been used.

As outlined in chapter 3, different types of evidence can be used to answer different research questions. For example:

- experimental studies (such as controlled trials) and observational studies (such as before-and-after studies) relate to the effectiveness of interventions
- correlation studies look at the relationship between exposure to particular factor(s) and an outcome of interest
- qualitative research studies (such as interviews or focus groups) examine the views of the target populations.

Reviewing is an explicit, systematic and transparent process that can be applied to qualitative and quantitative (experimental and observational) evidence (see chapter 3). The review team should follow the procedures outlined below to the best of its ability. However, at certain points, evidence reviews may well differ and flexibility may be needed (for example, to determine the strength of the evidence or the way evidence statements should be organised).

Further, as the presentation and interpretation of different types of evidence will always involve some degree of expert judgement, the review process relies on the expertise of the review team, any expert advisers they may consult and the CPHE project team. (For the appraisal and presentation of economic evidence, see chapter 6.)

Standard systematic review methodologies (for example, those used by Cochrane reviewers) prescribe exhaustive and thorough processes, normally relating to efficacy and effectiveness and often taking years to complete. This approach focuses on the precision and reliability of measurements used in the original science and tends to emphasise the limits of the evidence. However, while it is important to be aware of these limits, the process of interpretation is equally important.

NICE public health evidence reviews need to summarise and interpret evidence, in spite of its limitations, so that PHIAC/the PDG can make recommendations in areas of uncertainty. Rarely would it be necessary to undertake a full systematic review (and anyway, time and resource constraints would make this difficult). The key point is that

the evidence reviewed has to be good enough for the advisory committees to be able to make decisions about their recommendations.

All NICE evidence reviews (except mapping reviews) involve the following steps:

1. Select the relevant evidence.
2. Assess its quality.
3. Extract, synthesise and present it.
4. Derive evidence statements.
5. Assess its applicability.

(NICE mapping reviews may use purposive sampling to select data and do not necessarily assess its quality.)

5.2 Selecting relevant evidence

This section applies to both qualitative and quantitative evidence reviews.

Identifying and selecting all relevant studies is a critical stage in the evidence review process (see chapter 4 for identifying evidence). Prior to undertaking screening, the review team should discuss and work through examples of studies meeting the inclusion criteria (as set out in the agreed review protocol) to ensure a high degree of inter-rater reliability. Then studies meeting the inclusion criteria should be selected using the two-stage screening approach below:

1. Title/abstract screening: titles/abstracts should normally be screened independently by two reviewers (that is, they should be double-screened) using the parameters set out in the review protocol. Where reviewers disagree about a study's relevance, this should be resolved by discussion or by recourse to a third reviewer. If, after discussion, there is still doubt about whether or not the study meets the inclusion criteria, it should be retained.
2. Full-paper screening: once title/abstract screening is complete, the review team should assess full-paper copies of the selected studies, using a full-paper screening tool developed for this purpose. This should normally be done independently by two people (that is, the studies should be double-screened). Any differences should be resolved by discussion between the two reviewers or by recourse to a third reviewer.

The study selection process should be clearly documented and include details of the inclusion criteria. A flow chart should be used to summarise the number of papers

included and excluded at each stage of the process and this should be presented in the report. Each study excluded at the full-paper screening stage should be listed in the appendix of the review, along with the reason for its exclusion.

5.3 Assessing the quality of the evidence

5.3.1 Introduction

This section applies to the assessment of both qualitative and quantitative evidence (for the appraisal and presentation of economic evidence, see chapter 6).

The review team should assess the quality of evidence selected for inclusion in the review using the appropriate quality appraisal checklist (see section 5.3.2). This is a key stage in the guidance development process, since the quality rating of studies will be reflected in the evidence statements (see section 5.5). These, in turn, inform the recommendations (along with other factors and considerations [see section 7.2]).

Some of the more commonly used study types and their abbreviations are listed below:

Quantitative studies: experimental

- Before-and-after study.
- Non-randomised controlled trial (NRCT).
- Randomised controlled trial (RCT).

Quantitative studies: observational

- Before-and-after study.
- Case-control study.
- Cohort study.
- Correlation study.
- Cross-sectional study.
- Interrupted time series (ITS).

Qualitative studies

- Document analysis.
- Focus groups.

- Interview study.
- Observation and participant observation.

Economic studies

- Cost–benefit analysis.
- Cost–consequence analysis.
- Cost-effectiveness analysis.
- Cost–utility analysis.

The internal and external validity of quantitative studies should be assessed using either the quality appraisal checklist for intervention studies (see appendix F) or that for correlation studies (see appendix G). Appendix E includes an algorithm for identifying the quantitative study type and this terminology should replace any provided by the study author.

Qualitative studies should be assessed using the checklist in appendix H. This is important to maintain an audit trail. However, it is acknowledged that the concept of validity in qualitative research is less clearly defined than for quantitative research. As a result, the review team may wish to take account of other factors when judging the ‘trustworthiness’ of the study, its relevance to the research questions and how ‘convincing’ the results are. These factors should be clearly described in the review.

Some studies, particularly those using mixed methods, may report quantitative, qualitative and economic outcomes. In such cases, each aspect of the study should be separately assessed using the appropriate checklist. Similarly, a study may assess the effectiveness of an intervention using different outcome measures, some of which will be more reliable than others (for example, self-reported smoking versus a measure of plasma cotinine levels). In such cases, the study might be rated differently for each outcome, depending on the reliability of the measures used. For further information on how to integrate evidence from qualitative and quantitative studies, see Dixon-Woods et al. (2002).

5.3.2 Quality assessment

Quality assessment is a critical stage of the evidence review process. Prior to undertaking the assessment, the review team should discuss and work through some of the studies to ensure there is a high degree of inter-rater reliability.

Each full paper should be assessed by one reviewer and checked for accuracy by another. Periodically throughout the process, a random selection should be considered

independently by two people (that is, double-assessed). The size of the sample will vary from review to review, but a minimum of 10% of the studies should be double-assessed. Any differences in quality grading should be resolved by discussion or recourse to a third reviewer.

The composite inter-rater reliability scores should be reported, preferably as a kappa statistic and noting if it is good (between 0.60 and 0.74) or excellent (above 0.75) (see Cochrane Collaboration 2008). If the inter-rater reliability score is below 0.60, the reasons for digression should be explored and a course of action agreed.

5.3.2.1 Internal validity

The review team should use the relevant quality appraisal checklist to assess a study's internal validity: that is, to check if potential sources of bias have been minimised and to determine if its conclusions are open to any degree of doubt. Each study should be rated ('++', '+' or '-') to indicate its quality:

Quality rating

- ++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
- + Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

If a study is not assigned a '++' quality rating, the review team should record the key reasons why this is the case in the quality appraisal checklist comments column, alongside the overall quality rating. They should also record these reasons in the evidence table 'notes' column under 'Limitations identified by review team' (see appendix K) and highlight them in the narrative summary.

5.3.2.2 External validity

The review team should also use the quality appraisal checklist to assess the external validity of quantitative studies: the extent to which the findings for the study participants are generalisable to the whole 'source population' (that is, the population they were chosen from). This involves assessing the extent to which study participants are representative of the source population. It may also involve an assessment of the extent to which, if the study were replicated in a different setting but with similar population parameters, the results would have been the same or similar. If the study includes an 'intervention', then it will also be assessed to see if it would be feasible in settings other than the one initially investigated.

Studies should be given a separate rating for external validity (++, + or –) prefixed with 'EV' (external validity).

Qualitative studies do not seek to generalise specific findings and do not need to be assessed for external validity.

External validity is different to 'applicability' (see section 5.6).

5.3.2.3 Unpublished data, studies in progress and grey literature

Reviewers are not expected to search the grey literature or unpublished data as a matter of routine. However, if time and resources allow (or if the grey literature is particularly relevant), the review team may obtain such papers, particularly from stakeholders and experts in the topic area. Any unpublished data which the authors intend to publish as peer-reviewed literature should be quality-assessed in the same way as published studies. Ideally, if additional information is needed to complete the quality appraisal checklist, the authors should be contacted.

Grey literature may be assessed similarly, although this is not always appropriate. Where the grey literature has important insights to convey these should be reported in a manner to be agreed with the CPHE project team.

5.4 Extracting, synthesising and presenting the evidence

This section describes how to present data and develop related evidence statements for both qualitative and quantitative evidence reviews.

Any expert or value judgements that have been made (including expert advice from third parties) should be reported in the review.

Both qualitative and quantitative evidence reviews should incorporate narrative summaries of, and evidence tables for, all studies. Concise detail should be given (where appropriate) on: populations, interventions, settings, outcomes, measures and effects. This includes identifying any similarities and differences between studies, for example, in terms of the study population, interventions and outcome measures.

The summaries and evidence tables should be produced using the quality appraisal checklists for each study (see section 5.3.2) and original papers/reports.

Review authors should refer to 'Writing for NICE' and the 'NICE style guide' (noting that the Harvard referencing system is preferred). Both are available from the CPHE project team.

5.4.1 Data extraction and evidence tables

The evidence tables (see appendix K) can also be used as data extraction templates for quantitative (intervention and correlation studies), qualitative and economic studies and review-level material.

Each evidence review should include one main evidence table containing summaries of all the studies used in alphabetical order (by first author). If a review includes different types of studies (that is, quantitative, qualitative and economic), these may also be listed in separate evidence tables within the review. The review team should discuss any substantial changes to the structure of these tables with the CPHE project team.

Evidence tables can help determine whether it is possible to calculate a summary estimate of effect, if applicable (see section 5.4.4.2).

5.4.1.1 Evidence tables for quantitative studies

The quantitative evidence table template shows the type of data that should be included for quantitative studies (both experimental and observational) (see appendix K).

Concise details (sometimes in bullet point or another list form) should be given on: bibliography (authors, date); study aim and type (for example, RCT, case-control); population (source, eligible and selected); intervention, if applicable (content, intervener, duration, method/mode/timing of delivery); method of allocation to study group (if applicable); outcomes (primary and secondary and whether measures were objective, subjective or otherwise validated); and key findings (including effect sizes, confidence intervals and their significance, for all relevant outcomes).

Where given, exact p-values (whether or not significant) and confidence intervals must be reported, as should the test from which they were obtained. Where p-values are not given, any descriptive statistics indicating the direction of the difference between intervention and comparator should be presented. If no further statistical information is available, this should be clearly stated.

The quality ratings of the study's internal and external validity should also be given (see section 5.3.2). Where study details are not reported (or not applicable), this should be clearly stated.

5.4.1.2 Evidence tables for qualitative studies

The qualitative evidence table template shows the type of data that should be included for qualitative studies (see appendix K). Concise details should be given on: bibliography (authors, date); location (for example, UK); funding details (if known); population/participants; study design; theoretical perspective; key aims, objectives and research

questions; methods (including analytic and data collection technique); key themes/findings (including quotes from participants that illustrate these themes/findings, if appropriate); gaps and limitations; conclusions; and the study's quality rating.

5.4.2 Narrative summaries of quantitative or qualitative studies

The narrative summary provides an opportunity to place a study and its findings in context. It should highlight key factors influencing the results observed, an interpretation of the results and more on the detail presented in the evidence tables (see section 5.4.1). Each narrative summary should include:

- A brief description of the study design, methodology, population, setting and research questions/outcomes (if appropriate) for all relevant studies.
- A summary of the key findings.
- A summary of the quality ratings (expanding, where appropriate, on study strengths and weaknesses), applicability issues and any other relevant contextual points.
- Commentary on the scale and nature of the evidence base may also be useful.

The narrative summary should conclude with a short discussion, followed by one or more evidence statements. These should reflect the key findings, the quantity, quality and consistency of the evidence, and its applicability to the research question (including its applicability to the target population).

Narrative summaries of all studies and interventions should be incorporated within the main findings of the evidence review. They should be organised by research question and could be divided into smaller subcategories, such as outcome measure, setting or subpopulation.

5.4.3 Summary tables

If appropriate, short summary tables can be included with the main findings (usually preceding an evidence statement) or in the appendices. For example, these might:

- summarise the information gleaned for different research questions
- summarise the study types, populations, interventions, settings or outcomes for each study related to a particular research question
- organise and summarise studies related to different outcomes.

5.4.4 Other presentations of quantitative data

There are a range of ways to summarise and illustrate the strength and direction of quantitative evidence about the effectiveness of an intervention. Some of the most commonly used methods are described below, although this is not an exhaustive list: the review team should discuss the form any data presentations might take with the CPHE project team.

5.4.4.1 Graphical presentation

Results from relevant studies (whether statistically significant or not) can be presented graphically.

Forest plots should be used to show effect estimates and confidence intervals for each study (when available, or when it is possible to calculate them). If possible, they should be used even when it is not appropriate to do a meta-analysis and present a pooled estimate (see section 5.4.4.2). However, the homogeneity of the outcomes and measures in the studies needs to be carefully considered: the forest plot needs data derived from the same (or justifiably similar) outcomes and measures.

If a forest plot is not appropriate, other graphical forms may be used (for example, a harvest plot [Ogilvie et al. 2008]).

When outcome measures vary between studies, it may be appropriate to present separate summary graphs for each outcome. However, if outcomes can be transformed on to a common scale by making further assumptions, an integrated (graphical) summary would be helpful. In such cases, the basis (and assumptions) used should be clearly stated and the results obtained in this way should be clearly indicated.

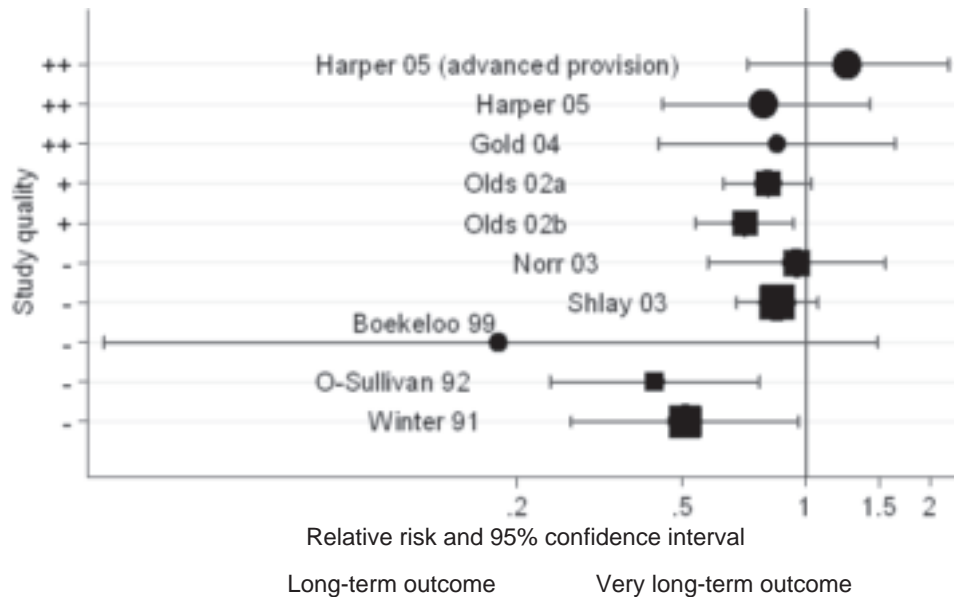
On any graph, the order of entries, symbols, line types and brief text may all be used to illustrate the study results. Sometimes, more than one graph may be needed to avoid undue complexity. If evidence from a meta-analysis is being presented, it is often appropriate to plot the pooled estimate and its confidence interval.

Figure 5.1 was drawn using the Stata® statistical package⁶ to plot symbols at the relative risk estimates, and lines ('error bars') between the corresponding upper and lower 95% confidence intervals.

RCT results (which can be expressed as relative risks) are displayed in order of decreasing study quality (and, within that, by publication date). This helps identify any links between effect-estimate and study quality. Different symbols are used to distinguish long-term from very long-term outcomes.

⁶ Many other statistical and data presentation packages could be used to draw comparable summary graphs.

Figure 5.1 Graphical presentation of quantitative evidence of effectiveness: effectiveness of 1-1 interventions on under-18 pregnancies



In this example, the symbol size has been used to give visual emphasis to the larger studies. Other dimensions of interest (such as standardised versus self-reported measures, or the country where the study was set) could also be represented in this or a supplementary graph.

5.4.4.2 Conducting and presenting a meta-analysis⁷

Meta-analysis data may be used to produce a graph if the data (usually from RCTs) are sufficiently homogenous and if there are enough relevant and valid data from comparable (or the same) outcome measures. Where such data are not available, the synthesis may have to be restricted to a narrative overview of individual studies looking at the same question. In such cases, a forest plot (see section 5.4.4.1) is one useful way of illustrating the results.

The characteristics and limitations of the data in the meta-analysis should be fully reported (for example, in relation to the population, intervention, setting, sample size and validity of the evidence).

Before pooling or combining the results of different studies, the degree of heterogeneity in the data should be assessed to determine how the results have been affected by the circumstances in which studies were carried out. The results of any homogeneity tests should be reported.

⁷Note that there is free Cochrane software called Review Manager, used for meta-analyses (www.cc-ims.net/RevMan).

Statistical heterogeneity can be addressed using a random (as opposed to fixed) effects model. The impact of known research heterogeneity (for example, population characteristics or the intensity or frequency of an intervention) can be managed using 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 relevant data, even if details 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). Sensitivity analysis can be used to investigate the impact of missing data.

Publication bias (studies, particularly small studies, are more likely to be published if they include statistically significant or interesting results) should be critically assessed and reported in the interpretation of the meta-analysis results. It may be helpful to inspect funnel plots for asymmetry to identify any publication bias (see www.cochrane-net.org/openlearning/HTML/mod15-3.htm; also Sutton et al. 2000).

Similarly, the possibility of selective reporting of outcomes (emphasising statistically significant results over others, for example) should be considered. In part, this can be done by examining which outcomes were described as primary and secondary in study reports or protocols.

A full description of data synthesis, including meta-analysis and extraction methods, is available in 'Undertaking systematic reviews of research on effectiveness' (NHS Centre for Review and Dissemination 2001).

5.4.5 Other presentations of qualitative data based on analytic and structured techniques

The nature of qualitative evidence is such that it is unhelpful to set a prescriptive method for its synthesis and description. Qualitative evidence occurs in many forms and formats. This section includes some of the methods that may be used to synthesise and present it. As with all data synthesis, the key is transparency. It is important that PHAC/the PDG and stakeholders can easily follow the method used. It should be written up in clear English and any analytic decisions should be clearly justified.

In some cases, the evidence may be synthesised and then summarised. In other cases, a narrative description may be adequate. The approach used depends on the volume and consistency of the evidence. If the qualitative literature is extensive, then a synthetic approach is preferable. If the evidence is more disparate and sparse, a descriptive approach may be more appropriate.

5.4.5.1 Reporting sparse, disparate qualitative evidence

In many cases, qualitative reviews will comprise relatively few papers compared to quantitative reviews and often their focus will be inconsistent (for example, they may involve different settings/populations/interventions). If the papers have little in common it would not be appropriate to synthesise them. Instead, the authors of the review should provide a narrative description of the key themes (including illustrative quotes) of each paper. They should also provide a quality appraisal and brief general description of each study (for example, describing the methods used, the participants involved and the underlying rationale).

Both the narrative summary and the evidence table should identify all the main themes reported: only themes that are not relevant to the review at hand should be left out and these omissions should be clearly documented. As in all qualitative research, particular attention should be paid to 'outliers' (other themes) and views that disagree with or contradict the main body of research.

The narrative description should be divided up under headings derived from the research question (for example, the settings of interest) unless good reasons are documented for not doing so. It should be summarised into evidence statements which note areas of agreement and contradiction (see section 5.5).

5.4.5.2 Synthesising qualitative evidence

The simplest and most rigorous approach to presenting qualitative data in a meaningful way is to analyse the themes (or 'meta' themes) in the evidence tables and write a narrative based on them. This 'second level' thematic analysis can be carried out if ample data are found and the papers and research reports cover the same (or similar) factors. (These should be relevant to the research question(s) and could, for example, include intervention/age/group/setting.)

It can be carried out in one of two ways. More simply, papers reporting on the same phenomena can be grouped together to compare and contrast themes, focusing not just on consistency but also on any differences. The narrative should be based on these themes.

A more complex but useful approach is 'conceptual mapping' (see Johnson et al. 2000). This involves identifying the key themes and concepts across all the evidence tables and grouping them into first level (major), second level (associated) and third level (subthemes) themes. Results are presented in schematic form as a conceptual diagram and the narrative is based on the structure of the diagram.

Alternatively, themes can be identified and extracted from the data itself, using a grounded approach (see Glaser and Strauss 1967). Other potential techniques

include meta-ethnography (see Noblit and Hare 1988) and meta-synthesis (see Barroso 2000).

5.4.5.3 Reporting 'bias' or variation

Any review or, in particular, any synthesis of qualitative data must, by its nature, mask some of the variations considered important by qualitative researchers (for example, the way the researcher interacts with research participants when gathering data). Reviewers should, as far as possible, highlight any significant causes of variation noted during data extraction.

5.5 Deriving evidence statements

5.5.1 Introduction

This section applies to both qualitative and quantitative reviews. As described in section 5.4.2, each evidence review should include a narrative summary and should conclude with a short discussion and one or more supporting evidence statements.

The evidence statements should reflect the strength (quality, quantity and consistency) of the evidence and make a statement about its applicability. They can also highlight a lack of evidence. They should provide an aggregated summary of the evidence (from one or more studies) in relation to a key question or issue. In the case of intervention studies, they should also reflect what is plausible, given the evidence available about what has worked in similar circumstances.

They are structured and written to help PHIAC/the PDG formulate and prioritise recommendations. They help it decide:

- whether or not there is sufficient evidence (in terms of strength and applicability) to form a judgement
- where relevant, whether (on balance) the evidence demonstrates that an intervention or programme can be effective or is inconclusive
- where relevant, the typical size of effect (where there is one)
- whether the evidence is applicable to the target groups and contexts being covered by the guidance.

Evidence statements which support the recommendations should be included in the final guidance.

5.5.2 Structure and content of evidence statements

One or more evidence statements are prepared for each review research question or its subsidiary questions. (Subsidiary questions may cover a type of intervention, specific population groups, a setting or an outcome.)

Once it has all the data, the review team should discuss with the CPHE project team how it intends to 'group' the evidence. For example, it could be grouped according to the similarity of the populations, interventions and outcomes covered in the studies. However, the decision will be highly context-specific and will depend on the amount, breadth and depth of evidence. The review team should avoid developing a separate evidence statement for each study while, at the same time, not grouping so many studies together that the evidence statements become too generic and therefore meaningless.

The evidence statements could comprise an overarching summary statement supported by various subsidiary statements. They should provide a clear, self-contained summary.

They should refer to the sources of evidence (study type and references) and their quality in brief descriptive terms and not just by acronyms. In addition, each statement should include summary information about the:

- **content** of the intervention, if applicable (for example, what, how, where?)
- **strength** of evidence (reflecting the appropriateness of the study design to answer the question and the quality, quantity and consistency of evidence)
- **direction and size of effect** (if applicable) both positive and negative
- **applicability** to the question/target population/setting (see section 5.6).

Note that the strength of the evidence and the direction and strength of the effects or correlations observed (if applicable) are reported separately.

5.5.3 Evidence statement terminology

Terms that describe the strength of the evidence should be used consistently within each review and their definitions should be reported in its methodology section. A set of standardised terms is given below. However, the evidence base for each review may vary, so the review team should define how these terms have been used.

5.5.3.1 Strength of evidence

The overall **strength** (quality, quantity and consistency) of the evidence may be summarised as:

- **No evidence** Be clear about the sources and inclusion criteria. For example, state: 'No evidence was found from English-language trials published since 1990...'
- **Weak evidence** For example, 'There was weak evidence from one before-and-after study (Jones 1990 [-])'.
- **Moderate evidence** For example, 'There was moderate evidence from two case-control studies (Smith 1992 [+]; Brown 1995 [+])'.
- **Strong evidence** For example, 'There was strong evidence from three randomised controlled trials (Green 2002 [++]; Black 2005 [++]; Blue 2000 [++])'.
- **Inconsistent evidence** Further commentary may be needed on the variability of findings in different studies. For example, when the results of (++) or (+) quality studies do not agree. In such cases, the review team may qualify an evidence statement with an explanatory sentence or section that gives more detail.

'Vote counting' (merely reporting on the number of studies yielding significant effects) is **not** an acceptable summary of the evidence.

5.5.3.2 Direction and size of effect

If appropriate, the **direction of effect** (impact) or **correlation** should be summarised using one of the following terms:

- positive
- negative
- mixed
- non-existent.

However, appropriate context/topic-specific terms (for example, 'an increase in HIV incidence', 'a reduction in injecting drug use' and 'smoking cessation') may be used.

If appropriate, the **size of effect** (impact) or **correlation** and, wherever possible, the degree of uncertainty involved, should be reported using the scale applied in the relevant study. For example, an odds ratio (OR) or relative risk (RR) with confidence interval (CI), or a standardised effect size and its standard error, may be quoted. Where an estimate cannot be explained, every effort should be made to relate it to interpretable criteria

or conventional public health measures. If it is not possible to provide figures for each study, or if there are too many studies to make this feasible, the **size of effect** or **correlation** can be summarised using the following standardised terms:

- small
- medium
- large.

These terms should be used consistently within each review and their definitions should be reported in its methodology section.

5.5.4 Quantitative evidence statements⁸

An example of an evidence statement about the effectiveness of an intervention is presented below:

There is strong evidence from four studies to suggest that educational interventions delivered by youth workers may reduce the incidence of hazardous drinking by young people. Two randomised controlled trials (RCTs) and one non-randomised controlled trial (NRCT) showed reduced risk (RR [95% confidence interval]) in the intervention group: 0.75 (0.58–0.94) (Jelley et al. 2004 RCT [++]); 0.66 (0.57–0.78) (Lake et al. 2003 RCT [++]); 0.42 (0.18–0.84) (Wagner et al. NRCT [+]). Another RCT showed reduced risk but was not statistically significant: 0.96 (0.84–1.09) (Blake et al. RCT [+]). However, one NRCT found increased risk of binge drinking in the intervention group: 1.40 (1.21–1.74) (Jensen et al. [–]).

The following is an example of an evidence statement from a 'correlates review':

There is moderate evidence from three cross-sectional studies about the correlation between young people's communication skills with safer sex and a reduction in the number of teenage pregnancies. The evidence about the strength of this correlation is mixed. One study (Jones et al. 2007 [+]) found that discussing condom use with new partners was associated with actual condom use at first sex (OR 2.67 [95% CI 1.55-4.57]). Another study (Buston et al. 2007 [–]) found that not talking to a partner about protection before first sexual intercourse was associated with teenage pregnancy (OR 1.67 [1.03-2.72]). However, another study (DiLorio et al. 2000 [+]) found small correlations between condom use, discussions about safer sex ($r = 0.072$, $p < 0.01$) and communication skills ($r = 0.204$, $p < 0.01$).

⁸Note these examples have been adapted from the originals and are for illustrative purposes only.

5.5.5 Qualitative evidence statements

Evidence statements developed from qualitative data do not usually report on the impact an intervention has on behaviour or health outcomes, nor do they report statistical effects or aggregate measures of strength and effect size. They should summarise the evidence, its context and quality, and the consistency of key findings and themes across studies. Areas where there is little (or no) concurrence should also be summarised. For example:

Teenage mothers interviewed in a family planning clinic (Smith 1999 [+]) and teenage mothers who responded to a questionnaire at their GP surgery (Jones 2003 [++]) agreed that access to education was the thing that helped them most after they had their child. However, Van Dinkleholm (2004 [-]) did not report this as a key theme in her study of health visitor perceptions of teenage mothers' needs.

5.6 Assessing applicability

This section applies to both quantitative and qualitative reviews. It describes how the review team should assess the applicability of evidence in qualitative or quantitative reviews. Health economic data are assessed differently (see chapter 6).

PHIAC/the PDG needs to judge the extent to which the evidence reported in the reviews is applicable to the areas for which it is developing recommendations. The review team should assess each evidence statement to judge how similar the population(s), setting(s), intervention(s) and outcome(s) of the underpinning studies are to those outlined in the review question(s). The studies should be assessed as a whole – rather than assessing each one individually.

The following characteristics should be considered:

- **Population** Age, sex/gender, race/ethnicity, disability, sexual orientation/gender identity, religion/beliefs, socioeconomic status, health status (for example, severity of illness/disease), other characteristics specific to the topic area/review question(s).
- **Setting** Country, geographical context (for example, urban/rural), healthcare/delivery system, legislative, policy, cultural, socioeconomic and fiscal context, other characteristics specific to the topic area/review question(s).
- **Intervention** Feasibility (for example, in terms of health services/costs/reach), practicalities (for example, experience/training required), acceptability (for example, number of visits/adherence required), accessibility (for example, transport/outreach required), other characteristics specific to the topic area/review question(s).

- **Outcomes** Appropriate/relevant, follow-up periods, important health effects.

Following this assessment, the review team should categorise each evidence statement as:

- directly applicable
- partially applicable
- not applicable.

A statement detailing the category it falls into (and the reasons why) should appear at the end of the evidence statement. It should state: 'This evidence is (directly, partially or not) applicable because ...'. An example⁹ of an applicability statement is presented below:

This evidence is only partially applicable to people in the UK who inject drugs. That is because all these studies were conducted in countries where needles are mainly sold by pharmacies (USA, Russia and France), rather than freely distributed, as is the norm in the UK.

The review team should note that PHIAC/the PDG needs to judge the extent to which the evidence reported in the reviews is applicable to the areas/topics for which it is developing recommendations and it may ask for additional information on the applicability of the evidence. The review team should also be aware that PHIAC/the PDG will draw upon a wide range of information in reaching its final judgement.

Although similar issues are considered when assessing the applicability of health economic data, there are some important differences. Details can be found in section 6.2.2.

5.7 Published guidance

The review team should identify relevant published guidance (from NICE and other agencies) in its data search, as well as relevant NICE guidance in development.

5.7.1 NICE guidance

NICE guidance (public health or clinical) should be fully referenced and the evidence underpinning the recommendations left unchanged, provided it is not out of date. If there is new published evidence that would significantly alter the existing recommendations, the review team should bring this to the attention of the CPHE project team.

⁹Note this has been adapted from the original and is for illustrative purposes only.

The CPHE project team, in turn, should pass it to the relevant team within NICE so that it can consider whether or not to update the guidance.

5.7.2 Other guidance

Other relevant published guidance should be assessed for quality using the AGREE instrument (AGREE Collaboration 2003) to ensure it is sufficiently documented. PHAC/ the PDG should set the cut-off point for accepting or rejecting other guidance and this should be documented in the guidance appendices.

5.8 Equality and diversity

In the discussion section of the evidence reviews, the following questions should be considered.

5.8.1 Are the evidence-review criteria inclusive?

All relevant inequalities data should be included in the reviews. At the data extraction stage, reviewers are prompted to refer to the PROGRESS-Plus criteria (age, sex, sexual orientation, disability, ethnicity, religion, place of residence, occupation, education, socioeconomic position and social capital) (Oliver et al. 2008). Review inclusion and exclusion criteria should also take the relevant groups into account.

5.8.2 Have the relevant data been appropriately extracted and presented in the evidence statements?

Equalities evidence should be considered during the drafting of reviews. It should be included in the data extraction process and should appear in the summary evidence statements.

5.8.3 What is the state of the evidence base?

This question aims to identify if there are any gaps in the evidence in relation to inequalities. It also aims to identify if the evidence has uncovered gaps in the scope of the guidance in relation to inequalities.

5.9 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
- Barroso J (2000) Meta-synthesis of qualitative research on living with HIV infection. *Qualitative Health Research* 10 (3): 340–53
- Bowling A (2002) *Research methods in health: Investigating health and health services*. Buckingham: Open University Press
- Cochrane Collaboration (2008) *Cochrane handbook for systematic reviews of interventions*. Higgins JPT, Green S editors [online]. Available from www.cochrane-handbook.org/ [Accessed 16 March 2009]
- Dixon-Woods M, Agarwal S, Young B et al. (2004) *Integrative approaches to qualitative and quantitative evidence*. London: Health Development Agency. Available from www.nice.org.uk/page.aspx?o=502695
- 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*. Oxford: Oxford Medical Publications
- Eccles M, Mason J (2001) *How to develop cost-conscious guidelines*. Health Technology Assessment 5: 16. Southampton: University of Southampton
- Edwards P, Clarke M, DiGuseppi C et al. (2002) Identification of randomized trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine* 21: 1635–40
- Egger M, Davey Smith G, Altman DG (2000) *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books
- Glaser BG, Strauss AL (1967) *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter
- 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 and 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*. Australia: Deakin University

- Johnson JA, Biegel DE, Shafran R (2000) Concept mapping in mental health: uses and adaptations. *Evaluation and Programme Planning* 23 (1): 67–75
- Kelly MP, Swann C, Morgan A et al. (2002) *Methodological problems in constructing the evidence base in public health*. London: Health Development Agency
- 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 (2008) *Systematic reviews: CRD's guidance for undertaking reviews in healthcare*. York: NHS Centre for Reviews and Dissemination. Available from www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf
- Noblit G, Hare RD (1988) *Meta-ethnography: synthesising qualitative studies*. London: Sage
- 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: 804–8
- 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: 886–92
- Ogilvie D, Fayter D, Petticrew M et al. (2008) The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology* 8: 8
- Oliver S, Kavanagh J, Caird J et al. (2008) *Health promotion, inequalities and young people's health: a systematic review of research*. London: University of London. Available from www.imbi.uni-freiburg.de/OJS/cca/index.php/cca/article/viewArticle/5068
- 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 (2003) Why certain systematic reviews reach uncertain conclusions. *BMJ* 326: 756–8
- Petticrew M, Roberts H (2003) Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57: 527–9
- 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
- Popay J, Rogers A, Williams G (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* 8 (3): 341–51

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

Sutton AJ, Jones DR, Abrams KR et al. (2000) *Methods for meta-analysis in medical research*. London: John Wiley

Swann C, Falce C, Morgan A et al. (2005) *HDA evidence base: process and quality standards for evidence briefings*. London: Health Development Agency

Tooth L, Ware R, Bain C et al. (2005) Quality of reporting of observational longitudinal research. *American Journal of Epidemiology* 161 (3): 280–8

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

6 Incorporating health economics¹⁰

6.1 Introduction

Health economics is about using resources efficiently to improve the population's health. Health economic analysis and evaluation forms an integral part of the public health guidance development process. The Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Groups (PDGs) are required to make decisions informed by the best available evidence of both effectiveness and cost effectiveness.

Economic evaluation compares the costs and consequences of alternative courses of action. The cost effectiveness of an intervention or programme is assessed to ensure maximum health gain from the available resources (which are finite within the public sector). If resources are used for interventions that are not cost effective, the population as a whole gains fewer health benefits (that is, there is a greater 'opportunity cost'). However, a balance must be struck between efficient allocation of resources on the one hand and an equitable allocation of those resources on the other.

This chapter describes how the health economic evidence should be collated and analysed. It also sets out the principles for conducting health economic modelling analyses if there is insufficient evidence in the literature to assess the cost effectiveness of interventions.

Public health recommendations should be based on the estimated cost of each intervention and how that relates to the expected health benefits (that is, recommendations should be cost effective). Recommendations should not be made on the basis of the total cost or the resource impact of implementing them. So, if the evidence suggests that an intervention provides significant health benefits and the cost per person is acceptable, it should be recommended, even if it would be expensive to implement across the whole population.

Although commissioners do need to know the resource and cost implications of implementing NICE guidance, this assessment is not within the remit of the economic analysis. NICE undertakes a separate cost-impact analysis after the public health guidance is published and this forms part of the implementation tool set.

¹⁰ Adapted from chapter 7 in 'The guidelines manual' (NICE 2009). Available from www.nice.org.uk

6.2 Reviewing economic evaluations

The review of economic studies should be systematic but focused. If a high-quality economic study has been published that addresses a structured public health question and is relevant to current practice, then further modelling will not be necessary. This frees up time to model other questions.

6.2.1 Critical appraisal of economic studies

The search for economic studies should have identified papers that are most relevant to current public sector practice and, hence, likely to inform PHIAC/PDG decision-making. It should have focused on 'full' economic evaluations that compare both the costs and the health consequences of the alternative options under consideration.

The process for sifting and selecting economic studies for critical appraisal is essentially the same as for effectiveness studies (see chapters 4 and 5). The CPHE project team and the review team should discuss and agree the inclusion criteria. These should include the populations and interventions that relate to the review question and should also specify:

- An appropriate date range, because older studies may reflect outdated practices.
- The country or setting, because studies conducted in other healthcare systems or sectors might not be relevant to the public sector in the UK. In some cases, it may be appropriate to limit inclusion to UK-based or Organisation for Economic Cooperation and Development (OECD) studies.
- The type of economic evaluation. This may include cost–utility, cost–benefit, cost-effectiveness, cost-minimisation or cost–consequence analyses. Non-comparative costing studies, 'burden of disease' studies and 'cost of illness' studies should usually be excluded.

6.2.2 Assessing economic evaluation studies

Estimates of resource use obtained from public health studies should be reviewed using the processes described above. Reservations about the applicability of these estimates to routine public sector practice should be noted in the evidence tables. PHIAC/the PDG should take these reservations into account.

The criteria for appraising costs, cost-effectiveness ratios and net benefits are different because these estimates are usually obtained using some form of modelling. Modelling can include:

- formal decision-analytic models
- economic evaluations conducted alongside effectiveness trials (these usually require some additional information, for example, unit costs, health-state valuations or long-term outcome probabilities)
- estimation procedures to predict long-term costs and outcomes.

The review team should appraise all the economic evaluation studies (both published and unpublished) used to inform public health recommendations, using the economic evaluations checklist (see appendix I). Unpublished economic evaluations may include studies submitted by stakeholders or pre-publication academic papers (see sections 4.4.4, 4.4.5 and 5.3.2.3). The same criteria should be applied to any additional economic evaluations conducted for the guidance.

Given the small number of economic evaluations usually obtained through the literature searches, title and abstract screening, full-paper screening and quality appraisal should normally be undertaken by two reviewers. Any disagreement between the two reviewers, at any stage, should be resolved by discussion or recourse to a third reviewer.

The checklist includes a section on the methodological quality of the study (that is, the extent to which it fulfils its stated objectives), a section on its applicability to the review question and a section on the NICE health economics methodology used for decision-making. It can be used to judge overall methodological quality as follows:

- **Very serious limitations** The study fails to meet one or more quality criteria and this is very likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from further consideration.
- **Potentially serious limitations** The study fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness.
- **Minor limitations** The study meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness.

The robustness of the study results (that is, its methodological limitations) may sometimes be apparent from reported sensitivity analyses. If not, judgement will be needed to assess whether a limitation is likely to change the results and conclusions.

The checklist can also be used to judge the overall applicability of the study in the context of the guidance:

- **Not applicable** The study fails to meet one or more applicability criteria and this is very likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from further consideration.
- **Partially applicable** The study fails to meet one or more applicability criteria and this could change the conclusions about cost effectiveness.
- **Directly applicable** The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness.

If necessary, the health technology assessment modelling checklist for decision-analytic models (Philips et al. 2004) may also be used to assess the methodological quality of modelling studies in more detail.

When assessing the evidence of effectiveness and cost effectiveness, PHIAC/the PDG should consider the implications of any unexplained differences between model results.

The review team should record its judgements based on the economic evaluation checklist (and health technology assessment modelling checklist, if appropriate) in an appendix to the guidance. The checklist 'comments' column should be used to record the reasons for these judgements and, where necessary, additional details about the studies.

6.2.3 Summarising and presenting cost-effectiveness results

The review team should present cost, cost effectiveness or net-benefit estimates from published or unpublished studies, or from economic analyses conducted for the guidance, in an economic evidence table. (See appendix K for the table template.) Where economic information is not available (or it is not thought to be relevant to the question) this should be explicitly stated. An overall assessment of any study limitations and its applicability should also be included. Issues that PHIAC/the PDG should consider when assessing the economic evidence should be noted in the 'Notes' column.

The economic evaluation results should include a 'best-available' estimate or range for the incremental cost, the incremental effect and, where relevant, the incremental cost-effectiveness ratio or net-benefit estimate. A summary of the degree of uncertainty in the estimates should also be presented, reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

The review team should include a short economic evidence statement at the end of the economic evidence synthesis. This should summarise key features of the evidence on cost effectiveness in a similar format to that used to summarise the effectiveness evidence (see section 5.5).

6.3 Prioritising areas for further economic analysis

Some further analysis will usually be required, as the health economic literature is rarely comprehensive or conclusive enough.

Additional economic analyses may involve adapting an existing model or developing a new one. For many public health interventions, it will be necessary to extrapolate effectiveness data over long time periods. It will also be necessary to derive long-term chronic and quality-adjusted life years (QALYs) outcomes from short-term, intermediate results.

Close collaboration between the health economist contractor, the CPHE project team and PHIAC/the PDG is essential early in the guidance development process to ensure the:

- most important questions or intervention areas are selected for economic analysis
- overall modelling approach is appropriate
- important health effects and resource costs are all included
- effects and outcomes not related to health are included (if they are important for the public sector)
- best available effectiveness, epidemiological and resource evidence is used
- model assumptions are plausible
- analysis results are interpreted appropriately and any limitations are acknowledged.

Economic analysis is potentially useful for any question involving comparisons between interventions or programmes. This includes comparisons of prevention methods used to achieve the same health outcomes. For instance, school-based interventions to reduce under-age drinking could be compared with enforcement of current legislation restricting the sale of alcohol to under-18s. It may also be appropriate when comparing different combinations or sequences of interventions. However, given the broad scope of much public health guidance, it will not be possible to conduct original analyses for every intervention or question.

The selection of interventions or questions for further economic analysis (including

modelling) should be a joint decision between the health economist, the CPHE project team and PHIAC/the PDG. Selection should be based on systematic consideration of the potential value of carrying out an economic analysis across all interventions. Note, an economic analysis will be more useful if it is likely to influence a recommendation – and the recommendation will have a significant health and financial impact. So, the decision about whether or not to carry out an economic analysis depends on the:

- expected net benefit of the recommendation (the number of individuals affected and the potential impact on costs and health outcomes per individual)
- degree of uncertainty in the cost effectiveness literature and the likelihood that economic analysis will clarify matters.

Economic modelling may not be warranted if it is not possible to estimate even a ball-park figure for cost effectiveness, or if the published cost effectiveness evidence is so reliable that further economic analysis would be superfluous. In addition, it may not be needed when it is obvious that the resource implications are modest in relation to the expected health gains.

6.3.1 Modelling approaches

A cost-effectiveness analysis should usually be undertaken with the health effects measured using a non-monetary outcome indicator. When cost-effectiveness analysis is not appropriate, other validated methods such as cost–benefit or cost–consequences analysis may be used.

A cost–utility analysis, cost-effectiveness analysis where units of effectiveness are expressed in QALYs, is a useful way of measuring and comparing the efficiency of different health interventions. A QALY is an overall measure of health outcome that weighs life expectancy against an estimate of the person's health-related quality of life (measured on a 0–1 scale). NICE uses cost–utility analysis for technology appraisals and clinical guidelines and, if suitable data are available, it should be used for public health guidance. This approach ensures baseline comparability across the UK healthcare sector and across NICE's programmes. It also helps to prioritise which recommendations should be implemented locally.

If there are not enough data to estimate QALYs gained, an alternative measure of cost effectiveness may be considered (such as life years gained, cases averted or a more disease-specific outcome).

It is acknowledged that, for some public health topics, such as mental wellbeing or the quality of life of children, the QALY approach is either under-developed or unsuitable. In such cases, a more pragmatic approach may be needed.

In addition to a cost–utility analysis, other approaches (such as cost–consequences or cost–benefit analyses) may be used to take into account the complex and multidimensional nature of public health interventions and programmes. The aim is to give PHIAc/the PDG the opportunity to consider the multiple, non-health related and non-quantifiable outcomes that may be important to the public sector.

Productivity costs and any costs borne by patients and carers that are not reimbursed by the public sector should not be routinely included in the analyses. However, a wider societal perspective may be considered for some topics. For example, in guidance on how to help people back to work after sickness absence, the benefits of any increase in productivity resulting from their return to work was taken into account.

These analyses could be used to:

- modify a decision based solely on cost-per-QALY considerations
- explain to stakeholders and the public the additional costs and benefits for organisations outside the NHS and personal social services (PSS).

The review team should provide a rationale for using these types of analysis, highlighting what additional information it will provide compared to the usual cost–utility analysis. This should ensure the guidance development process remains fair and transparent.

A cost-effectiveness analysis could be modelled on a single well-conducted randomised controlled trial (RCT), or it could use decision-analytic techniques to analyse probability, cost and health-outcome data from a variety of published sources.

There is often a trade-off between the range of new analyses that can be conducted and the complexity of each piece of analysis. Simple methods may be used if these provide PHIAc/the PDG with sufficient information on which to base a decision. For example, if an intervention is associated with better health outcomes and fewer adverse effects, then a simple decision tree may provide a sufficiently reliable estimate of cost effectiveness. In other situations a more complex approach, such as Markov modelling or discrete-event simulation, may be warranted.

NICE's 'Guide to the methods of technology appraisal' recommends using the following 'reference-case' assumptions as a basis for the cost effectiveness analysis:

- All health effects on individuals are included.
- Equity weightings are not applied to QALYs.
- Costs and health outcomes are discounted at 3.5%.
- Health-related quality of life is valued using choice-based elicitation methods, a representative sample of the general population and validated, generic health-state

instruments. (There is unlikely to be time to collect original quality-of-life valuations, so data collected by alternative methods may be used, but this should be suitably justified.)

- The time horizon should be chosen to incorporate all important costs and effects.

Departures from the reference case may be made, but this must be agreed with the CPHE project team prior to implementation and must be highlighted. The reasons why the reference case was not followed must be given in the report.

Where available, consider linking the model structure to the logic model used to develop the research questions. The logic model illustrates the causal pathways between human behaviour, the social, environmental and biological determinants of health and potential interventions and outcomes (see chapter 2 and appendix A for details).

6.3.2 Perspective

Often the costs of public health interventions – and the benefits – will be borne outside the NHS, predominantly by other public sector organisations. As a result, it is necessary to adopt a public sector perspective. As defined by NICE's statutory instruments¹¹, it shall perform: 'such functions in connection with the promotion of excellence in public health provision and promotion and in that connection the effective use of resources available in the health service and other available public funds'.

This public sector perspective differs from that used for NICE technology appraisals and clinical guidelines: these only consider the health service. However, for some public health guidance, an NHS and PSS perspective may be sufficient to capture all the major costs and benefits.

When appropriate, and following agreement with the CPHE project team, results may also be presented from other perspectives. For example, an employer's perspective could be taken to demonstrate the business case for a public health intervention.

Sensitivity analysis should be used to explore the impact that potential sources of bias and uncertainty could have on model results.

Deterministic sensitivity analysis should be used to explore key structural assumptions; testing whether and how the model results change under alternative, plausible scenarios. It should also be used to test any bias resulting from the data sources selected for key model parameters.

Probabilistic sensitivity analysis should be used to explore the uncertainty arising from imprecision in model parameters. Any uncertainty associated with all parameters can

¹¹ National Institute for Clinical Excellence (Establishment and Constitution) Amendment Order 2005 (www.opsi.gov.uk/si/si2005/20050497.htm)

Table 6.1 Summary of the public health reference case¹²

Element of assessment	Reference case
Defining the decision problem	The scope developed by NICE
Comparator	Interventions routinely used in the public sector, including those regarded as best practice
Perspective on costs	Public sector, including the NHS and PSS
Perspective on outcomes	All health effects on individuals
Type of economic evaluation	Cost-effectiveness analysis (primary analysis) Cost-consequence analysis (secondary analysis) Cost-benefit analysis (secondary analysis)
Synthesis of evidence on outcomes	Based on a systematic review
Measure of health effects	QALYs
Source of data for measurement of health-related quality of life (HRQL)	Reported directly by patients and/or carers
Source of preference data for valuation of changes in HRQL	Representative sample of the public
Discount rate	An annual rate of 3.5% on both costs and health effects
Equity weighting	An additional QALY has the same weight, regardless of the characteristics of the individuals who gain the health benefit

¹² Adapted from table 5.1 in 'Guide to the methods of technology appraisal' (NICE 2008) Available from www.nice.org.uk

be simultaneously reflected in the results. In non-linear decision models, probabilistic methods also provide the best estimates of mean costs and outcomes. However, models incorporating probabilistic methods are more time-consuming to construct and may not always be a priority for health economists working on public health guidance. In such cases, the decision not to use probabilistic methods should be clearly justified in the guidance and deterministic sensitivity analysis should be used.

The 'Guide to the methods of technology appraisal' document includes other useful advice for health economists developing economic models for use in public health guidance.

6.3.3 General principles

Regardless of the modelling approach adopted, economic analysis should be based on the following principles:

- The question being asked for the economic analysis should be clearly specified and appropriate and include comparison of all relevant alternatives for specified population groups.
- Analysis should be carried out by the health economist in collaboration and following agreement with the CPHE project team and PHIAC/the PDG.
- It should be underpinned by the best-quality, publicly available effectiveness evidence.
- There should be the highest level of transparency in the reporting of methods and results. Conventions on reporting economic studies should be followed (see Drummond and Jefferson 1996).
- Potential sources of bias and uncertainty should be explored using an appropriate sensitivity analysis and discussed with PHIAC/the PDG.
- The limitations of the approach and methods used should be discussed with PHIAC/the PDG and presented in the report.

6.3.4 Identifying and selecting model inputs

The review team and the CPHE project team should discuss whether or not any additional searches are needed. Additional searches may be necessary if the effectiveness searches do not provide the information needed for economic modelling. For example, information on:

- the relationship between short- and long-term outcomes

- the links between behaviour and disease (for example, intermediate and longer-term outcomes)
- quality of life
- resource use or costs.

It is not necessary to conduct formal, systematic literature searches for all the types of information required for economic modelling (although effectiveness data used in the modelling should be taken from the effectiveness reviews). For example, it is better to get information on unit costs from: national list prices, the Personal Social Services Research Unit (PSSRU) 'Unit costs of health and social care' report or the Department of Health tariff. Information on costing can also be found in the NICE document 'Developing costing tools: methods guide' (2008) and through discussion with the costing analyst in the NICE implementation team. In addition, it might be better to obtain some information about epidemiology or public services from national statistics or databases, rather than from literature studies.

Ideally, systematic literature reviews should be conducted for other model inputs, for instance to model the relative risk of coronary heart disease in relation to physical activity levels. However, this is time-consuming and the health economist should look at other options first. Possibilities include: searching the public health effectiveness evidence that was used to structure the key question(s) (and perhaps other relevant questions) in the scope; or liaising with the review team, PHIAC/the PDG and other experts.

If an additional literature search is necessary, the review team should discuss and agree this with the information scientist and the CPHE project team. If longer-term follow-up data are required, a literature search to identify cohort studies may be appropriate. The report by Philips et al. (2004) is a useful guide to data searching methods for economic models.

6.4 Economic evidence and guidance recommendations

For an economic study or analysis to be useful, it must inform the public health guidance recommendations. PHIAC/the PDG should discuss cost effectiveness in parallel with general effectiveness when formulating recommendations (see chapter 7).

If there is strong evidence that an intervention dominates the alternatives (that is, it is both more effective and less costly), clearly it should be recommended. However, if one intervention is more effective but also more costly than another, then the incremental cost-effectiveness ratio (ICER) should be considered. The cost per QALY gained should be calculated as the difference in mean cost divided by the difference in mean QALYs for one intervention compared with the next most effective alternative.

If one intervention appears to be more effective than another, PHIAC/the PDG will have to decide whether any increase in cost associated with the increase in effectiveness represents reasonable 'value for money'. In doing so, it should refer, as appropriate, to the principles outlined in NICE's report 'Social value judgements: principles for the development of NICE guidance. Second edition' (2008). It states that:

'NICE has never identified an ICER above which interventions should not be recommended and below which they should. However, in general, interventions with an ICER of less than £20,000 per QALY gained are considered to be cost effective. Where advisory bodies consider that particular interventions with an ICER of less than £20,000 per QALY gained should not be provided by the NHS they should provide explicit reasons (for example that there are significant limitations to the generalisability of the evidence for effectiveness). Above a most plausible ICER of £20,000 per QALY gained, judgements about the acceptability of the intervention as an effective use of NHS resources will specifically take account of the following factors.

- The degree of certainty around the ICER. In particular, advisory bodies will be more cautious about recommending a technology when they are less certain about the ICERs presented in the cost-effectiveness analysis.
- The presence of strong reasons indicating that the assessment of the change in the quality of life inadequately captured, and may therefore misrepresent, the health gain.
- When the intervention is an innovation that adds demonstrable and distinct substantial benefits that may not have been adequately captured in the measurement of health gain.

As the ICER of an intervention increases in the £20,000 to £30,000 range, an advisory body's judgement about its acceptability as an effective use of NHS resources should make explicit reference to the relevant factors considered above. Above a most plausible ICER of £30,000 per QALY gained, advisory bodies will need to make an increasingly stronger case for supporting the intervention as an effective use of NHS resources with respect to the factors considered above'.

However, such thresholds do not exist outside the health sector, making it difficult to judge whether the benefits accruing to the non-health sectors are cost effective. PHIAC/ the PDG should take into account the factors it considers most appropriate. These could include non-health related outcomes that are valued by the rest of the public sector (like reduced levels of civil disobedience resulting from reduced alcohol consumption).

Decisions about whether to recommend interventions should not be based on cost effectiveness alone. The advisory committees responsible for developing the guidance should also take into account other factors, such as the need to prevent discrimination

and to promote equality. They will consider trade-offs between efficient and equitable allocations of resources. These factors should be explained in the considerations section of the guidance (see section 7.5). If a structured public health question is not considered for further economic analysis, PHIAC/the PDG should still consider the likely cost effectiveness of the associated recommendations. If available, this assessment may be based on published estimates of cost effectiveness or, if necessary, a qualitative judgement.

6.5 References and further reading

Briggs A, Claxton K, Sculpher M (2006) *Decision modelling for health economic evaluation*. Oxford: Oxford University Press

Cooper NJ, Sutton AJ, Ades AE et al. (2007) Use of evidence in economic decision models: practical issues and methodological challenges. *Health Economics* 16: 1277–86

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

Drummond MF, McGuire A (2001) *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press

Drummond MF, Sculpher MJ, Torrance GW et al. (2005) *Methods for the economic evaluation of health care programmes*, 3rd edition. Oxford: Oxford University Press

Eccles M, Mason J (2001) How to develop cost-conscious guidelines. *Health Technology Assessment* 5: 1–69

NHS Centre for Reviews and Dissemination (2001) *Improving access to cost-effectiveness. Information for health care decision making: the NHS Economic Evaluation Database*. CRD report number 6, 2nd edition. York: NHS Centre for Reviews and Dissemination, University of York. Available from www.york.ac.uk/inst/crd/report6.htm

Philips Z, Ginnelly L, Sculpher M et al. (2004) Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 8: iii–iv, ix–xi, 1–158

Raftery J, editor (1999–2001) *Economics notes series* [online]. BMJ. Available from www.bmj.com

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

7 Developing recommendations

7.1 Introduction

Developing recommendations is at the heart of the work of the Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Groups (PDGs). It is not a straightforward task and it may not always be easy to reach agreement.

It is vital to ensure the recommendations:

- are informed by the most appropriate and available evidence (both scientific and other evidence)
- are set within a framework that acknowledges a range of social value judgements (see 'Social value judgements: principles for the development of NICE guidance. Second edition' [2008])
- take account of relevant theories of public health and behaviour change
- reflect the views and experiences of both those being advised to take action (for example, health professionals or teachers) and the people who might be affected by that action (the target population and their families or carers)
- are clear
- are practical (that is, they can be implemented).

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

1. Considering the evidence and other factors.
2. Creating the format and wording of recommendations.
3. Revising the recommendations following fieldwork and stakeholder consultation.
4. Developing considerations.
5. Equality and diversity.
6. Prioritising recommendations (if there are more than eight and it is deemed necessary).
7. Formulating research recommendations.

7.2 Considering the evidence and other issues

Recommendations are developed using a range of scientific evidence (with and without 'context') and other evidence – such as expert testimony, stakeholder and practitioner views, committee discussions and debate (see chapter 3).

The evidence may be assessed for validity, reliability and bias, however evidence is not the sole determinant of the content of recommendations. It requires interpretation, especially an assessment of its implicit and explicit value base. It also needs to be assessed in light of the conceptual framework for public health (appendix A) and theories relating to individual and organisational behaviour change.

PHIAC/the PDG's 'considered judgement' should take account of a range of issues (including any ethical issues and social value judgements) and policy imperatives, as well as equality and diversity legislation to ensure the recommendations are ethical, practical and specific. There are no hard-and-fast rules or mechanisms for doing this: PHIAC/the PDG should make conscious and explicit use of its members' skills and expertise to make assumptions and apply inductive and deductive reasoning.

PHIAC/the PDG, with support from the Centre for Public Health Excellence (CPHE) project team, should examine the evidence and related documents and discuss whether or not they address the issues under consideration. They should focus on the decision that needs to be made, the recommendations that are required as well as the research questions used for the NICE evidence reviews and the key questions from the scope. (This should be documented in the minutes.) To do this, they should consider the following issues (these are not in hierarchical order).

7.2.1 Strength (type, quality, quantity and consistency) of the evidence

Statistical and methodological issues and the study types available should all be taken into account, along with the degree of bias in the findings. The evidence statements will describe the number, type and quality of studies and summarise the strength of evidence. PHIAC/the PDG should agree that this is a fair summary of the evidence and be mindful that the strongest available evidence does not necessarily translate to important areas for action. The impact of the potential benefits and harms also needs to be taken into account (see sections 7.2.4 to 7.2.6 below).

Members may want to discuss any inconsistencies in the findings (different studies may relate to slightly different interventions, populations or settings). In addition, they should use the NICE conceptual framework for public health (appendix A) to identify which causation vectors are applicable. (See also chapter 2.)

7.2.2 Applicability of evidence to the target populations and settings

The evidence statements conclude with a summary statement on applicability. This describes the evidence as 'directly', 'partially' or 'not applicable' and why. PHIAC/the PDG should reach its own conclusion, based on its members' knowledge, experience and understanding of the target populations. They may choose to make a recommendation on evidence that is weaker, but more applicable, than stronger evidence from another context. (For more detail on the assessing applicability, see section 5.6.)

7.2.3 Availability of evidence to support implementation – including evidence from practice

PHIAC/the PDG should assess the extent to which the available evidence is about efficacy, effectiveness or both. Often the distinction between the two is not made clear in reports of public health interventions, not least because the number of efficacy studies is relatively small (compared to clinical studies).

PHIAC/the PDG should also judge whether or not it will be possible to put the recommendations into practice. They can decide by using the fieldwork reports, expert testimony and by drawing on their own experience. They may also be able to draw on qualitative studies from the reviews or other forms of evidence relating to organisational and political processes.

In addition, PHIAC/the PDG should assess the degree of change in practice required, staff training needs, policy levers and funding streams.

7.2.4 Relative value of the outcomes (including impact on inequalities)

PHIAC/the PDG should assess the extent to which the recommendations may impact on health inequalities. This needs to be made clear, regardless of whether the recommendation is aimed at the whole population, specific subgroups or a combination of both.

7.2.5 Trade-off between harms and benefits

Where possible, PHIAC/the PDG should assess any potentially negative effects and whether these are offset by the anticipated benefits.

7.2.6 Size of effect and potential impact on individual and population health (if applicable)

PHIAC/the PDG should consider whether it is possible to anticipate effect sizes at individual or population-level. If this is the case, it will be important to consider effect

sizes along the whole causal chain, not just at the end points. This may be difficult because of the current state of the evidence.

7.2.7 Cost effectiveness

PHIAC/the PDG should consider cost-effectiveness evidence and the economic models (see chapter 6).

7.2.8 Target groups

PHIAC/the PDG should assess the particular characteristics of the target group, paying attention to social differences (including class, gender, ethnicity, disability, culture and sexual orientation). It should assess how these characteristics will impact on the effectiveness of interventions.

7.2.9 Equality and diversity

(See section 7.6.)

7.2.10 Philosophical basis for making recommendations

All evidence requires interpretation as evidence alone cannot determine the content of a recommendation. The development of evidence-based recommendations involves inferential, inductive or deductive reasoning:

- inferential because it involves moving from that which is known (the evidence) to uncertainty about what is reasonably expected to happen as a consequence of implementing a recommendation
- inductive when it is derived from evidence
- deductive when it is drawn from theory or methodological principles.

Other information is needed to make evidence comprehensible – no methodology can take the place of imaginative interpretation. However, judgement and interpretation can only be based on past experience. This involves making comparisons with previous observations and evidence about the same or similar issues, as well as taking into account general principles, ideas and values.

NICE's social value judgements paper ('Social value judgements: principles for the development of NICE guidance. Second edition' [2008]) explicitly acknowledges that non-scientific values are brought to bear and all of NICE's advisory committees are encouraged to take account of (and to make explicit) the value judgements they make. The advisory committees may also draw upon the principles outlined in the Nuffield

Committee on Bioethics report on public health when making its judgements¹³.

PHIAC/the PDG should make explicit the kinds of interpretive methods they are using during the inferential process.

7.2.11 Conceptual framework and logic model

When the advisory committee is developing its recommendations, it should keep in mind the overarching conceptual framework, how it applies to the topic in question and the resulting logic models. This will help to clarify the practical issues involved in bridging the gaps between the evidence and producing a recommendation. It will also demonstrate how evidence drawn from the organisational and political sciences can help inform the decision-making process. (See appendix A.)

7.2.12 Additional documents

In addition to the NICE evidence review(s) (particularly the evidence statements and the economic analysis and modelling), other documents available to help PHIAC/the PDG develop recommendations include (not in hierarchical order):

- the scope
- related NICE guidance
- policy reports and guidance produced by other organisations
- position papers on current practice
- evidence (including practice-based evidence) and opinions from expert witnesses
- evidence reviews about the views and experiences of the target population
- stakeholder comments on the draft recommendations
- fieldwork report (when available, following consultation and fieldwork on the draft recommendations)
- documents which help members consider equity issues – such as NICE’s documentation on its responsibilities in relation to equalities legislation¹⁴.

¹³Nuffield Council on Bioethics (2007) Public health: ethical issues. London: Nuffield Council on Bioethics. Available from www.nuffieldbioethics.org/fileLibrary/pdf/Public_health_-_ethical_issues.pdf

¹⁴The equality scheme and action plan for implementation (approved by NICE’s Board in March 2007) are available at www.nice.org.uk/aboutnice/howwework/NICEEqualityScheme.jsp

7.2.13 Challenges

Table 7.1 summarises the challenges that may arise during the deliberations – and possible approaches to overcoming them. In each case, PHIAC/the PDG will need to make its approach explicit in the ‘Considerations’ section of the guidance, stating the basis for its decision(s) and the assumption(s) made (see section 7.5).

Table 7.1 Interpreting the evidence: possible challenges and ways to overcome them

Challenge	Possible approach
No evidence, weak evidence or it is only partially applicable	<ul style="list-style-type: none"> Consider the ‘direction of travel’ of the evidence available. Make a tentative recommendation and develop a ‘consideration’ which explains why weak or partially applicable evidence has been used Consider evidence from practice (see below)
Only evidence of a similar type and quality is available and the findings conflict (inconsistent or mixed evidence)	<ul style="list-style-type: none"> Consider the reasons for conflict. For example, if this is because different groups of people might respond differently to an intervention or programme, consider making recommendations for specific groups Identify studies that are most applicable to the target population and setting and, where appropriate, use them as a basis for recommendations
Evidence not directly applicable to the target population (for example, it covers a different age group)	<ul style="list-style-type: none"> Consider the degree to which the findings can be extrapolated to the target population. For example, this may be possible if it is high-quality evidence drawn from a largely similar but different population group
Evidence conflicts with existing government policy or NICE guidance	<ul style="list-style-type: none"> Consider the reason(s) for conflict. For example, was the policy/guidance evidence-based? Has the evidence changed substantially since the policy/guidance was developed? Were the goals/intentions of the policy/guidance different? The CPHE project team may be able to discuss the conflict(s) with the relevant policy guidance team, as necessary, to help resolve this issue. However, be mindful that this latest NICE guidance might directly inform changes in government policy or supercede previous NICE guidance

Table 7.1 Continued

Challenge	Possible approach
Limited information on cost effectiveness	<ul style="list-style-type: none"> • For recommendations that are likely to have a significant resource impact, consider using economic modelling to develop an estimate of cost effectiveness
Unclear how to make best use of the different types of evidence from practice (including evidence provided by committee members, expert witnesses, stakeholders and the target population)	<ul style="list-style-type: none"> • Consider how evidence from practice can help answer the key questions • Consider what weight should be given to evidence from practice compared to evidence from the NICE evidence reviews • Consider how evidence from practice can: 1) support the NICE evidence reviews of effectiveness and cost effectiveness and 2) address gaps in the evidence on effectiveness and cost effectiveness • Consider whether it is possible to record the conclusions drawn from practice in a consistent and transparent way. Specifically, can the conclusions be developed into evidence statements and discussed in the considerations section of the guidance?

7.2.14 The process in practice

As soon as members have discussed the findings of a NICE evidence review (or any expert testimony) PHIAC/the PDG should start drafting recommendations. This is an iterative process; the recommendations are likely to be revised on a number of occasions before the wording is finalised.

First, PHIAC/the PDG should decide what they want to recommend and which sectors (including which professionals within those sectors) should act on the recommendations. (As an example, the recommendations could be aimed at practitioners within the NHS, schools, workplaces or local authorities.)

In the early stages, it can be helpful to work in small groups, supported by the CPHE project team and using sample templates. It may also help if the CPHE project team develops a first draft of the recommendations as a starting point for discussion, based on PHIAC/the PDG's initial deliberations as a group. However they are developed, the CPHE project team should ensure the draft recommendations are clearly linked to evidence statements.

Between PHIAC/PDG meetings, the nuances of the words can be refined via email discussions among members (again, supported by the CPHE project team). (See 'The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public [second edition]' [NICE 2009] for terms of reference and the standing orders which define quoracy for advisory committees.)

The recommendations may be prioritised (see section 7.7).

Where evidence on effectiveness or cost effectiveness is lacking or conflicting, PHIAC/the PDG may decide that further research should be a condition for implementation.

Decisions can be made using a variety of approaches: discussion, informal or formal consensus or formal voting (for example, in instances when members disagree). The proceedings should be recorded and a clear statement made about the factors that have been considered and the methods used to achieve consensus. This ensures the process is as transparent as possible.

A summary of the generic and specific issues considered and the key deliberations should be given in the 'Considerations' section of the guidance (see section 7.5).

7.3 Format and wording of recommendations

Writing the recommendations is one of the most important and difficult steps in developing guidance. Great care should be taken. Each recommendation should answer the reader's main question: 'What does this mean for me?' In addition, it should clearly specify the intervention/action to be taken (what, how often and for how long?) and the context/circumstances (where and when?).

The wording must be concise and unambiguous so that the target audience knows what to do in practice and the public know what is being recommended. The CPHE project team should ensure that PHIAC/the PDG is supplied with a copy of the booklet 'Writing for NICE'.

7.3.1 Format of recommendations

The recommendations are grouped under three main headings as below.

Who is the target population?

All those who will be affected by the recommendation. This may include:

- individuals
- communities or families

- larger population groups defined by a range of factors (for example, by age, gender, ethnicity, setting).

Who should take action?

The professionals and others who should take action. These may be:

- practitioners
- commissioners
- policy makers
- researchers.

They may be subdivided by sector and setting, for example:

- NHS
- other public sector bodies (government, government agencies and local government, arm's length bodies, armed forces, prisons, police service, education)
- private and voluntary organisations (large, medium and small).

They may refer to specific job titles, such as:

- teachers
- GPs.

Ensure those taking action are listed by type of organisation or by job title; do not mix organisations and job titles in the same list.

What action should they take?

Actions should be as specific as possible, although how prescriptive they are will be decided on a case-by-case basis and will depend on the evidence available. They may cover:

- strategy, policy and planning
- service management and delivery
- individual practice
- research priorities.

7.3.2 Wording of recommendations

Each recommendation should:

- Stand alone and be understood without reference to supporting material (supporting information can be included in the 'Considerations' section of the guidance or as part of the implementation materials).
- Be as specific as possible about the action and who should take it (the ability to check whether it is being implemented properly [audit] should be considered when finalising the wording).
- Only contain one main action in each bullet point.
- Provide a clear link to the supporting evidence statements and evidence reviews, preferably with a numeric reference (to review number and evidence statement number).
- Avoid, wherever possible, terminology and jargon – where this is not possible, it needs to be clearly defined and unambiguous (see 'Writing for NICE').
- Avoid trade names. Any reference to products (for example, pedometers) and services (for example, slimming clubs) should be made in general terms to avoid giving the impression that NICE endorses a particular brand.
- Avoid implying that interventions/actions should be 'done' to people (that is, use 'offer' and 'discuss' rather than 'prescribe' or 'give', also avoid 'subjects' – use 'people', 'patients', 'clients' or 'service users' instead).
- Avoid labelling people (that is, don't describe someone as a 'drug user' or 'smoker', use instead, 'someone who takes drugs or smokes').
- Emphasise the need for lay people to be involved – by acknowledging their role in decision-making.
- Include cross-references to recommendations from other NICE guidance to avoid the need to repeat information. It should be clear where the recommendations come from (refer to the guidance template for instructions). Recommendations from other NICE guidance or NHS policy (for example, national service frameworks) can be quoted verbatim (as appropriate). Recommendations from other (non-NICE) guidance should not be quoted verbatim.

7.3.3 Reflecting the strength of recommendations

NICE's public health recommendations are not graded, but PHIAc/the PDG's view of how important they are should be clear from the wording (see table 7.2). The importance of a recommendation should not necessarily reflect the strength of the evidence available to support it. Other important factors (for example, ethics, principles, potential outcomes and equality issues) all need to be considered by PHIAc/the PDG (see sections 7.2.1 to 7.2.11).

Table 7.2 Reflecting the strength of recommendations

Level of certainty	Wording
Interventions that must be used	<p>PHIAc/PDG members are encouraged to opt for advisory terms such as 'should' or 'could' (see below).</p> <p>'Must' should only be used when, for example, the recommendation links to enforceable legislation (such as health and safety regulations). It can also be used if PHIAc/the PDG believes there will be serious repercussions if the recommendation is not followed. In such a case, a clear rationale for using 'must' should be set out and discussed with the CPHE director</p>
Interventions that should be used	<p>PHIAc/the PDG is confident that the intervention will do more good than harm and is likely to be cost effective.</p> <p>Word recommendations of this type as direct instructions using verbs such as 'offer', 'assess', 'refer'.</p> <p>If these recommendations refer to all people and situations (where the evidence is clear and uncontested) they should be worded, 'always do this'. They can include caveats (where the evidence is less clear/mixed) such as, 'do this when'.</p> <p>Example (from 'Community-based interventions to reduce substance misuse among vulnerable and disadvantaged children and young people' [NICE 2007]):</p> <ul style="list-style-type: none"> ● Offer one or more motivational interviews, according to the young person's needs. Each session should last about an hour and the interviewer should encourage them to: <ul style="list-style-type: none"> – discuss their use of both legal and illegal substances – reflect on any physical, psychological, social, education and legal issues related to their substance misuse – set goals to reduce or stop misusing substances

Table 7.2 Continued

Level of certainty	Wording
Interventions that could be used	<p>PHIAC/the PDG has determined that the intervention is effective and/or cost effective, but other options may be similarly effective and/or cost effective. Or the choice of intervention (or the decision whether to have one at all) is likely to vary depending on the client's values and preferences.</p> <p>Word recommendations of this type as direct instructions, but add 'consider' or 'could' – for example, 'consider referring'.</p> <p>Example (from 'One to one interventions to reduce the transmission of sexually transmitted infections (STIs) including HIV, and to reduce the rate of under 18 conceptions, especially among vulnerable and at risk groups' [NICE 2007]):</p> <ul style="list-style-type: none"> • Identify individuals at high risk of STIs using their sexual history. Opportunities for risk assessment may arise during consultations on contraception, pregnancy or abortion, and when carrying out a cervical smear test, offering an STI test or providing travel immunisation. Risk assessment could also be carried out during routine care or when a new patient registers. <p>In exceptional circumstances, the committee/group may wish to consider making 'only in research' recommendations</p>
Interventions that should not be used	<p>State explicitly if a particular action should not be carried out or should be stopped (because, for example, it is ineffective or not cost effective).</p> <p>Example (from 'Smoking cessation services in primary care, pharmacies, local authorities and workplaces, particularly for manual working groups, pregnant women and hard to reach communities' [NICE 2008]):</p> <ul style="list-style-type: none"> • If a smoker's attempt to quit is unsuccessful using NRT, varenicline or bupropion, do not offer a repeat prescription within 6 months, unless special circumstances have hampered the person's initial attempt to stop smoking, when it may be reasonable to try again sooner

7.4 Revisions following fieldwork and stakeholder consultation

Fieldwork and stakeholder consultation is carried out to test the draft recommendations with practitioners.

The fieldwork method is described in chapter 8. A summary of the collated and analysed data and the key implications for the recommendations – including the implications for local and national public health policy and practice – is presented as a fieldwork report. This makes use of direct quotations from practitioners.

The stakeholder consultation is described in ‘The NICE public health guidance development process: an overview for stakeholders including public health practitioners, policy makers and the public (second edition)’ (NICE 2009). Registered stakeholders normally include professional organisations and statutory agencies representing practitioners, as well as voluntary organisations run by, or representing the interests of, the target populations. The CPHE project team should prepare a summary of stakeholder responses.

The CPHE project team may occasionally commission work to test out the draft recommendations directly with the target population.

7.4.1 PHIAC/PDG meeting following fieldwork and stakeholder consultation

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

- fieldwork report
- summary of how fieldwork data impact on the draft recommendations (as necessary)
- summary of stakeholder responses to the guidance consultation
- summary of how stakeholder responses impact on the draft recommendations (as necessary)
- report on direct consultation with the target population (if conducted)
- equality and diversity assessment carried out on the draft recommendations (see section 7.6).

If it appears from the consultation or fieldwork that professionals or others (as appropriate) do not endorse a recommendation, PHIAC/the PDG should consider:

- the possible reason(s) (for example, they may have concerns over training issues or capacity)

- whether to amend the recommendation or associated recommendations to support implementation.

7.5 Considerations

The 'Considerations' section of the guidance should clearly illustrate the range of issues PHIAC/the PDG has considered in developing the recommendations. It should also make explicit the criteria used to create and prioritise them.

This section can be developed using the issues and documents listed in section 7.2, the minutes from all its meetings (including any subgroup meetings) and records of any email discussions.

The following information may be included:

- How the evidence statements were developed into recommendations – such as detail on the decisions made in relation to issues raised in table 7.1 and, in particular, on the strength and applicability of the evidence available.
- How recommendations were prioritised – for example, whether this was based on the strength of evidence or policy imperatives.
- The rationale for making recommendations which do not answer the key questions in the scope.
- The rationale for making recommendations where there is a lack of evidence of effectiveness or cost effectiveness.
- How evidence from practice was defined and the relative weight it was given compared to the evidence of effectiveness or cost effectiveness.
- Testimony from expert witnesses.
- Key facts for example, in relation to legislation, policy, funding and organisational issues.
- Issues outside the remit of the guidance – to re-emphasise them.
- Evidence not considered due to its quality or focus or due to time constraints.

The 'Considerations' section should not include any text which could be construed as a recommendation. However, it may be used to point out where PHIAC/the PDG would have liked to have made a recommendation but felt that there was insufficient evidence or lack of a rationale for doing so.

7.6 Equality and diversity

NICE guidance on assessing recommendations in line with equality and diversity legislation focuses on two questions:

1. Does the guidance avoid unlawful discrimination?
2. Are there ways in which the guidance could better promote equality?

Equality issues are considered during the normal guidance development process and it is likely that many will have been considered during PHIAC/the PDG's deliberations. These should be recorded in the 'Considerations' section of the guidance. (Where a direction suggested by the evidence has been altered in the interests of promoting equality, this should also be recorded.)

Recommendations should be formally assessed against equality considerations after the draft guidance has been issued for consultation. The findings should be considered by PHIAC/the PDG when it reviews stakeholder comments and fieldwork.

To avoid unlawful discrimination, four issues are considered:

- whether specific groups may be denied access to an intervention
- whether specific groups will, in practice, find it more or less easy to access the intervention
- whether any assessment needed to access the intervention will make it more or less difficult for specific groups to gain access
- whether any features make it impossible or unreasonably difficult for people with disabilities to access the intervention.

In addition, the assessment checks whether stakeholders have raised any areas of possible discrimination.

Ideally, opportunities to promote equality will have been maximised during development of the draft recommendations. Nevertheless, after the consultation PHIAC/the PDG should reconsider the draft recommendations specifically from this perspective: would, for example, changes to the wording (or the deletion of a recommendation or inclusion of new ones) further promote equality?

Suggested changes to the recommendations should be recorded and presented to PHIAC/the PDG, together with stakeholder comments and fieldwork on the draft guidance. Decisions on these issues should also be noted in the minutes of the relevant committee meetings.

7.7 Prioritising recommendations for implementation

Occasionally, if there are more than eight recommendations, some of them may be identified as key priorities. NICE provides implementation support for the key priorities (or all of the recommendations if none have been prioritised). This can take the form of promotional slides and other tools that might aid implementation.

No more than 10 (or 25% of the total) recommendations should be prioritised. PHIAC/the PDG can use many criteria to select them. The criteria should be transparent and should take into account the issues outlined in section 7.2. Some of the possible reasons why a recommendation will be prioritised include the fact it:

- is not part of current practice
- highlights the need for practice to change
- requires retraining or the development of new skills
- requires implementation by a broad range of agencies or across a range of settings
- may be viewed as potentially contentious or difficult to implement for other reasons
- is particularly likely to benefit from implementation support.

Information from various sources, including national audit data, the fieldwork report and the experience of PHIAC/the PDG should be used to identify implementation priorities. The choice should also be informed by a broad range of stakeholders, as part of the formal consultation process.

There should be a clear record detailing the criteria and the process used for agreeing them (a brief overview should be given in the 'Considerations' section).

7.8 Formulating research recommendations

Research recommendations provide an opportunity to highlight how the public health evidence base can be improved. More important, they ensure CPHE has access to the best possible evidence when the relevant guidance is being revised.

Research recommendations are a set of specific questions which aim to gather new evidence or strengthen the existing knowledge base, where it was previously equivocal. Both primary and secondary research (for example, systematic reviews) can be recommended.

The recommendations are drawn from a broader set of 'gaps in the evidence' that should be listed in an appendix to the guidance. The latter is a list of areas for further research by those wishing to contribute to the development of a wider public health

evidence base. These may have arisen out of the original evidence reviews or through general discussion at meetings.

This section provides a framework for formulating and prioritising research recommendations.

7.8.1 Principles for formulating research recommendations

There will probably be a large number of potential research recommendations. They 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 should be formulated as one question or as a set of closely related questions. It should consider the importance of issues relating to equality and diversity (for example, gender, ethnicity and people with special needs) and take into account the criteria set out in table 7.4 below. Each recommendation should also be evaluated and prioritised according to these criteria and this information should be presented as an appendix.

Only high-priority research recommendations should be included in the guidance.

7.8.2 Selecting research recommendations

PHIAC/the PDG should select a maximum of five research recommendations for guidance produced using the public health intervention process and 10 for guidance produced using the public health programme process. A research recommendation has two components:

- a well-formulated, answerable question (see below)
- a statement about the importance of the recommended research (see table 7.4).

It should comprise a question with explanatory text of not more than 150 words. See below for an example from 'Brief interventions and referral for smoking cessation in primary care and other settings' (NICE 2006).

Question:

Which brief interventions work best for increasing smoking cessation among lower socioeconomic 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 and have the fewest resources with which to cope.

Alternatively, it may take the form of a recommendation which will answer: 'who should take action?' (for example, the Medical Research Council); and 'what action should they take?' (for example, to ensure certain outcome measures are used in studies that it funds in a particular topic area). If more than one organisation and one action is involved, these should be broken down into a bullet point list under the two headings.

PICO is a widely used mnemonic summarising the four major components of a research question about an intervention: patient (population), intervention, comparison and outcome (see table 7.3).

Table 7.3 Checklist for formulating research questions about interventions using the PICO model

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>Consider:</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>Consider:</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>Consider:</p> <ul style="list-style-type: none"> • all the parameters mentioned above, where applicable
Outcome	<p>What will the researcher need to measure, improve, influence or accomplish? What intervention outcomes should be measured?</p> <p>Consider:</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

7.8.3 Why is this question important?

PHIAC/the PDG should draft a paragraph explaining the need for research and select the five most important recommendations using the following headings.

Table 7.4 Research priorities

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 national service framework or 'Choosing health' [DH 2004])? Specify the relevant document
5. Lack of 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

The CPHE project team should note that continuous feedback from PHIAC/the PDG is essential to improve the above list of questions. They should be asked:

- how user-friendly, acceptable, reliable and valid they are
- how useful they are for generating clear, precise recommendations for research that could feasibly be carried out.

All the research recommendations contained in the guidance are added to a database on the NICE website. Those classed as high priority are highlighted. High-priority research recommendations are put through a second prioritisation process at NICE.

7.8.4 Other research recommendations

Important research recommendations which fall outside the scope of the guidance are communicated to research and development funders such as:

- National Coordinating Centre for Health Technology Assessment
- NHS Service Delivery and Organisation Programme
- Medical Research Council (MRC)
- Economic and Social Research Council (ESRC)
- DH Policy Research Programme (PRP).

7.8.5 Equality and diversity

Research recommendations should aim to address any gaps in the evidence base in relation to the groups identified in the Equality Act (or groups who are particularly disadvantaged with respect to the topic under consideration).

7.9 Further reading

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

Michie S, Johnston M (2004) Changing clinical behaviour by making guidelines specific. *BMJ* 328: 343–5

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

Nuffield Council on Bioethics (2007) Public health: ethical issues. London: Nuffield Council on Bioethics

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

Schunemann HJ, Best D, Vist G et al. (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

Tannahill A (2008) Beyond evidence – to ethics: a decision making framework for health promotion, public health and health improvement. *Health Promotion International* 23 (4): 380–90

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

8 Testing the recommendations: fieldwork

8.1 Introduction

This chapter outlines the fieldwork phase of public health guidance development. During this phase, the draft recommendations are tested with policy makers, commissioners and practitioners to see how easy they are to implement. This chapter describes the:

- aims and guiding principles of fieldwork
- commissioning process
- approaches used
- way ethical approval is achieved
- standards used
- analysis of data
- structure of the fieldwork report.

It concludes by describing how the fieldwork findings are used to finalise the recommendations.

8.2 Aim of fieldwork

The fieldwork phase tests how easy it will be for policy makers, commissioners and practitioners to implement the draft recommendations and how the recommendations will work in practice. Practitioners may include members of the community, volunteers, parents and carers as well as professionals such as GPs, nurses and teachers.

Sometimes the views of people who are affected by the recommendations ('end users') will also be sought, either at this stage or earlier in the process. This is quite separate from fieldwork and would be commissioned as primary research or a separate 'end user consultation'. (Please refer to chapter 3 for more details.)

Practitioners' experience and views are used to fine-tune the recommendations. The aim is to ensure they are understood and interpreted as the Public Health Interventions Advisory Committee (PHIAC) or the Programme Development Group (PDG) intended, even without supporting information. This is important as they should stand alone: the quick reference guide and the implementation tools developed for the guidance contain little supporting information.

8.3 Guiding principles

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)
- 'Effective action briefings' developed by the HDA's public health collaborating centres during 2004/5
- an evaluation of HDA work commissioned by the Centre for Public Health Excellence (CPHE) in 2005
- literature reviews on qualitative research methods and approaches to use for market and social research, commissioned by the CPHE in 2005
- the experiences of the CPHE project team in commissioning and managing fieldwork from 2005–8.

Recommendations developed by PHIAC/the PDG are based largely on evidence presented in the reviews, particularly evidence about the effectiveness and cost effectiveness of interventions and programmes. This allows assessments to be made about the plausibility of an intervention or programme, under different circumstances and with different populations.

However, the context in which practitioners will be implementing the recommendations also needs to be considered. The aim is to determine: how feasible the recommendations are, what challenges practitioners will face, what opportunities exist to develop practice and what support might be needed. Practitioner knowledge and experience is gathered during fieldwork to provide this type of evidence.

The following principles underpin the fieldwork phase.

- Successful implementation depends on evidence-based recommendations informed by practical experience.
- Practitioners know what is relevant to their current practice and the policy context.
- Practitioners know what is feasible in practice.
- High-quality fieldwork elicits practitioners' knowledge in a transparent, reliable and systematic way.
- Draft recommendations are presented as indicative only of what might work.
- Fieldwork allows participants' needs to be taken into account in the final recommendations.

- Fieldwork seeks to test the impact of draft recommendations on inequalities in health, equity and diversity.

The exact questions participants are asked will vary from topic to topic, but **three areas always need to be considered**:

- 1. Content of the recommendations** Are the recommendations relevant and appropriate for the groups responsible for delivering them? Are they clear and easy to understand for all these groups?
- 2. Practice** What is current practice in the area? How might these recommendations build on or change current practice and/or service provision? What are the implications of this? Should any other elements be covered in the recommendations?
- 3. Impact** What are the barriers to, and opportunities for, implementing the recommendations? What further resources, training or support might be needed? Are the recommendations relevant to other professional groups? What is the best way to get information to the range of professional groups involved? Are the recommendations sustainable? How would participants prioritise the draft recommendations?

8.4 Fieldwork commissioning

The CPHE project team should commission an academic or research organisation to carry out the fieldwork. This organisation should be separate from the review team involved in compiling evidence reviews for the guidance, unless there are exceptional circumstances. (An exception might be made if, for example, specific expertise in the topic, or access to specialist networks, is needed). However, the review team may be asked to help the fieldwork contractor, for example, by generating a list of participants.

The fieldwork contractor should have a good track record in health-related qualitative and/or participatory research and, ideally, should have experience in the guidance topic area.

The commissioning process should adhere to NICE's Standing Financial Instructions. This involves developing a project specification, inviting expressions of interest, issuing invitations to tender and selecting a contractor based on clear and auditable criteria.

8.5 Approaches to fieldwork

Fieldwork is a qualitative exercise; a range of views are required and it can involve a number of methods. The CPHE project team and PHIAC/the PDG should consider the

choice of method(s) carefully, taking into account the topic, the professional groups involved and other issues. It may include the use of groups, one-to-one or paired in-depth interviews or surveys. In some cases – for example, if a range of professional groups are involved – a combination of approaches may be used.

The fieldwork contractor should agree the approaches to use with the CPHE project team. Similarly, the methodology and any questions or support materials used must be developed and agreed with the CPHE project team.

8.5.1 Group-based methods

Group-based methods include focus groups, participative workshops and 'virtual' (electronic) groups. These may be appropriate when:

- potential participants have clear professional 'identities' and the 'field' is well established
- NICE can contact enough professionals in a geographical region to set up a focus group or workshop
- the issues discussed are unlikely to be confidential or professionally sensitive and anonymity will not be necessary.

8.5.2 One-to-one or paired in-depth interviews

Interviews may be carried out face-to-face, by telephone or electronically. They may be appropriate when:

- it is not possible to get groups of professionals together because it's a relatively new area, the number of possible participants is limited or there are geographic or time constraints
- the issues discussed are likely to be confidential or professionally sensitive and anonymity may be needed
- in-depth responses are needed.

8.5.3 Surveys

Group-based methods and interviews (as outlined in sections 8.5.1 and 8.5.2) are the best way to find out participants' opinions. But they are not suitable in all circumstances, for example, because of the sensitivity of the topic, confidentiality issues, or difficulties in recruiting participants. In such cases, surveys that use semi-structured and open-ended questions could be more appropriate. Surveys may be carried out by telephone, electronically, on paper or by using vote casting or polling.

Surveys gather opinions in a quick, less obtrusive manner than group-based approaches and interviews. In addition, the responses can be quantified. But they do not allow the same depth of exploration and, generally, should only be used if other methods are unsuitable.

8.6 Ethical approval

In principle, fieldwork falls into the category of 'service evaluation' and so is outside the remit of NHS research ethics committees. However, the CPHE project team and the fieldwork contractor should consider the ethical issues each time fieldwork is planned. (If there is any doubt, the contractor should consult www.corec.org.uk). The fieldwork contractor is responsible for seeking ethical approval, if required.

8.7 Quality standards for fieldwork methods

The fieldwork contractor should agree any changes to the protocol for fieldwork with the CPHE project team before the fieldwork starts.

8.7.1 Fieldwork participants

Participants should be chosen to represent a broad range of professional groups in the statutory, non-statutory and voluntary sectors. This may include people who work with the target populations covered by the guidance and other users of the guidance, such as service commissioners. It may also include professionals working indirectly to promote the aspect of health covered by the guidance. Please note, fieldwork participants do not have to be from an organisation that is registered as a NICE stakeholder.

Equality issues should be fully considered when choosing fieldwork participants. This may mean getting a representative spread of professionals, but it may also mean focusing on participants with recent experience of working with disadvantaged groups. The approach should be based on the content of the recommendations, whether or not they refer to the whole population or subgroups of it and service delivery and policy issues.

Participants can include commissioners, policy officers, outreach workers, the police, representatives of the emergency services and the probation service, GPs, health visitors and educational welfare officers.

Participants should also:

- have relevant experience and knowledge of the guidance topic
- operate mainly (but not exclusively) in a regional or local capacity.

8.7.2 Sampling

Sampling should be guided by the topic. It will depend on the:

- professional groups identified as being responsible for taking action (listed in the 'Who should take action?' part of the recommendation)
- ultimate beneficiaries of the guidance (listed in the 'Who is the target population?' part of the recommendation)
- scope
- research questions
- inclusion criteria for the evidence reviews.

'Snowballing' (gathering participants via other participants or networks) and purposive or other non-random techniques may be used to ensure all relevant professional groups are represented.

Random sampling (randomly selecting participants from the relevant professional groups) or quota sampling (selecting a fixed number of participants, randomly or purposively from these groups) may be useful for large-scale surveys. Random and quota sampling may also be useful where there are a large number of professionals, but there are not enough of them in each relevant geographical area.

The fieldwork report should explicitly state how participants were identified, sampled and recruited. The fieldwork contractor should ensure the final sampling frame and sample take account of the equity focus. It should be agreed with the CPHE project team.

8.7.3 Conducting the fieldwork

Once the method(s) has been chosen (see section 8.5), a summary of the issues to be covered should be developed. The summary should be based on the draft recommendations (and related evidence statements) and the three key areas of inquiry outlined in section 8.3. The quality of the fieldwork approach (including sample and method[s] selected) should be quality assured by the CPHE project team. The team should:

- brief the fieldwork contractor in detail before work begins
- agree final documents and comment on draft recruitment letters
- help develop topic guides (summaries of the recommendations and key questions for discussion)

- agree interview schedules or sampling frames and samples, and other supporting materials
- discuss how to get participants who work with key or vulnerable groups involved
- attend fieldwork groups and/or observe interviews (where possible)
- have access to transcripts of all data
- discuss and agree techniques for data analysis and themes for data presentation
- comment on the fieldwork report before the final draft is submitted.

8.7.4 Group-based approaches

The criteria for organising group-based fieldwork are outlined below. They can be adjusted to accommodate a particular fieldwork process, although the fieldwork contractor should agree any deviations from the set criteria with the CPHE project team. For further information about preliminary work in this area, see Kelly et al. (2004).

- The final sampling frame, recruitment and sample must be agreed with the CPHE project team and must take into account any planned comparisons between, for example, professions or geographical areas.
- At least six fieldwork workshops or focus groups should be convened for guidance produced using the programme development process; at least four should be convened for guidance produced using the intervention process. These should take place in more than one geographical region and may be up to a day long. If it is not feasible to organise this many workshops or groups, the decision on how many should be convened must be agreed with the CPHE project team.
- Independent professional facilitators should be selected to lead each meeting. A background in the topic under consideration is not essential, but it is essential to have knowledge of the public health community and the methods used to translate research evidence into practice. Facilitators also need an understanding of the wider public sector (and other sectors, as appropriate, according to the range of practitioners attending the fieldwork sessions). In addition, they need some knowledge of public health as a discipline.
- A sample of up to 35 professionals (not including researchers) with a remit for the relevant public health topic should be invited to each meeting. If it is not feasible to invite 35 participants, the decision on how many to invite must be discussed and agreed with the CPHE project team.

- If it suits the needs of the project, separate workshops or focus groups can be arranged for different professional groups. This will depend on the number of participants and should be agreed with the CPHE project team.
- For some topic areas, researchers need to be included in the fieldwork. In such cases, a separate meeting should be convened for them, using the same processes. This should be agreed with the CPHE project team.
- Topic guides, prompts or supporting materials (such as the draft recommendations, supporting evidence statements and the key areas of concern – see section 8.3) must be developed in collaboration with, and agreed by, the CPHE project team.
- A member of the NICE implementation team should attend at least one fieldwork meeting.
- If possible, a member of the CPHE project team should attend each workshop or focus group to answer any technical questions and present the evidence statements and draft recommendations.

8.7.5 One-to-one or paired, in-depth interviews

The sampling for one-to-one or paired, in-depth interviews should follow the procedures outlined in section 8.7.2. Specifically:

- Sampling frames, techniques and recruitment must be agreed with the CPHE project team.
- Samples and sample sizes must be agreed with the CPHE project team, taking into account any planned comparisons of professional groups that may be needed.
- Interview schedules must be developed in collaboration with, and agreed by, the CPHE project team.
- Any prompts or supporting materials (such as the draft recommendations, supporting evidence statements and the key areas of concern – see section 8.3) should also be developed with, and agreed by, the CPHE project team.

Interviews may be structured or semi-structured, depending on the topic and the professional groups involved. Semi-structured interviews allow complex or difficult issues to be explored and so are likely to be more useful than a fixed-format interview. They should focus on the draft recommendations (and related evidence statements) and the three key areas of inquiry, as outlined in section 8.3.

Individual or paired interviews are usually more expensive to set up than group work, and the need for in-depth or individual contact should be weighed against the available resources at the planning stage.

8.7.6 Survey methods

If survey methods are used (see section 8.5.3), sampling and recruitment should follow the principles outlined in section 8.7.2. The fieldwork contractor should agree the approach, the sampling frame, final sample, how the survey will be done, and survey questions and/or the use of supporting materials with the CPHE project team.

Distribution and day-to-day management of any surveys used is the responsibility of the fieldwork contractor.

8.7.7 Recording groups and interviews

The way groups and interviews are recorded depends on the method(s) used. For example, software that automatically produces transcripts is available for online focus groups. This would differ from the way a 'traditional' face-to-face focus group would be recorded. The plenary discussions, group work and interviews should all be recorded (for example, on tape or digitally and then transcribed, or by using a scribe). Previous experience has shown that stenography is the best way for a scribe to record the points. All participants should consent in writing to the recording – and to its use in discussions and group work.

If a scribe is used, they should accurately record points raised against each question (mediating factors, barriers and solutions). They may also categorise each point according to whether it relates to strategy and policy development, commissioning, management or individual practice.

8.7.8 Structure of the fieldwork

The structure of fieldwork events depends on the method(s) adopted. Presentations may be used, for example, to give an overview of the recommendations (and if appropriate, the guidance document). In such cases, they should be kept succinct to make best use of time and strike the right balance between passive and active participation.

Topic guides, generated by the contractor (and approved by the CPHE project team), should be circulated to fieldwork participants before or at the beginning of sessions to facilitate discussions.

Workshops, focus groups and interviews should be based on the structure outlined below. The structure will differ for surveys, but the elements outlined below should still be covered. If the fieldwork contractor wants to make any changes to the content and structure of sessions, it should agree them beforehand with the CPHE project team.

8.7.8.1 Session 1: purpose of the meeting

The lead facilitator explains the aims and objectives for the day and the values underpinning NICE's fieldwork process (see sections 8.2 and 8.3). 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 target population. The initial presentations may take around 20 minutes.

8.7.8.2 Session 2: 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 may last for up to an hour.

8.7.8.3 Session 3: appraisal of draft recommendations and evidence statements

Group work can be run in different ways, depending on the project. For example, a large group convened on the same day can be subsequently divided into four or five multidisciplinary groups, each working with a facilitator. Alternatively, different sessions set up on different days can be arranged for each professional group – each with a facilitator.

If interviews (rather than groups) are being used, a series of one-to-one or paired sessions may also be set up.

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?'

Social and marketing research techniques can sometimes be useful (such as role play). It can also be useful to develop tools to help participants assess the feasibility and impact of each recommendation (for example, electronic key pads or q-sort techniques to help prioritise and sort sets of standard statements). Any techniques or tools used would need to be agreed with the CPHE project team.

Participants should also be asked to address the implications for health inequalities and generally for their own practice or profession. To conclude, discussion could consider barriers to – and facilitators of – 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 information services
- performance management
- prioritisation of recommendations
- examples of local good practice that may support the recommendations.

8.7.8.4 Session 4: feedback

At the end of group sessions, the facilitator may provide plenary feedback about the participants' view on implementation barriers, opportunities and solutions. Case study templates should be distributed for participants to note any points that have come up during the day or to submit case studies of local 'good practice'. These should be collected at the close of the session, or returned to the team by a specified date. Forms should be clearly marked with instructions for completion and return.

8.7.9 Evaluation and follow-up

At the end of each workshop, an evaluation of it may be completed by all participants. The facilitators and members of the CPHE project team should then liaise to share notes and transcripts. The fieldwork contractor should use these as a basis for the fieldwork report. All original notes should be retained. The draft fieldwork report should be circulated to all participants to check for accuracy. If separate reports are produced for each event, participants should check the report about the event they attended. These individual reports should form the basis for the fieldwork report considered by PHIAC/the PDG.

8.8 Fieldwork analysis

Fieldwork analysis is dependent on the method(s) adopted but the following points are a guide. The fieldwork contractor should agree the way data are presented and analysed with the CPHE project team.

8.8.1 Data presentation

The fieldwork groups' discussions and/or interviews should be transcribed in full (electronically or by hand). If a survey approach has been used, responses should be collated, transcribed or recorded in full.

8.8.2 Analysis

The methods used for the fieldwork will affect how the analysis is done.

- For group-based fieldwork, data analysis should begin as soon as possible, and preferably in time to be included at the next fieldwork session. This approach may also be feasible if interviews and surveys are being used.
- For one-to-one interviews and surveys, analysis is usually carried out at the end of data collection.

Analysis may be performed using qualitative research software, or by hand, but the method should be fully reported in the fieldwork report.

The fieldwork data should be broken down into common and consistent themes, framed by the research questions, using a content analysis approach. Usually, one researcher should prepare an initial analysis. This should be verified by 'blind' coding and sorting of a sample of the transcript by a second researcher. For examples of this kind of analysis, see part three (chapters 7–13) of Silverman (2004) or Ritchie and Spencer (1993).

Once the analysis is complete, participants' quotes should be selected to illustrate each theme. These quotes should be coded to keep participants anonymous and to allow the quotes to be distinguished (see section 8.9.1).

8.9 Fieldwork report

The fieldwork report sent to PHIAC/the PDG is a summary of analysed fieldwork data and key points arising from it. A copy of the fieldwork report should also be sent to all participants.

8.9.1 Style and transcription notation

The fieldwork report should describe the aims of the fieldwork, the methodology used and the findings, drawing conclusions about how the guidance can be improved. The main section should cover the findings, summarising the emerging themes. It should be illustrated with verbatim quotes from participants as follows:

- Quotes should not be edited, other than to clarify where text is not clear. If 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 double quotation marks (" "). Longer quotes should be presented as inset paragraphs with double quotation marks.
- If words from the quote are omitted, they should be replaced by '(...)', but the omission must not alter the meaning of the original quotation.
- Quotes should be coded to keep the participant anonymous and allow for the distinction between different participants' comments.
- As with data from clinical trials, transcripts should be kept for at least 5 years (see www.ct-toolkit.ac.uk).

8.10 Using fieldwork findings to finalise recommendations

The fieldwork contractor should present a summary of the findings to PHIAC/the PDG. It should use this information to refine and prioritise the recommendations after the consultation. (This includes making them more specific for different groups of practitioners, where appropriate.)

For further details about developing and prioritising recommendations, see chapter 7.

8.11 Equality and diversity

Equality and diversity issues should be considered at every stage of the fieldwork process – from commissioning the contractor to finalising the fieldwork report. For example, the fieldwork contractor should make every effort to ensure equality and diversity issues are considered when generating the fieldwork sample and the questions to be asked. These

issues should also be considered when deciding on the approaches to use. In addition, fieldwork should specifically seek to determine:

- Does the guidance avoid unlawful discrimination?
- Are there ways in which the guidance could better promote equality?

For further information, see section 7.6.

8.12 References and further reading

Green J, Thorogood N (2004) *Qualitative methods for health research: qualitative methods for health*. London: Sage

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. Available from www.nice.org.uk/page.aspx?o=507904

Pope C, Ziebland S, Mays N (2000) Qualitative research in health care: analysing qualitative data. *BMJ* 320 (7227): 114–16

Ritchie J, Spencer L (1993) Qualitative data analysis for applied policy research. In: Bryman A, Burgess R editors. *Analysing qualitative data*. London: Routledge 173–94

Silverman D, editor (2004) *Doing qualitative research: a practical handbook*. London: Sage Publications

Tashakkori A, Teddlie C (2002) *Handbook of mixed methods in social and behavioural research*. London: Sage

Appendix A Conceptual framework for the work of the Centre for Public Health Excellence (CPHE)

Introduction

This appendix describes NICE's conceptual framework for public health and health equity. This framework informs CPHE methods and processes, including the economic analysis, which will be used to produce guidance as described in the second edition of the process and methods manuals.

There are many different models and frameworks used to describe public health and the ways in which the health of the population is shaped (see for example, Cockerham 2007; Evans and Stoddart 2003; Krieger 2008; Levine et al. 2004; Solar and Irwin 2007; Starfield 2006; 2007). This paper draws on these in various ways but it also develops its own particular approach. The framework also draws on work undertaken by CPHE for the World Health Organization (WHO) as part of the WHO commission on the social determinants of health (Bonney et al. 2007; Kelly et al. 2007), and on the NICE guidance on behaviour change (NICE 2007).

The subject matter of public health is broad and diverse. It involves disease prevention, health promotion, protecting individuals and populations from hazards, and it is concerned with health improvement. It has a population rather than an individual focus. It draws on social models of health as well as biomedical ones. Its discipline base includes epidemiology, medical statistics, medical informatics, health psychology, management, medical sociology, health economics, medical anthropology, medical geography, political science and infectious and communicable disease, among others.

The practice of public health in the UK is also diverse. It involves the management of the public health service at district, regional and national levels, responses to major incidents, emergency planning, managing outbreaks of infectious disease, health informatics, data handling, collection and interpretation, certain aspects of environmental and occupational health, smoking cessation, health economics, health services research, managing some service delivery, the provision of some services in local government as well as health protection, health education and health promotion.

The people who practice public health are both medically and non-medically qualified and work in NHS and non-NHS settings. This variation is reflected in some of the different definitions of public health and health promotion. According to the Institute of Medicine (1988), the role of public health is to contribute 'to the health of the public through assessment of health and health needs, policy formation, and the assurance of the availability of services'. The Faculty of Public Health in the UK defines it more broadly as 'the bigger picture' involving actions 'to

promote healthy lifestyles, prevent disease, protect and improve general health, and improve healthcare services' (2006). Sir Donald Acheson, former Chief Medical Officer defined it as the science and art of preventing disease, prolonging life and promoting health through the organised efforts of society (Acheson 1988) The International Union of Health Promotion and Education in its document 'Shaping the future of health promotion: priorities for action' (2007), define health promotion as aiming 'to empower people to control their own health by gaining control over the underlying factors that influence health. The main determinants of health are people's cultural, social, economic and environmental living conditions, and the social and personal behaviours that are strongly influenced by those conditions.' An earlier generation of scholars found the definition similarly broad (See Downie et al. 1990 for a review).

Given this diversity the task of producing public health guidance, as NICE has responsibility to do, is very wide-ranging and has to meet the needs of disparate audiences. So finding a way to encapsulate this variation in such a way that is both coherent and strategic is important. This paper develops both a coherent and strategic framework.

The conceptual framework is based on a number of principles. These are as follows. First, that there are determinants of health and disease which are much broader than, but include, biomedical causes. Second, these determinants operate in highly patterned ways which reflect inequalities in society. Third, the determinants work through causal pathways to disease. Fourth, the causal pathways help to identify ways of preventing and ameliorating disease. Fifth, there are also causal pathways for the promotion of health. Sixth, positive and negative causal pathways cross physical, biological, social and psychological boundaries.

In the first edition of the manuals produced by NICE for public health, a simple and pragmatic definition of public health was adopted. Public health guidance was defined as being about the promotion of good health and the prevention of ill health and as being for those working in the NHS, local authorities, and the wider public and voluntary sector. The methods manual noted that:

'...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... NICE public health guidance considers a variety of approaches, from traditional health education and public campaigns to community development' (NICE 2006).

The first NICE public health process and methods manuals recognised the wide spectrum of determinants of health. These included social, economic and environmental factors, through to individual choice and ease of access to services. The methods manual also noted that recommendations could be made at population, community, organisational, group, family or individual level, as appropriate (NICE 2006).

The multi-disciplinary base of public health was also described in the original manuals. It was stated in the methods manual that NICE public health guidance would be 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.

The determinants of health

The unifying factor in topics of public health interest is the linkage of social, psychological and biomedical phenomenon. The core of the NICE conceptual framework for public health deals with this linkage in the form of causal pathways from the determinants of health to the biological changes in the human body. As a result of differential contextual stimuli and their respective interactive chains, the cells in the human body may behave differently according to the social position someone occupies, the country they live in, the global political situation around them. The cells behave differently according to the job a person does, according to their experience of class, gender and ethnic relations, according to their education, and according to a range of social factors which affect them over their life course. Their genetic structure and their immunity, their nutritional status, their resilience, their ability to cope – act as mediating factors, but there are causal pathways from a range of determinants to biological structures in the individual human body. In the areas of particular interest to public health pathological changes in the human body occur, but, and this is of fundamental importance, in highly patterned ways in whole populations or sub-population groups. Both the pathologies **and** their patterning have causes. In other words, social and biological causes can work in tandem as well as interactively. The underlying principle is that the origins or the causes of patterns of disease **and** their social patterning are mainly to be found in the social determinants (Kelly et al. 2007) (see also [Hamlin 1995] for a discussion of the nineteenth century controversies over these matters between two public health pioneers, Edwin Chadwick and William Farr). As will be shown, the link between the determinants and the disease outcomes as the early pioneers realised, is neither a simple nor a single causal pathway. This remains the case today.

This is not to say that there are not biological disease pathways, nor that some aetiological processes are only or principally biological in operation. The aetiology and the biological manifestations of disease may be described very precisely and in ways which permit diagnostic clarity and therapeutic action. However, for a number of very important causes of mortality

and morbidity, social factors play a significant role in the aetiology. The public health task is to describe that role. This conceptual framework helps to do that.

The ways in which the determinants of health operate is an area of considerable research interest. This has received particular attention with respect to the determinants of health differences, health inequalities and health inequities in populations. Much is known. It is clear, and has been so for a century and a half, that at population and individual level poor health is linked to social and economic disadvantage (Checkland and Lamb 1982; Frazer 1947). The distribution of income, employment, education, housing and environment links to inequities in health (Graham 2000). Social disadvantage and marginalisation do likewise (Braveman 2003; 2006). However, while the general relationship between social factors and health is well established (Marmot and Wilkinson 1999; Solar and Irwin 2007), the relationship is not precisely understood in causal terms (Cockerham 2007; Link and Phelan 2005; Shaw et al. 1999).

At least four groups of theories have been proposed to explain inequities in health and its relation to socioeconomic position and so the relationship between health outcomes and the social determinants of health. The **materialist/structuralist** theory proposes that inadequacy in individual income levels leads to a lack of resources to cope with stressors of life and so causes ill health (Frohlich et al. 2001; Goldberg et al. 2003; Macintyre 1997). The **psychosocial** model proposes that discrimination based on one's place in the social hierarchy causes stressors of various kinds which lead to a neuroendocrine response that causes disease (Evans and Stoddart 2003; Goldberg et al. 2003; Karasek 1996; Siegrist and Marmot 2004). The **social production of health** model is based on the premise that capitalist priorities for accumulating wealth, power, prestige and material assets are achieved at the cost of the disadvantaged. The **ecosocial** theory brings together psychosocial and social production of health models, and considers how social and physical environments interact with biology and how individuals embody aspects of the contexts in which they live and work (Goldberg et al. 2003; Krieger 2001). It builds on the 'collective lifestyles' approach and the neo-Weberian theory that lifestyle choices are influenced by life chances defined by the environment in which people live (Cockerham et al. 1997; Frohlich et al. 2001; Weber 1948). (For a review see Cockerham 2007).

Rather less attention has been focused on the social determinants of health improvement or health promotion. However, some years ago Antonovsky (1983; 1984; 1985; 1987) proposed the theory of salutogenesis. Salutogenesis literally means the origin of health. He offered an alternative to what he saw as the pathogenic approach (that is, an approach focusing on the origins of disease and system breakdown). Pathogenesis was, according to Antonovsky, the dominant paradigm in medicine, and the behavioural and social sciences. Antonovsky's salutogenic argument majored on the factors which protected people from trauma and the generally noxious physical, social and biological world which they inhabit. This has both spawned and dovetailed with a range of research focusing on coping, resilience, adaptation, social capital and social support. So although there is more literature on the negative impacts

on health of social determinants, the health enhancing nature of social factors will also be part of the conceptual framework to be outlined here. This distinction also highlights that the factors causing health improvement are not necessarily the converse of those causing disease. Consequently, the causes of health inequities are not the same as the causes of health. The framework presented here allows for the full range of causes to be explored.

What is generally missing in both pathogenic and salutogenic accounts, is the underlying certainty about cause and effect associated with some other branches of science including clinical medicine. We see instead mostly associational or probabilistic types of explanations (Link and Phelan 2005; Mechanic et al. 2005). Clinical medicine has its own uncertainties in relation to causation. Aetiology is sometimes unknown, tenuous, partial and often multifaceted, and morbidities are frequently present in ways which are not typical and as co- or multiple morbidities. The effects of treatments are also sometimes uncertain (Chalmers 2004). The disease categories used by medicine to describe pathology are nominalist rather than essentialist and therefore they change and evolve over time, reflecting new knowledge and understanding. Data and evidence are surrounded by uncertainty (Griffiths et al. 2005), and in the end the skill of the clinician is in working through and with these uncertainties, not resolving them.

Despite the uncertain and contingent nature of the understanding of biomedical processes, medicine operates successfully with an underlying epistemological principle: health outcomes have preceding causes and the isolation of cause is the basis of effective preventive or curative intervention. This logic can be applied, subject to all the uncertainties just outlined, to public health. The task is to map the determinants and the biological processes and the interaction between them.

The causal pathways of interest here traverse a number of levels of analysis which academic disciplines traditionally keep separate. However, biologically, sociologically and psychologically plausible pathways need to be developed with reference to each other. This will allow for the development of explanatory systems which cross the traditional discipline boundaries and the different levels of explanation. Sociology must stop its explanations ending at the level of the social; psychologists must move beyond a focus only on the individual and on treating social factors (if they do so at all) as residual characteristics of individuals; and medicine must draw itself away from the fetishism of the gene and acknowledge more readily the powerful physical, social and psychological forces impacting on the biology of human life.

Within the existing literature there are many models and theories which help to provide a potential way of mapping the determinants. It is possible to identify some of the necessary and the sufficient conditions involved in causation but their nature, under what circumstances, and how they operate is not always very clear. The core candidates can be listed relatively easily because the research has explored them at length:

- poverty
- hunger

- occupational exposure to hazards
- occupational experience of relations at work
- the social and economic effects of ageing
- the experience of gender relations
- the experience of ethnic relations including direct experience of racism
- home circumstances
- the degree and ability to exert self-efficacy especially through disposable income
- dietary intake
- habitual behaviours relating to food, alcohol, tobacco and exercise
- position now and in the past in the life course
- the accumulated deficits associated with particular life courses
- schooling
- marital status
- socioeconomic status and social class.

These are the media through which the external world impacts directly on life experiences and exerts direct effects on the human body. They in turn are linked to macro variables like the class system, the housing stock, the education system, the operation of markets in goods and labour, and so on (see Solar and Irwin 2007). In the next section a classification of these media is described in the vectors of public health.

Vectors of public health

Public health has a unique contribution to make to the vectors through which the causal pathways operate. In these vectors the interaction of the structures of society including material, physical and economic ones, interact with individual human behaviour and biomedical processes in the human body.

The vectors are not distinct but are part of an overlapping and interacting set of forces. But for simplicity's sake they can be considered separately.

Population vector

The population vector includes those elements which affect, impinge, or impact on the total population. States, governments, supra-national organisations and corporations play critical roles. The elements in this vector include the structure of the state, and concomitant legislation, taxation, and the rules and regulations it uses to manage relations within civil society and between civil society and the state. The degree to which the state permits democratic engagement, political and economic freedoms, free speech, the degree to which it is itself fragile or secure, corrupt or efficient sets a context and directly determines positive or negative health outcomes as well as configuring a range of other vectors of health (Espelt et al. 2008). In the UK things like legislation to ensure the wearing of seat belts, the ban on smoking in public places, legislation prohibiting the sales of cigarettes, tobacco and alcohol to persons under the age of 18 are elements in the population vector. Another example of a public health element in this vector from the UK, is enshrined in formal rules and regulations governing road use in the 'Highway code'. In societies which are totalitarian, authoritarian, dictatorial, and the state is not regulated by principles of equality before the law, the impacts on the health and wellbeing of the population are generally in varying degrees malign.

A very significant element in the population vector is the economy including the size and distribution of gross domestic profit (GDP) and incentives offered through the market as well as barriers to opportunity enshrined in market arrangements and practices. Incentives in the market and the regulation of market failure are two very significant aspects of the economy and its operation is fundamental to human health. Economic growth, rates of employment and economic freedom promote market opportunity as well as cause damage when people lose their jobs or businesses fail. These things have direct effects on the livelihoods and life chances of people. The extent to which markets are regulated and managed, and the degree to which protection is offered against the vicissitudes of the market are fundamental. As part of the regulatory structure of the economy the taxation system is core. Related public health elements in the vector in the UK include the duty on beer, cigarettes, wine and spirits. The general fiscal structure especially its regressive or progressive qualities, the amount of VAT on food and possibly on different types of food are also good examples of elements in the population-level public health vector.

Legislation and rules will be mediated by the degree to which laws and regulations are enforced and are complied with. Just because the population vector is supposedly universal in its reach, it does not follow that there will always be population-wide epidemiological effects. There may be matters of degree, in that some sectors, groups or individuals may seek to avoid the impact of these actors and may be able to resist the effects, but potentially at least such factors reach the total population.

Environmental vector

The second vector is environmental. Environmental elements in this vector include all those potentially noxious substances and particles which might be present in macro and micro-environments, like dust, lead, asbestos, and other things associated with industrial, agricultural, transport or construction activities. They may be present in the micro-environments of homes or workplaces or in atmospheres in the wider environment. The environmental vector also embraces meteorological, tidal, and geophysical hazards such as radiation, floods, and drought as well as longer-term climatic threats and dangers. It includes microbiological agents, germs, viruses, bacteria, prions and other biological stressors. It includes some psychological stressors and mediators like noise, working conditions, and so on. It also includes transport systems, buildings, homes and the structural organisation of the workplaces and schools. It includes the systems of sanitation and provision of clean water.

Some of this sits as the specialist concern of environmental health professionals, building planners and engineers, microbiologists, geophysicists and meteorologists. The interest of public health is in the detail of these specialties. But it is also the totality of the environmental elements described that provide both a macro and micro context for the world of experience, vulnerability and risk. These factors will be mediated, in part by the actions of the state in the population vector, but also by various economic actors like businesses and trades unions. Some of the hazards in the environment are more amenable to amelioration and control than others, through regulation and management. Others, like climate change, are less easily amenable to control through regulation and legislation.

Social vector

The social vector consists of all those elements and factors which are linked to social, economic and cultural circumstances (and therefore are closely associated with the population vector). They also include the nature of relationships between social groups in civil and economic society. The conventional way of describing the social vector in public health is by way of describing the epidemiological differences between social groups. So class, status, ethnicity, age, gender, and disability, religion, caste, tribe are typically the familiar axes of social differentiation, and also the ones where epidemiologically there are plentiful data showing the different health states of different social classes, different ethnic groups and men and women. These social categories are well known to be correlated in a graded way with almost all health outcomes.

For the purposes of understanding how these groupings work in this vector, it is helpful to see beyond the statistics and the epidemiological aggregations, to the relationships of power, discrimination, disadvantage and exploitation that are the relational correlates of social position. Similarly, these groupings contain within them patterns of social behaviours or what is sometimes called lifestyle. There are groupings or clusters of ways of living associated with social position

which are good ecological predictors of future health outcomes and states. Much of the work of medical sociology and social epidemiology has been about plotting the excess morbidity and mortality associated with these social positions and the sub categories of these positions. However, these groupings are not just the manifestation of the way social factors determines health, they are the core of the operation of these elements in this vector (Cockerham 2007; Link and Phelan 2005).

Organisation

Closely linked, practically and conceptually with the social vector is the organisational vector. Most human activity which is not domestic, takes place in social organisations of one type or another. Social organisations provide much of the framework or the architecture of social life in institutions like bureaucracies, schools, factories, businesses, clubs, societies, and religious organisations. There are libraries full of detailed descriptions of the structure and functioning of such organisations (Burns and Stalker 1961; Etzioni 1961; 1964; 1968). Clearly they define important parts of the vectors associated with working and environmental factors which impinge directly on health. These were described briefly above in the vector dealing with environmental elements. What this vector does distinctly however, is to provide a causal pathway from the structure of organisations and in particular in the way they function or they behave as actors. This is most easily done in this context by describing the way the organisation of healthcare directly affects health with reference to access to – and exclusion from – services.

These patterns have been famously described as ‘the inverse care law’ (Tudor Hart 1971). Tudor Hart argued that the need for care varied inversely with the care provided. In other words, those in most need received the worst care and those in least need the best. Tudor Hart saw this as contributing significantly to health inequality. His observation is widely replicated in many healthcare settings, including those where there is no fee for service and care is free at the point of delivery, like the UK.

Tudor Hart’s observation at first seems to fly in the face of older evidence which suggested that services were of relatively minor importance when compared to sanitation, housing and nutrition (McKeown 1976). The answer to this apparent contradiction is that historically, and especially in the era of rampant infectious disease, health services probably played a relatively minor role in maintaining the overall health of populations (although they sometimes relieved suffering at the individual level). However, as technologies have improved and become more effective, services have increasingly become a critical variable in health outcomes, health experience and ultimately mortality, at population and at individual level (Bunker 2001). Therefore, services constitute an important gateway to health life chances individually and at population level. The way people can get to the whole range of care, from preventive services to acute and primary care, mediates health outcomes.

There are a number of dimensions within this organisational vector that apply to health services. The first is availability. People can only use a service if it is there. In the UK system there is universal provision. The second is entitlement. In the UK, entitlement is universal regardless of any other social or economic factor. This is not the case in market systems or others which in some way limit entitlement through other mechanisms. Even with universal provision, it does not follow that there will be universal access. So the third element is the service configuration and the way it affects access. Included here are the ways the service is organised, delivered, and the behaviour of the employees in the service to the clients and patients and to each other. Configuration also includes flexibility and responsiveness to the client group, innovation in care and new pharmaceuticals and its ability to implement new ways of working. Fourth is the relationship between the professional and managerial cadres and of both cadres to the bureaucratic or other mechanisms of organisation. In organisational terms, all of these things have a profound impact on effectiveness of care at all levels and all have a profound impact on the way that clients engage with the service (Friedson 1970).

The fifth element is the behaviour of the client groups themselves. For well-documented and rational reasons people make differential use of all types of service. They delay seeking treatment, they avoid preventive opportunities, they overuse services or use them inappropriately. They can act in ways which will not necessarily maximise the benefits they may derive individually from the service and in ways which may diminish the effectiveness of interventions at population level (Becker et al. 1977; Mechanic 1962; Rosenstock 1974).

The interaction of the vectors and human behaviour

The principal elements of human behaviour of particular interest in public health in Britain are: smoking, eating (and associated consumer behaviour), alcohol consumption, physical activity (including active travel), and sexual behaviour. This is because these are the behaviours most closely associated with disease patterns. But in order to understand the interaction between behaviour and the vectors described above, it is important to conceptualise human behaviour beyond strictly speaking health-related behaviours (NICE 2007). Although the social structure shows strong associations with patterns of health and disease, the public health conceptual framework requires not mere association, but causation and needs to embrace behaviour as a whole rather than behaviours directly associated with disease and disease prevention.

The dynamic of causation of disease lies in the interaction between agency and structure (NICE 2007). Giddens (1979; 1982; 1984) argued that society was the product of interaction between individual human behaviour and the social structure. He further argued that the billions and billions of individual human actions produce societal patterns. The patterns repeat themselves to such a degree that structures emerge. Although these structures change, sometimes gradually, sometimes rapidly, individuals are aware of them and orient their actions in line with them (and are constrained by them). The vectors described above are structural. That is, they

are components of the social structure. Those vectors (with the possible exception of physical environmental elements) are themselves the product of human behaviour and then in turn impact on it. Even the environmental vectors are considerably affected by human actions, from climate change to the mutations of viruses and bacteria in the face of antivirals and antibiotics.

It is sometimes mistakenly asserted that the structures of society determine human behaviour. It is considerably more complex than that. A completely determinist position, that is, one which fails to acknowledge the power of human agency to be creative and ingenious, inventive and non-conformist as well as more mundane choices in everyday life, is deficient. Behaviour, although patterned and linked to social structure, is still under some degree of human individual control. Behaviour is not pre-programmed according to social position. For example, in 2006, 33% of people in routine or manual occupations smoked compared with 16% of people in managerial or professional occupations. Members of all social groups exceed their energy requirements in the form of calories consumed, against energy used in the form of physical activity. Excessive alcohol consumption is not confined to the poor and manual workers. And many manual workers do enough physical activity to gain health benefits. In other words, notwithstanding well-defined patterns of behaviour at group level and strong associations between social position and health outcomes, this is neither a programming nor a deterministic effect. The social patterns of health and disease are subject to wide degrees of individual and subgroup variation. This variation is in part accounted for by the enormous variability in human behaviour. The important conceptual trick is to find a means of capturing variation but at the same time capturing the pattern. The conceptual vehicles to do this are the 'life course' and the 'life world'. These in turn provide a means to explain causation from the determinants of health to the microbiology of the human body.

'Life course' sociology and 'life course' epidemiology have accumulated a significant body of evidence which shows that from the moment of conception to the moment of death, the human organism accumulates insults and benefits (Kuh et al. 2003). In health terms, these insults and benefits are a kind of health profit and loss account which determines the health state of the individual. Some of these things are biological and are determined by the hereditary structure of the organism and the microbiological environment; others are a consequence of the vectors described above and their interaction with behaviour. They reflect the immediate physical, social, psychological and emotional environment of the growing child, and then the adult. The 'life course' approach also demonstrates that at critical points on life's journey, which are very highly socially patterned, benefits and insults can be greatly magnified, past insults can be cancelled out, and new benefits can come into play. It is also clear that these changes may be self-reinforcing, producing and reproducing patterns of health advantage and disadvantage. Those critical points on life's journey are like gateways or forks in the road, setting in train patterns that may endure and have long-lasting effects.

It is also clear that the 'life course' follows quite distinct patterns for different social groups. The trajectory through life for the child of a single mother in receipt of state benefit in public

sector housing in Scotland will be very different to that of a child born to a professional couple in Surrey, and both will be quite different to that of a Bangladeshi girl born in Tower Hamlets and an African-Caribbean boy born in Lewisham. The direction people go at each gateway has a profound effect on their future. The gateways and where they lead are markedly determined by social factors.

On life's journey the experience of benefits and insults to health occur in what some philosophers call the 'life world'. The notion of the 'life world' in the context used here was developed in the work of Schutz in particular (Schutz 1964; 1967; 1970). The notion also draws upon the work of Mead (1934). The 'life world' is a social space, which can be drawn biologically and physically but which is predominantly cognitive and subjective. It is the place where we make our own decisions, decide upon our immediate actions, judge ourselves and others, experience the social structure first hand in the form of opportunities, barriers, difficulties, disadvantage, and it is where our emotions are played out and our feelings are expressed.

Every human being inhabits his or her own personal 'life world'. At its core is the subjective self, which is experienced as a continuous 'self' existing through time and space within a more or less familiar world of places and people. Although the 'life world' is uniquely personal, it is also inhabited by others who are recognised as physically and subjectively similar to, but separate from, the self. These others who inhabit the centre of our 'life world' are those individuals whom we meet and interact with, or think about and relate to, on a recurring basis. The people with whom we share our domestic arrangements, some of our workmates and perhaps friends and family, as well as those who are not intimates or friends but whom we meet with regularly, make up this world. It is the interaction, real or imagined, on a repetitive basis which defines the inner zones of the 'life world'. The level of intimacy is not the crucial issue. It is the repetitive and routine nature of the contacts with others that is important.

Schutz (1967; 1970) conceptualised the 'life world' as a series of concentric circles. The innermost circle is the one where the everyday contacts and routines are highly predictable and are therefore taken for granted. They are salient and immediate and tend most of the time to be the most important. The more distant parts of the 'life world' are inhabited by things and people we can recognise even though we do not know them, and whom we could and would understand were we to meet and interact with them. We therefore have some sense of these persons and things but their impact on us is nil or negligible. Schutz described the concentric circles of the 'life world' as zones of relevance (Schutz 1970). The closer to the centre, the greater the relevance of what goes on there to the 'I'. The values and prescriptions of the circles closest to the centre are important. The stock of knowledge or assumptions that an individual has of those parts of the 'life world' is a crucial resource for making sense of things (Schutz 1967).

It is very important to note that the innermost circle of the 'life world' may not be (and Schutz never suggested it would be) a place that was benign and cosy. It may be violent and bullying. It

may be cold and unforgiving. It may be unpleasant and chronically difficult. It will be the place where discrimination and disadvantage, poverty and unemployment are experienced. However, it constitutes the centre of the existence of the person. 'Life worlds' change as individuals move through space and time. Groups of intimates change, children grow up, leave home and move to a more distant part of the individual's 'life world'. New people come into our orbit of friends and acquaintances. The social group in the everyday 'life world' of contacts – direct and indirect, real, imaginary or virtual – is continually in a state of flux. The possible variability is enormous.

Schutz (1967; 1970) believed that central realities and experiences of everyday life are the building blocks of social life and individual behaviour originates and is rationalised and explained in the 'life world'. The experiences and meanings attributed to disadvantage are constructed in the 'life world'. It is the point where social structure impacts on the individual. It is the highly localised manifestation of the social structure and is where that social structure is experienced, is made meaningful and constrains human action in a very direct way.

The 'life world' is the locus of experience: social, psychological and physical. The 'life world' is also about the physical space which we inhabit. It is where the social meets the biological. 'Life worlds' are the point at which stressors are moderated, mediated or exacerbated. It is the point where insults are parried or where they have their noxious effects. It is the point where vulnerabilities translate stressors into physical and emotional damage. It is where immunities – biological, physical or psychological – work their protective powers. Social disadvantage is characterised by the inability or lesser ability to control the 'life world'. Social advantage is characterised by the ability to make control of the 'life world' sustainable.

There are four types of resources that help to control the 'life world'. First, there are technical things like skills, knowledge, money and access to services and resources. Second, there are interpersonal resources constituted from the relationships, social support, safety and ease of communication. Third, there are intrapersonal resources – the ability to deal with the emotions of life and its psychological distresses with equanimity or otherwise. Finally there are the resources of being able to make sense of the 'life world', of being able to make it meaningful. If humans can do that, they seem better able to cope with the ups and downs of human existence (Antonovsky 1984; 1985; 1987).

The trajectory through the 'life course', mediated through the 'life world', is how structural factors, the vectors, determine health. The 'life world' is where the causal mechanisms of health inequities operate, and the pathways to ill health can be described. Disadvantage may be viewed as a differential opportunity (life chance) to control one's 'life world' (Weber 1948). Differences between 'life worlds' are the social manifestations of differences in physical life chances. 'Life worlds' operationalise the differential experiences of power, exploitation and access to resources. Where 'life worlds' abut, the experience of discrimination and disadvantage originates and within the 'life world' the experience of pain and suffering are located. 'Life worlds' also group together and although each one is unique, there are patternings and clusterings which produce

shared experiences and what philosophers call intersubjectivity, that is, a shared understanding and set of common meanings. It is the group properties of aggregated 'life worlds', the clustering of similar experiences that produces the patterns of disease which epidemiology eloquently demonstrates. It is because of the individual operation of the factors which are damaging to health that we can observe the causal pathway from the social to the biological. The pattern is manifest in the differential exposure and vulnerabilities to disease and conversely protection from disease which are familiar territory to public health. The summative effect is the degree of total exposure to pathogens and risks. Vulnerability may be biological, reflecting for example pre-existing nutritional status, immunological status, or illness. It may be psychological in that the ability to be resilient to stressors is at least in part a consequence of psychological processes. It will be social in that supportive social relations and economic security for example, are considerably advantageous when dealing with stressors and their absence both amplify and sometimes directly lead to the inability to cope with stressors. It is not possible to predict individual health outcomes, and the reason for this is that the agency structure system is both patterned and has enormous variability. To borrow an analogy from physics, what we are dealing with here is something akin to the uncertainty principle, and as with physics the uncertainty operates at the micro level rather than at the system level.

Equity

Whitehead describes health inequality as 'measurable differences in health experience and health outcomes between different population groups – according to socioeconomic status, geographical area, age, disability, gender or ethnic group' (1992). Inequality is about objective differences between groups and individuals measurable by mortality and morbidity. Whitehead defines 'health inequity' as differences in opportunity for different population groups which result in, for example, unequal life chances, access to health services, nutritious food, adequate housing. These differences may be measurable; they are also judged to be unfair and unjust (Whitehead 1992). Health equity is defined following the WHO Commission on the Social Determinants of Health as 'the absence of unfair and avoidable or remediable differences in health among social groups' (Solar and Irwin 2007). Health inequity is therefore defined as unfair and avoidable or remediable differences.

There are several further important distinctions which need to be included in a discussion of equity. The terms health gaps and health gradients and absolute and relative differences need to be noted. Health gaps simply refer to the difference measured between two different individuals or groups say between men and women, or social class 1 and 5. Conventionally a health gaps approach focuses on extreme differences between the most and least advantaged in a society, for example. The measure could be one of mortality, morbidity, or some subjective measure of health state. The fact of difference, aside from the social injustice involved, is of rather less interest than the trends in the differences between different time points, for example, over a period of 1, 5, or 10 years. Imagine two groups with different mortality at point time one. Then

let's assume that at point time two the difference between the two has remained the same. In absolute terms both groups have remained the same. We would also note that over time the relative difference between them had also remained the same.

However, in contemporary Britain we do not see the health states of groups remaining constant over time. What we observe is that in absolute terms all groups tend to show a long-run trend to improve. Health improvement has been the order of the day for nearly two centuries. So in absolute terms everyone is getting better. However, what also has happened has been, at least over the last 40 years or so, a trend for the health of the higher social classes to improve at a faster rate than the health of the less advantaged, resulting in the poorer becoming worse-off in relative health terms. This is because their health has not improved as quickly as those above them on the social ladder.

The other feature of health inequity is that the absolute differences at any given point in time manifest themselves as a gradient, with each group as we move up the social hierarchy being a bit better off in health terms than the one immediately below them. This gradient shows a remarkable consistency with income and social class, for example. The steeper the gradient at any given time, the greater the differences between people. However, because of the trend over time for health to improve faster nearer the top than at the bottom, the gradient tends to shift, producing relative disadvantage in spite of overall health improvement in the population. This may fluctuate, however, and absolute and relative differences can vary across time (Krieger et al. 2008).

Universal interventions which produce overall health improvement tend to exacerbate this and therefore moderated or targeted interventions are needed in order to make the rate of improvement greater in the more disadvantaged groups. The problem is that this tends to be both more difficult and less cost effective than universal programmes in which the well off make more use of available services. It is also difficult to do in practical terms because the social differences which are the constituent parts of the gradient, are not very well described, and certainly not in sufficient detail to allow pin-point accuracy with interventions to change things. Unfortunately, the factors which lead to general health improvement – improvements in the environment, good sanitation and clean water, better nutrition, high levels of immunisation, good housing – do not necessarily reduce health inequity. This is because the determinants of good health are not necessarily the same as the determinants of inequities in health (Graham and Kelly 2004). It is necessary to distinguish between the causes of health improvement and the causes of health inequities. Both operate through the vectors. There appears to be an almost inevitable process where the better off always benefit disproportionately and earlier when universal interventions are applied (Kelly 2006). Sometimes there is a catching up effect with the less well off making up ground later, but a differential remains (Antonovsky 1967; Victora et al. 2000). It may be argued that the widening differential does not matter as everyone is benefiting to some degree, so the differential is not a reason not to carry out general health

improvement. In fact the decision here will need to be made on a case-by-case basis. Sometimes in some areas of public health work universal approaches designed to produce overall population health improvement will be appropriate, in other cases a focus on the most disadvantaged, and in others a graduated approach across the gradient will be needed. This cannot be decided in advance. It is also important not to define universal and targeted approaches as simple alternatives. Hybrid actions which contain elements of, for example, universal actions with targeted follow through, will sometimes be the most appropriate.

Whether and how particular interventions will impact on subgroups in the population in particular ways is difficult to determine. The present state of the evidence base means that the data available to make such assessments are very limited. In other words, the amount of data available about differential population effects outcomes in different populations is limited because there is a paucity of outcome research in the first place and there is relatively little which deals with subgroups in ways which allow this to be done. Several academic groups are currently working towards the better construction of the evidence base, the Cochrane Collaboration for example, but the real problem lies in the primary research and its analysis. Although data are not infrequently collected on gender and socioeconomic status and age (although this is by no means universal), the analysis is usually done in such a way that these variables are used to control for confounding rather than to identify differential population effects. This is particularly so at the level of systematic review.

Building the map of public health

In this last section a classification system based on the vectors will be described to map the key areas of interest for CPHE and to develop a means of clarifying the scoping process and topic selection. The starting point for this is a list of key areas of interest for public health. The feature which links these topics together is their connection to the vectors described above and the fact that there are clear social patterns in the epidemiology and gradients in related health outcomes (Bonney et al. 2007; Cockerham 2007). From this starting point it is possible to draw up a list of key general areas – some linked to disease, some to populations, some to behaviours and lifestyle and some to available technologies. This list helps to capture the main, but not necessarily complete, areas of interest (see below).

The main public health areas of interest:

- Accidents and injuries
- Alcohol
- Cancer
- Cardiovascular disease
- Child health

- Chronic illness
- Diabetes
- Drugs
- Environmental health
- Housing
- Maternal health
- Mental health
- Obesity
- Occupational health
- Oral health
- Physical activity
- Screening
- Sexual health
- Smoking and tobacco
- Transport
- Vaccine preventable diseases.

Obviously this list overlaps with some clinical areas and some of the topics which the NICE clinical guidelines programme would concern itself with. Also, there are clear overlaps within the list: diabetes and obesity are closely related and they both overlap with cardiovascular disease, and this in turn impinges on chronic illness. Environmental health abuts and overlaps with housing and transport and so on. However, the purpose is not to build a taxonomy but rather to describe a variety of potential public areas of interest. This pragmatic approach also allows us to extricate ourselves from the argument about whether topic, population or setting ought to be the basis for public health and health promotion. It is all of these things and the list captures that.

The vectors described above: population, environment, society and organisation, can be cross-classified with the topics in the list. This allows for different nuances in respect of a topic to be articulated. So with cardiovascular disease, population-level issues directs attention to tax on cigarettes or controlling levels of salt in manufactured food for example, whereas a focus on the organisation would lead attention to the screening or case finding in primary care. To take another example, if the topic was physical activity, a focus on the environment would lead to an emphasis on the natural and built environment and its links to the ability to take exercise, while a focus on the social vector in this topic would lead to an emphasis on the ways in which, for example, young women and some ethnic minority groups find getting active very difficult for cultural and religious reasons. If transport was the topic, then a population focus would be on

regulation of traffic and vehicle and pedestrian safety, whereas an organisational focus would lead to the manner in which public systems of transport facilitate accessibility to services, for example. In each of these areas human behaviour plays a part in various ways, and in each of these areas the problem can be interrogated for its equity dimensions, and considerations as to whether the gap, the gradient and the absolute and relative values are the most germane and salient. The complete cross-classification produces the following matrix.

	Population	Environment	Social	Organisation
Accidents and injuries				
Alcohol				
Cancer				
CVD				
Child health				
Chronic disease				
Diabetes				
Drugs				
Environmental health				
Housing				
Maternal health				
Mental health				
Obesity				
Occupational health				
Oral health				
Physical activity				
Screening				
Sexual health				
Smoking				
Transport				
Vaccine preventable				

Public health activity can be delivered at a number of different levels. These are population, community, organisation/setting, group, household/family, and individual. It is important to note that outcomes are not necessarily at the same level as the intervention. So population-level interventions may have effects at individual and group level, while individual-level interventions may have outcomes at population level if done on a wide enough scale.

Each of the potential levels of delivery can be applied to each of the cells in the matrix of programmes. It is helpful then to describe the matrix in three dimensional terms where the topics are on the y axis, vectors on the x axis and the levels of intervention on the z axis.

Each of the cells in the three dimensional matrix represents a potential subset of public health interventions. Clusters of these interventions constitute programmes. Some of these programmes would constitute a cell or cells in the three or two dimensional matrix. Some programmes will map onto different parts of the three dimensional matrix. The matrix also allows us to plot the pieces of work already completed by the public health team at NICE.

Topic selection and scoping

The purpose of the forgoing discussion is to map the possible areas of NICE's public health work unified by a conceptual understanding of the causal pathways to disease. As far as topic selection goes, the principles are that when an idea is suggested as a possible topic for NICE to produce public health guidance, it will be plotted first in the long list of key topic areas. Then the topic will be assessed against the vectors and then the potential delivery areas will be considered and the behavioural and equity issues articulated. At the point when the briefing paper gets written for the topic selection panel, the briefing paper will have to determine which of the causal vectors are of prime interest. The extent to which the interaction with behaviour, the 'life world' and the 'life course' will also need to be specified, along with an account of the equity issues involved. The briefing paper will then go on to describe the programme theory and or theories of change which describe the interaction between the vector, the structure, the human behaviour, as a logic model to map the behaviour in the 'life course' and the 'life world' (Kelly 2006; NICE 2007). This will also be the basis for the initial economic modelling. It will provide a total picture of the placement of the topic on the public health map, provide a means of modelling the key behaviours, provide a means of starting the economic modelling, and provide the basis for the start of forensic searching of the evidence base.

Conclusion

This appendix has mapped the key vectors used for the determinants of health and disease that are of interest to public health. The dynamic interactions between the vectors and human behaviour and the ways they operate through the 'life course' and the 'life world' have been developed. The importance of distinguishing between the ideas of health gaps, health gradients

and absolute and relative understandings of equity have also been described. This framework, when combined with key public health areas of interest and levels of delivery, provides a means to map the potential areas where NICE can produce public health guidance.

References and further reading

- Acheson D (1988) Public health in England: the report of the Committee of Inquiry into the Future Development of the Public Health Function. London: The Stationery Office
- Antonovsky A (1967) Social class, life expectancy and overall mortality. *Millbank Memorial Fund Quarterly* 45: 31–73
- Antonovsky A (1983) Health promoting factors at work: the sense of coherence, in Kalmo R. editor *Psychosocial factors of occupational health*. Geneva: World Health Organization
- Antonovsky A (1984) The sense of coherence as a determinant of health, in Matarazzo JD, Weiss SM, Herd JA et al. *Behavioral health: a handbook of health enhancement and disease prevention*. New York: Wiley
- Antonovsky A (1985) *Health stress and coping*. San Francisco: Jossey Bass
- Antonovsky A (1987) *Unravelling the mystery of health: how people manage stress and stay well*. San Francisco: Jossey Bass
- Becker MH, Haefner D, Kasl SV et al. (1977) Selected psychosocial models and correlates of individual health related behaviours. *Medical Care* 15: 27–46 (supplement)
- Bonnefoy J, Morgan A, Kelly MP et al. (2007) *Constructing the evidence base on the social determinants of health: a guide*. Geneva: World Health Organization
- Braveman P (2003) Monitoring equity in health and health care: a conceptual framework. *Journal of Health Population Nutrition* 21(3): 181–92
- Braveman P (2006) Health disparities and health equity: concepts and measurement. *Annual Review of Public Health* 27: 167–94
- Bunker J (2001) *Medicine matters after all: measuring the benefits of medical care, a healthy lifestyle, and a just social environment*. London: The Stationery Office
- Burns T, Stalker GM (1961) *The management of innovation*. London: Tavistock
- Chalmers I (2004) Well informed uncertainties about the effects of treatments: how should patients and physicians respond? *British Medical Journal* 328: 475–6
- Checkland O, Lamb M editors (1982) *Health care as social history: the Glasgow case*. Aberdeen: Aberdeen University Press

- Cockerham WC (2007) *Social causes of health and disease*. Cambridge: Polity Press
- Cockerham WC, Ritten A, Abel T (1997) Conceptualizing contemporary lifestyles: moving beyond Weber. *Sociological Quarterly* 38: 601–22
- Department of Health (2004) *Choosing health: making healthy choices easier*. London: The Stationery Office
- Downie RS, Fyfe C, Tannahill A (1990) *Health promotion: models and values*. Oxford: Oxford University Press
- Espelt A, Borrell C, Rodriguez M et al. (2008) Inequalities in health by social class dimensions in European countries of different political traditions. *International Journal of Epidemiology*. Available from doi:10.1093/ije/dyn051
- Etzioni A (1961) *A comparative analysis of complex organisations: on power, involvement and their correlates*. New York: Free Press
- Etzioni A (1964) *Modern organisations*. New Jersey: Prentice Hall
- Etzioni A (1968) *The active society: a theory of social and political processes*. New York: Free Press
- Evans RG, Stoddart GL (2003) Consuming research, producing policy. *American Journal of Public Health* 93: 371–9
- Faculty of Public Health (2006) *Public health: specialise in the bigger picture*. London: Faculty of Public Health
- Frazer WM (1947) *Duncan of Liverpool: being an account of the work of Dr W H Duncan medical officer of health of Liverpool 1847–63*. London: Hamish Hamilton
- Friedson E (1970) *Professional dominance: the social structure of medical care*. Chicago: Aldine
- Frohlich KL, Corin H, Potvin L (2001) A theoretical proposal for the relationship between context and disease. *Sociology of Health and Illness* 23: 776–97
- Giddens A (1979) *Central problems in social theory: action, structure and contradiction in social analysis*. Berkeley: University of California Press
- Giddens A (1982) *Profiles and critiques in social theory*. London: Macmillan
- Giddens A (1984) *The constitution of society: outline of the theory of structuration*. Berkeley: University of California Press
- Goldberg M, Melchior M, Leclerc A et al. (2003) Épidémiologie et déterminants sociaux des inégalités de santé. *Revue d'épidémiologie et de santé publique* 51: 381–401

- Graham H editor (2000) *Understanding health inequalities*. Buckingham: Open University Press
- Graham H, Kelly MP (2004) *Health inequalities: concepts, frameworks and policy*. London: Health Development Agency. Available from www.nice.org.uk/page.aspx?o=502453
- Griffiths F, Green E, Tsouroufli M (2005) The nature of medical evidence and its inherent uncertainty for the clinical consultation: qualitative study. *BMJ* doi:10.1136/bmj.38336.482720.8F
- Hamlin C (1995) Could you starve to death in England in 1839? The Chadwick – Farr controversy and the loss of social in public health. *American Journal of Public Health* 85: 856–66
- Institute of Medicine (1988) *The future of public health*. Washington DC: Institute of Medicine
- International Union for Health Promotion and Education (2007) *Shaping the future of health promotion: priorities for action*. France: International Union for Health Promotion and Education
- Irwin A, Valentine N, Brown C et al. (2006) The commission on social determinants of health: tackling the social roots of health inequalities. *PLoS Medicine* 3(6): e106
- Karasek R (1996) Job strain and the prevalence and outcome of coronary artery disease. *Circulation* 94 (5): 1140–1
- Kelly MP (2006) Mapping the life world: a future research priority for public health, in Killoran A, Swann C, Kelly MP editors. *Public health evidence: tackling health inequalities*. Oxford: Oxford University Press
- Kelly MP, Morgan A, Bonnefoy J et al. (2007) *The social determinants of health: developing an evidence base for political action*. Chile: Universidad del Desarrollo; London: National Institute for Health and Clinical Excellence
- Krieger N (2001) Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology* 30 (4): 668–77
- Krieger N (2008) Proximal, distal and the politics of causation: what's level got to do with it? *American Journal of Public Health* 98: 221–30
- Krieger N, Rehkopf DH, Chen JT (2008) The fall and rise of US inequalities in premature mortality. *PLoS Medicine* 5 (2): e46 doi:10.1371/journal.pmed.0050046
- Kuh D, Ben-Shlomo Y, Lynch J et al. (2003) Life course epidemiology. *Journal of Epidemiology and Community Health* 57: 778–83
- Leon D, Walt G, Gilson L (2001) International perspectives on health inequalities and policy. *British Medical Journal* 322: 591–4

- Levine R, The What Works Working Group with Kinder M (2004) Millions saved: proven successes in global health. Washington DC: Center for Global Development
- Link BG, Phelan JC (2005) Fundamental sources of health inequalities in Mechanic D, Rogut LB, Colby DC et al. editors. Policy challenges in modern healthcare. New Jersey: Rutgers University Press
- McKeown T (1976) The role of medicine: dream, mirage or nemesis? London: Nuffield Provincial Hospitals Trust
- Macintyre S (1997) The Black report and beyond: what are the issues? *Social Science and Medicine* 44 (6): 723–45
- Marmot M, Wilkinson R editors (1999) Social determinants of health. Oxford: Oxford University Press
- Mead GH (1934) *Mind, self and society: from the standpoint of the social behaviorist*. Chicago: Chicago University Press
- Mechanic D (1962) The concept of illness behaviour. *Journal of Chronic Disease* 15: 189–94
- Mechanic D, Rogut LB, Colby DC et al. (2005) Introduction in Mechanic D, Rogut LB, Colby DC et al. editors. Policy challenges in modern healthcare. New Jersey: Rutgers University Press
- National Institute for Health and Clinical Excellence (2006) *Methods for development of NICE public health guidance*. London: National Institute for Health and Clinical Excellence
- National Institute for Health and Clinical Excellence (2007) *Behaviour change at population, community and individual levels. Public health guidance 6*. London: National Institute for Health and Clinical Excellence
- Rosenstock IM (1974) Historical origins of the health belief model. *Health Education Monograph* 2: 328–35
- Schutz A (1964) *Collected papers: II studies in social theory*. The Hague: Martinus Nijhoff
- Schutz A (1967) *The phenomenology of the social world*. Translated by Walsh G, Lehnert F. Illinois: North Western University Press
- Schutz A (1970) *On phenomenology and social relations: selected writings*. Chicago: Chicago University Press
- Shaw M, Dorling D, Gordon D et al. (1999) *The widening gap: health inequalities and policy in Britain*. Bristol: Policy Press
- Siegrist J, Marmot M (2004) Health inequalities and the psychosocial environment – two scientific challenges. *Social Science and Medicine* 58: 1463–73

Solar O, Irwin A (2007) Towards a conceptual framework for analysis and action on the social determinants of health [draft]. Geneva: World Health Organization

Starfield B (2006) State of the art in research on equity and health. *Journal of Health Politics, Policy and Law* 31: 11–32

Starfield B (2007) Pathways of influence on equity in health. *Social Science and Medicine* 64: 1355–62

Tudor Hart JT (1971) The inverse care law. *Lancet* 1 (7696): 405–12

Victora CG, Vaughan JP, Barros FC, et al. (2000) Explaining trends in inequalities: evidence from Brazilian child health studies. *Lancet* 356: 1093–98

Weber M (1948) Class, status, party In Gerth HH, Mills CW editors. *Max Weber: essays in sociology*. London: Routledge and Kegan Paul

Whitehead M (1992) Perspectives in health inequity. *International Journal of Health Services* 22: 429–45

Appendix B Electronic resources

Core databases

Recommended core databases; these should be considered for each topic.

- 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 Central Register of Controlled Trials
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness (DARE; 'other reviews' in Cochrane Library)
- Current Contents
- EMBASE
- HMIC (or King's Fund catalogue and DH data)
- MEDLINE
- UK Clinical Research Network Portfolio Database
- PsycINFO
- Sociological Abstracts
- Social Policy and Practice
- Social Science Citation Index

Additional topic-specific databases for consideration

This list, which is not exhaustive, suggests additional databases which could be considered, depending on the focus of the review.

- ABI Inform (business/workplace issues)
- Ageline (older people)

- Campbell Collaboration reviews
- Enviroline (environment)
- EPPI Centre databases (qualitative reviews)
- ERIC (education)
- PAIS International (Public Affairs Information Service)
- SportDiscus (sport)
- Social Care Online (SCIE) (social care)
- Transport
- Transportation Research Information Services (TRIS) (transport)

As public health topics can potentially cover a wide range of subjects, it is recommended that advice is sought from an appropriate subject librarian or subject specialist. For example, a business librarian (either academic or corporate) could provide advice on databases relevant to the workplace. A useful list of databases which may be relevant to public health topics is maintained by University of Massachusetts Medical School Lamar Soutter Library at: <http://library.umassmed.edu/ebpph/dblist.cfm>

Core websites

This is a short list of core websites, and additional subject-relevant websites should be considered.

- NICE website. In addition, former Health Development Agency documents should be searched at: www.nice.org.uk/page.aspx?o=hda.publications
- Public health observatories. These should be searched using the interoperable search facility.
- Scottish Government and Welsh Assembly Government (where policy for the topic is devolved).
- Relevant National Library for Health Specialist Library (where appropriate).
- Joseph Rowntree Foundation (where appropriate).

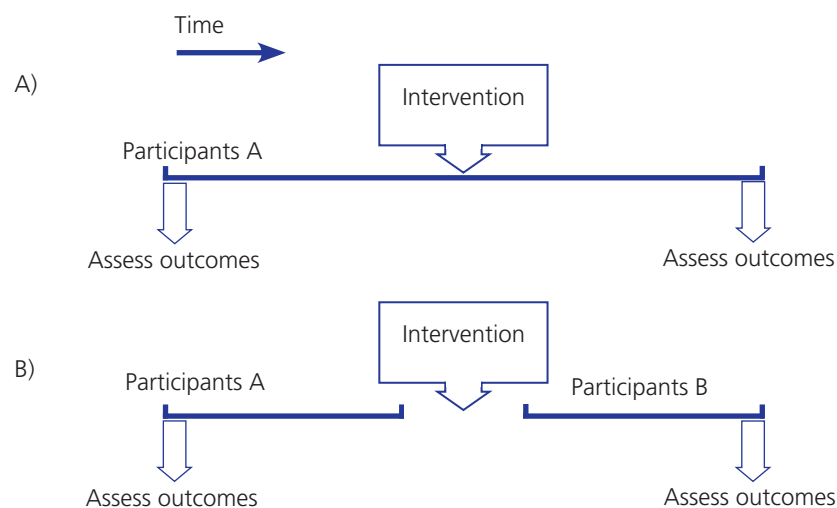
Appendix C Example of audit information to accompany search strategies

Database name	(name of database)
Database host	(name of host or environment in which the database was searched)
Database coverage dates	(dates that the database covers)
Searcher	(name of staff who conducted the search)
Search date	(date that the database was searched)
Search strategy checked by	(names of people who checked the search strategy. At least one person in addition to the searcher should check this)
Number of records retrieved	(total number of records retrieved)
Name of RefMan library	(file name of RefMan results)
Number of records loaded into RefMan	(total number of records retrieved)
Reference numbers of records in RefMan library	(range of unique reference numbers assigned to the records by RefMan)
Number of records after de-duplication in RefMan library	(total number of records from database after de-duplication)

Appendix D Glossary of study designs

Before-and-after (BA) studies

An approach where the dependent variables are measured before and after an intervention has been delivered. The intervention can either be delivered by the investigator or by others (observational before and after study). An approach that is often called a pre–post study. Study participants in pre- and post-intervention stages can either be the same (A) – as is often the case for simple one-to-one intervention studies – or different (B) – as is often the case for assessing large-scale interventions.



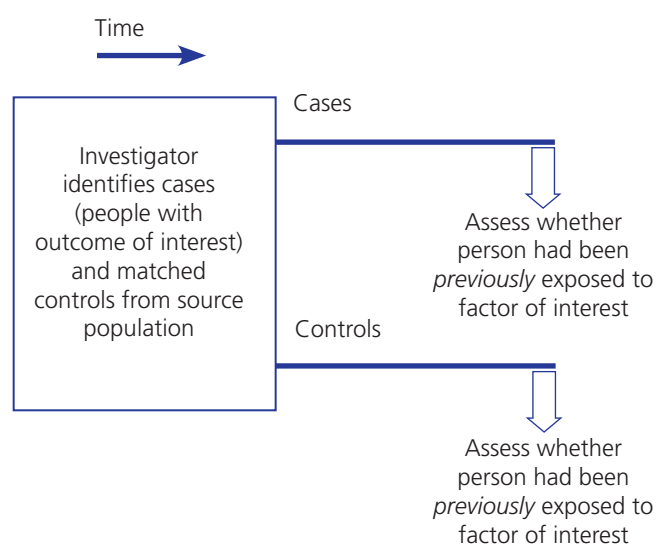
Quality appraisal of BA studies

BA studies are generally considered to have lower internal validity than study designs in which outcomes in the intervention (exposed) group are compared with outcomes in a concurrent (unexposed) control group. Important things to consider in appraising a BA study is whether or not there is any evidence for a prevailing 'temporal trend' (for example, a general reduction in population smoking rates) that may confound study findings and whether or not there is any indication of selection bias (for example, have the most or least motivated individuals been selected for participation?). A good quality BA study will demonstrate clear and consistent inclusion and exclusion criteria in subject selection. Likewise, causality between intervention and outcome will often be strengthened when observed changes are significant and occur soon after the intervention.

BA studies should not be confused with non-randomised controlled trials.

Case–control studies

A comparative observational study in which the investigator selects people who have an outcome of interest (for example, developed a disease) and others who have not (controls), and then collects data to determine previous exposure to possible causes. Case–control studies are often reserved for early hypothesis testing or for investigating the causes of rare outcomes.



Quality appraisal of case–control studies

Case–control studies are generally considered to have lower internal validity than study designs in which outcomes in the intervention (exposed) group are prospectively compared with outcomes in a concurrent (unexposed) control group. Important things to consider in appraising a case–control study are whether or not there is any indication of potential confounding factors; whether there is any indication of selection bias (for example, have the same selection and exclusion criteria been applied equally to both cases and controls?), and whether or not there is likely to be significant recall bias (for example, whether the investigator is able to reliably determine previous exposure to factor of interest for both cases and controls).

Cluster randomised controlled trial

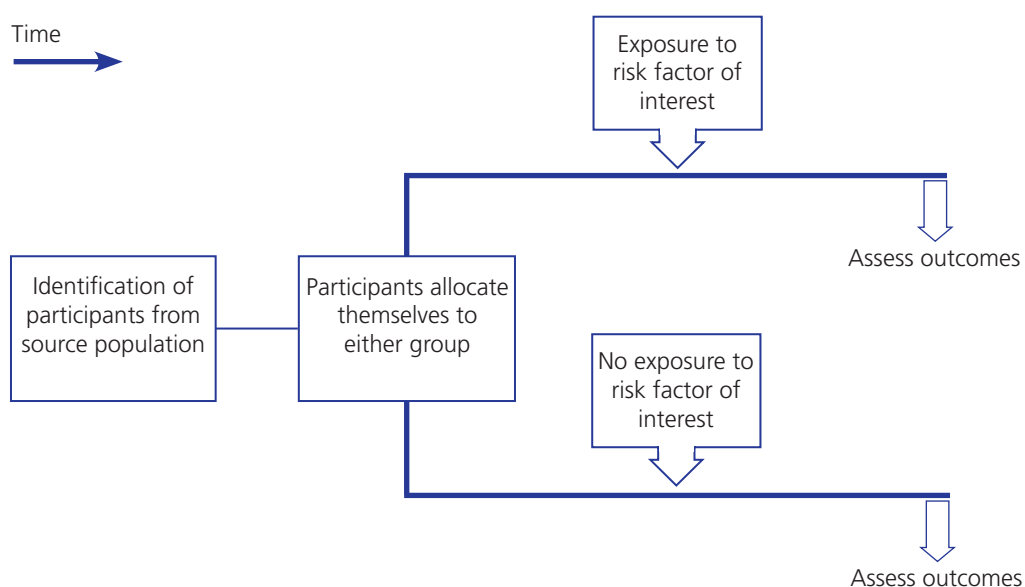
A trial where the unit of randomisation is a cluster of participants (for example, a school). See randomised controlled trial (RCT).

Quality appraisal of cluster randomised controlled trials

As for (individual) randomised controlled trials, important things to consider in appraising a cluster RCT is whether or not the method of randomisation is truly random (compared to pseudo-randomisation procedures); whether the allocation of clusters to either intervention or control could have been influenced by the person doing the allocation (allocation concealment); and whether the trial is externally valid – that is, the extent to which the findings of a study are applicable or generalisable beyond the confines of the study itself. It is also important to consider whether appropriate analyses were conducted (that is, for cluster design, analyses of sample size, power, and effect size should be performed on clusters – as the unit of randomisation – rather than individuals).

Cohort studies

An observational study in which a group or 'cohort' of people are observed over time in order to see who develops the outcome of interest. An approach that is often called a longitudinal study. Cohort studies differ from experimental studies such as randomised or non-randomised controlled trials because individuals effectively allocate themselves according to the extent of their exposure to the risk factor of interest. Prospective cohort studies involve following groups of people forward in time to assess who develops the outcome of interest, often by conducting a series of cross-sectional studies. Conversely, in retrospective cohort studies, both the exposure and outcomes of interest all take place in the past relative to the starting point of the study.



Quality appraisal of cohort studies

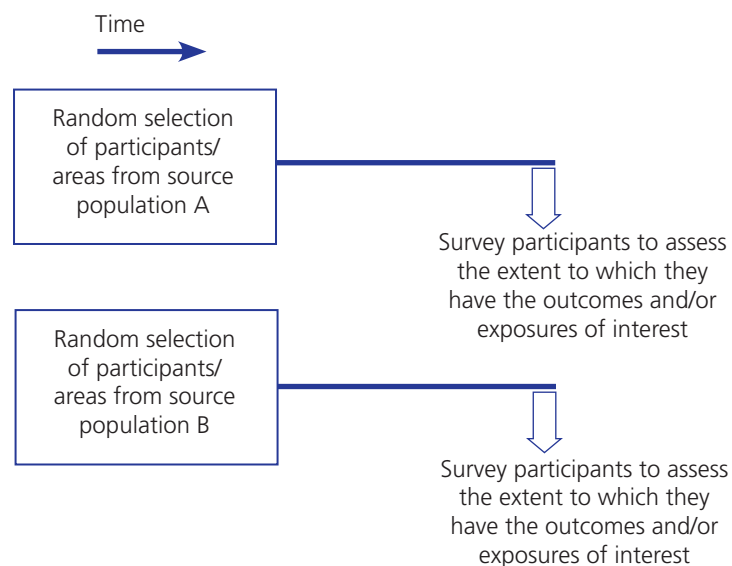
Cohort studies are generally considered to be the most reliable observational study design and are particularly useful for examining the effects of harmful exposures. Although they do not allow for the random allocation of study participants to receive or not receive the exposure of interest, they do allow for random sampling from the source population which can minimise some sources of selection bias and confounding. Important things to consider in appraising cohort studies include whether or not there is any indication of selection bias or confounding; in the case of retrospective cohort studies, whether or not there is likely to be significant recall bias (that is, whether the investigator is able to reliably determine previous exposure to factor of interest for both cases and controls), and in the case of prospective cohort studies, whether or not there is likely to be significant withdrawal bias (that is, where a significant number of participants have been lost to follow-up).

Controlled before-and-after (CBA) study

See non-randomised controlled trial.

Correlation study

A correlation study is an observational study in which the association (or correlation) between two or more variables is investigated. An approach that is often called an ecological or association study. Like cross-sectional studies, correlation studies are descriptive rather than analytical and cannot be used to estimate the relationship between cause and effect.

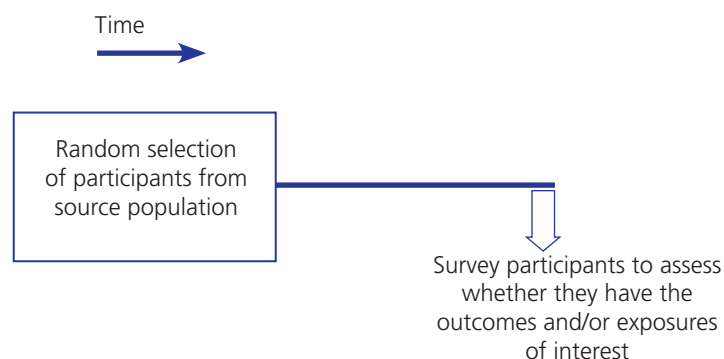


Quality appraisal of correlation studies

The most important things to consider in appraising the quality of correlation studies is whether or not there are any potential confounding factors and whether there is likely to be significant sources of measurement bias (that is, whether exposures and outcomes are assessed using reliable criteria).

Cross-sectional study

A cross-sectional study is an observational study in which the source population is examined to see what proportion has the outcome of interest, or has been exposed to a risk factor of interest, or both. Cross-sectional studies are generally used to determine the population prevalence of outcomes or exposures. An approach that is often called a survey or a prevalence study. Although cross-sectional studies can often provide useful estimates of disease burden for a particular population, they are less reliable for determining prevalence of very rare conditions or conditions with a short duration. As cross-sectional studies are descriptive rather than analytical, they cannot be used to estimate the relationship between cause and effect.



Quality appraisal of cross-sectional studies

The most important things to consider in appraising the quality of cross-sectional studies is whether or not there are any potential confounding factors and whether or not there is likely to be significant sources of measurement bias (that is, whether exposures and outcomes are assessed using reliable criteria).

Ecological study

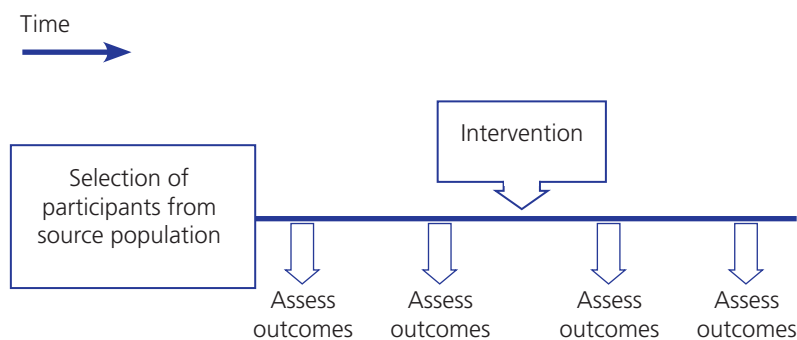
See correlation and cross-sectional studies.

Focus group study

A specific type of interview study (see interview study) where a group of people (usually 6–12) is interviewed by one or more facilitators or interviewers. The method explicitly includes and uses the group interaction to generate data. Data are managed in a similar way to other interview data, in that the discussions are usually recorded and transcribed and analysis is undertaken on the transcripts rather than the verbal data.

Interrupted time series

An approach in which multiple (more than two) observations are made on the same individuals, or groups of individuals, over time.



Quality appraisal of interrupted time series

The most important things to consider in appraising the quality of interrupted time series are whether or not outcomes were assessed before and after an intervention was delivered; whether it is clear precisely when an intervention took place (that is, in order to compare outcomes before and after intervention); and like before-and-after studies, whether or not there is any evidence for a prevailing 'temporal trend' (for example, a general reduction in population smoking rates) that may confound study findings.

Interview study

A qualitative method of data collection where participant's views are elicited via verbal interviews. Interviews can be structured (that is, reliant on set questions), semi-structured (the interviewer has a general idea of topics and themes to cover and will steer the interview towards these participants) or unstructured (the format and direction of the interview is set by the participant). In addition, interviews can be one-to-one, or with couples or small groups (see focus

groups). They can be conducted in-person, on the telephone, or more recently online. Interviews are usually recorded and transcribed and analysis is undertaken on the transcripts rather than the verbal data.

Longitudinal studies

See cohort studies.

Meta-analysis

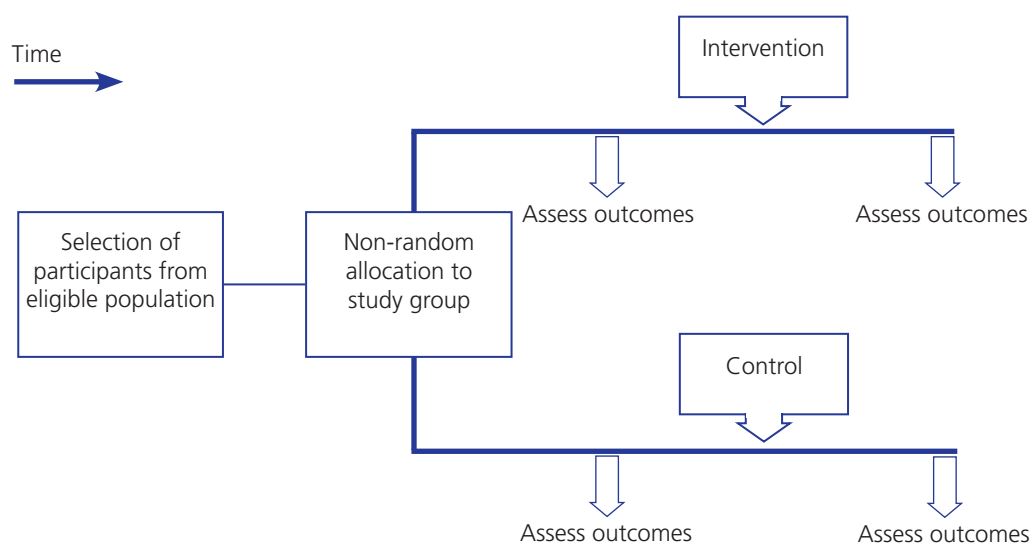
Should not be confused with systematic review. Meta-analysis is a statistical technique that enables the findings from multiple primary studies (often identified during a systematic review) to be combined.

Non-participant observation

A qualitative methodology where the researcher observes participants as they engage with the phenomenon being researched.

Non-randomised controlled trial (NRCT)

These are trials where participants (or clusters) are allocated to receive either intervention or control (or comparison intervention) but the allocation is not randomised – an approach often called a controlled before-and-after (CBA) study.



Quality appraisal of non-randomised controlled trials

NRCTs are generally considered to be less reliable than randomised controlled trials as non-randomisation of participants makes selection, participant and observer biases much more likely and confounding factors may exist. The most important thing to consider in appraising NRCTs is whether or not there are likely to be significant baseline differences between groups.

Participant observation

A qualitative methodology where the researcher joins in with the participants as they engage with the phenomenon being researched.

Pre–post study

See before-and-after (BA) study.

Prevalence study

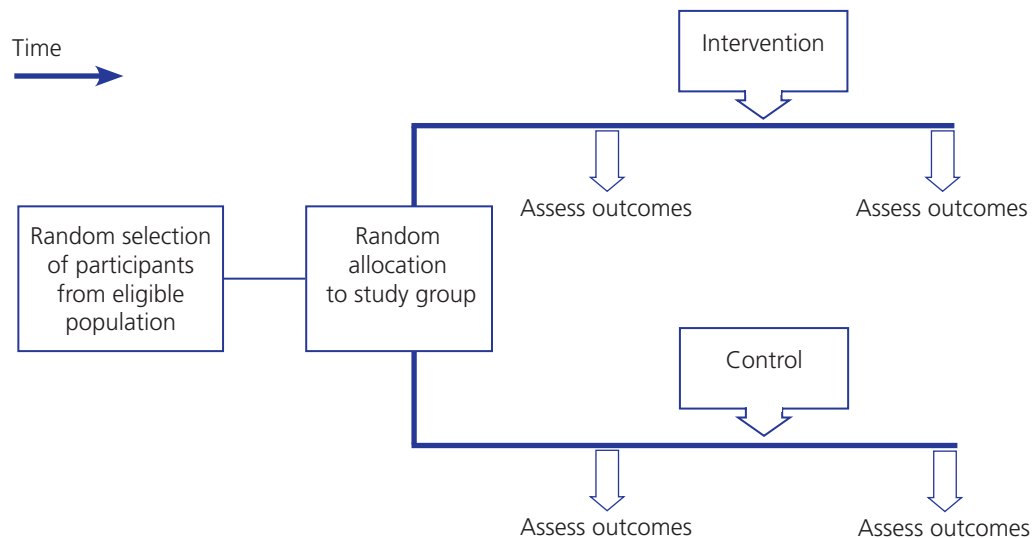
See cross-sectional study.

Prospective cohort study

See cohort study.

Randomised controlled trial (RCT)

These are trials where participants (or clusters) are randomly allocated to receive either intervention or control. If well implemented, randomisation should ensure that intervention and control groups only differ in their exposure to treatment.



Quality appraisal of RCTs

RCTs are generally considered to be the most rigorous experimental study design as the randomisation of participants helps to minimise confounding and other sources of bias. Important things to consider in appraising an RCT are whether or not the method of randomisation is truly random (compared to pseudo-randomisation procedures such as consecutive admissions to a clinic); whether the allocation of participants to either intervention or control could have been influenced by the person doing the allocation (allocation concealment); and whether the trial is externally valid – that is, the extent to which the findings of a study are applicable or generalisable beyond the confines of the study itself. It is not unusual for an RCT to have strong internal validity, but poor external validity, for example, if stringent selection criteria for entry into the study mean that the study participants fail to reflect the characteristics of the source population.

Retrospective cohort study

See cohort study.

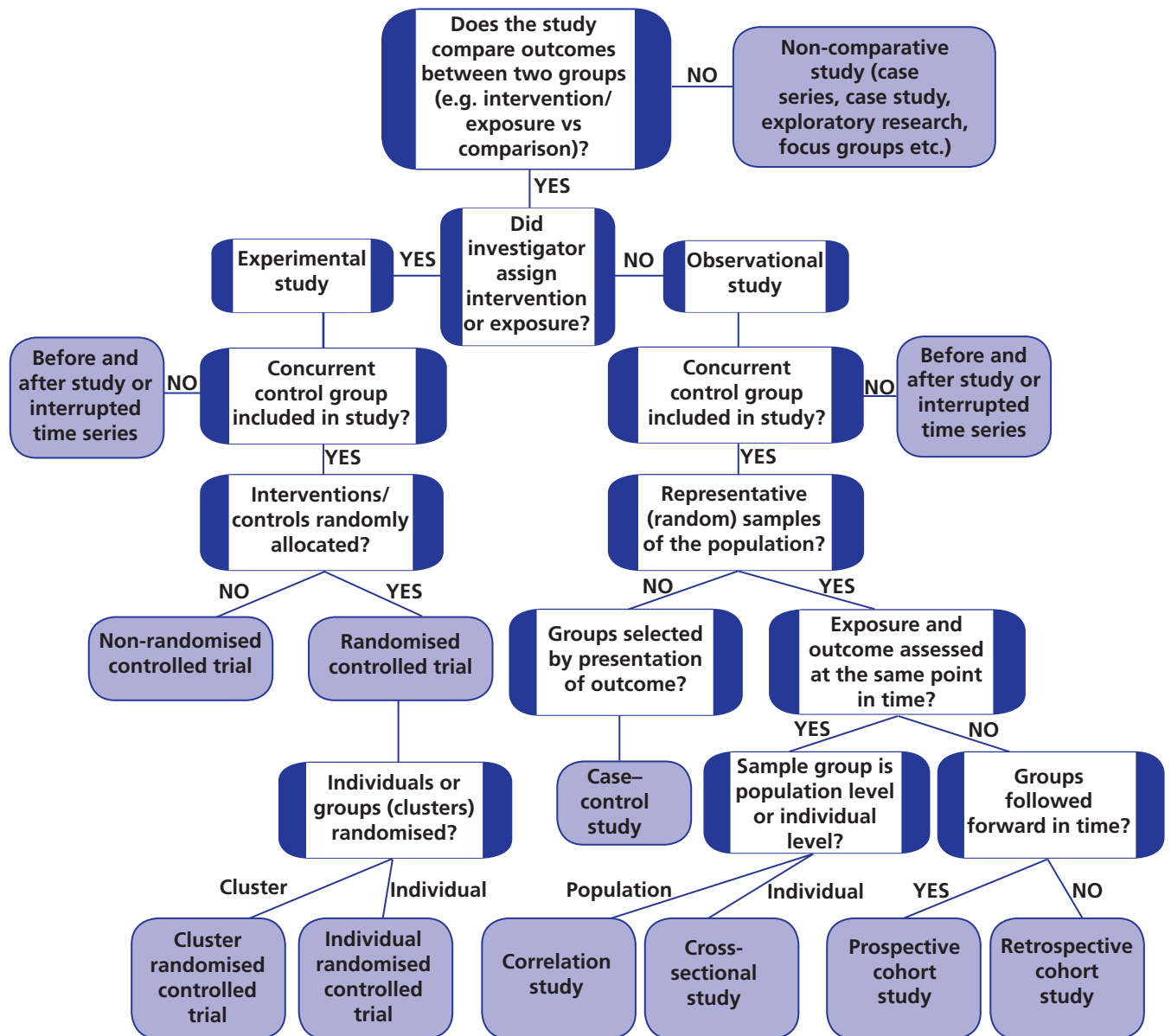
Survey

Surveys can be used as part of a quantitative or a qualitative methodology. For quantitative surveys, see cross-sectional study. Qualitative surveys try to elicit qualitative data from surveys or questionnaires by providing open-ended questions that try to encourage a lengthy response and then treat this data as qualitative data by coding and analysing it.

Systematic review

A systematic review can be defined as a summary of the literature that uses explicit and systematic methods to identify, appraise and summarise the literature according to predetermined criteria. If this description is not present, it is not possible to make a thorough evaluation of the quality of the review.

Appendix E Algorithm for classifying quantitative (experimental and observational) study designs



Primary study designs are noted in shaded boxes. For further description of study designs, see appendix D for a glossary of terms.

Appendix F Quality appraisal checklist – quantitative intervention studies¹⁵

Public health interventions comprise a vast range of approaches, from the relatively simple through to complex national policy interventions. As a consequence, research questions about the effectiveness and efficacy of public health interventions will typically rely on quantitative evidence from a range of sources (see section 3.2). This will include evidence from small (experimental) randomised controlled trials through to large-scale observational studies (see appendix E for an algorithm outlining the range of experimental and observational quantitative study designs). Rather than include an exhaustive list of critical appraisal tools for each individual study design, we have chosen to adopt a single checklist assessing key aspects of studies designed to determine the effect of an intervention on a (quantitative) outcome.

This checklist replaces those for randomised controlled trials, case-control studies, cohort studies, controlled before-and-after studies and interrupted time series from appendix A of the 'Public health guidance: development, process and methods manual' (2006).

A number of generic quality appraisal checklists have been developed, some specifically for public health studies (Effective Public Health Practice Project 2008; Heller et al. 2008). After consultation with both internal and external reviewers, the 'Graphical appraisal tool for epidemiological studies (GATE)', developed by Jackson et al. (2006) emerged as the preferred checklist.

GATE has been revised, tailoring it to be more suitable for public health interventions. It is anticipated that the majority of study designs used to assess public health interventions will be amenable to critical appraisal with this revised tool.

It enables a reviewer to appraise a study's internal and external validity after addressing the following key aspects of study design: characteristics of study participants; definition of, and allocation to, intervention and control conditions; outcomes assessed over different time periods and method(s) of analyses.

GATE is intended to be used in an electronic (Excel) format which will facilitate both the sharing and storage of data, and through linkage with other documents, the compilation of research reports. Much of the guidance to support the completion of the critical appraisal form which is reproduced on pages 174–85 also appears in 'pop-up' windows in the electronic version¹⁶.

¹⁵ Appraisal form derived from: Jackson R, Ameratunga S, Broad J et al. (2006) The GATE frame: critical appraisal with pictures. *Evidence Based Medicine* 11(2): 35–8.

¹⁶ Available from CPHE on request.

There are five sections of the revised GATE. Section 1 seeks to assess the key population criteria for determining the study's **external validity** – that is, the extent to which the findings of a study are generalisable beyond the confines of the study to the study's source population.

Sections 2 to 4 assess the key criteria for determining the study's **internal validity** – that is, making sure that the study has been carried out carefully, and that the outcomes are likely to be attributable to the intervention being assessed, rather than some other (often unidentified) factor. In an internally valid study, any differences observed between groups of patients allocated to receive different interventions may (apart from the possibility of random error) be attributed to the intervention under investigation. Biases are characteristics that are likely to make estimates of effect differ systematically from the truth. Each of the critical appraisal checklist questions covers an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

Checklist items are worded so that one of five responses is possible:

- ++ Indicates that for that particular aspect of study design, the study has been designed/conducted in such a way as to minimise the risk of bias.
- + Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias for that particular aspect of study design.
- Should be reserved for those aspects of the study design in which significant sources of bias may persist.
- Not reported (nr) Should be reserved for those aspects in which the study under review fails to report how they have/might have been considered.
- Not applicable (na) Should be reserved for those study design aspects which are not applicable given the study design under review (for example, allocation concealment would not be applicable for case control studies).

In addition, the reviewer is requested to complete in detail the comments section of the quality appraisal form so that the grade awarded for each study aspect is as transparent as possible.

Each study is then awarded an overall study quality grading for internal validity (IV) and a separate one for external validity (EV):

- ++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
- + Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Checklist

Study identification (Include full citation details)		
Study design: Refer to the glossary of study designs (appendix D) and the algorithm for classifying experimental and observational study designs (appendix E) to best describe the paper's underpinning study design		
Guidance topic:		
Assessed by:		
Section 1: Population		
1.1 Is the source population or source area well described? Was the country (e.g. developed or non-developed, type of healthcare system), setting (primary schools, community centres etc.), location (urban, rural), population demographics etc. adequately described?	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
1.2 Is the eligible population or area representative of the source population or area? Was the recruitment of individuals/clusters/areas well-defined (e.g. advertisement, birth register)? Was the eligible population representative of the source? Were important groups under-represented?	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
1.3 Do the selected participants or areas represent the eligible population or area? Was the method of selection of participants from the eligible population well described? What % of selected individuals/clusters agreed to participate? Were there any sources of bias? Were the inclusion/exclusion criteria explicit and appropriate?	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:

Section 2: Method of allocation to intervention (or comparison)

<p>2.1 Allocation to intervention (or comparison). How was selection bias minimised?</p> <p>Was allocation to exposure and comparison randomised? Was it truly random (++) or pseudo-randomised (+) (e.g. consecutive admissions)? If not randomised, was significant confounding likely (–) or not (+)?</p> <p>If a cross-over, was order of intervention randomised?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.2 Were interventions (and comparisons) well described and appropriate?</p> <p>Were intervention/s and comparison/s described in sufficient detail (i.e. enough for study to be replicated)?</p> <p>Was comparison/s appropriate (e.g. usual practice rather than no intervention)?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.3 Was the allocation concealed?</p> <p>Could the person(s) determining allocation of participants/clusters to intervention or comparison groups have influenced the allocation?</p> <p>Adequate allocation concealment (++) would include centralised allocation or computerised allocation systems.</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.4 Were participants and/or investigators blind to exposure and comparison?</p> <p>Were participants AND investigators – those delivering and/or assessing the intervention kept blind to intervention allocation? (Triple or double blinding score [++])</p> <p>If lack of blinding is likely to cause important bias, score (–).</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

<p>2.5 Was the exposure to the intervention and comparison adequate?</p> <p>Is reduced exposure to intervention or control related to the intervention (e.g. adverse effects leading to reduced compliance) or fidelity of implementation (e.g. reduced adherence to protocol)?</p> <p>Was lack of exposure sufficient to cause important bias?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.6 Was contamination acceptably low?</p> <p>Did any in the comparison group receive the intervention or vice versa?</p> <p>If so, was it sufficient to cause important bias?</p> <p>If a cross-over trial, was there a sufficient wash-out period between interventions?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.7 Were other interventions similar in both groups?</p> <p>Did either group receive additional interventions or have services provided in a different manner?</p> <p>Were the groups treated equally by researchers or other professionals?</p> <p>Was this sufficient to cause important bias?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.8 Were all participants accounted for at study conclusion?</p> <p>Were those lost-to-follow-up (i.e. dropped or lost pre-/during/post-intervention) acceptably low (i.e. typically <20%)?</p> <p>Did the proportion dropped differ by group? For example, were drop-outs related to the adverse effects of the intervention?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>2.9 Did the setting reflect usual UK practice?</p> <p>Did the setting in which the intervention or comparison was delivered differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) condition in a hospital rather than a community-based setting?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

<p>2.10 Did the intervention or control comparison reflect usual UK practice?</p> <p>Did the intervention or comparison differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) delivered by specialists rather than GPs? Were participants monitored more closely?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>Section 3: Outcomes</p>		
<p>3.1 Were outcome measures reliable?</p> <p>Were outcome measures subjective or objective (e.g. biochemically validated nicotine levels [++] vs self-reported smoking [–]).</p> <p>How reliable were outcome measures (e.g. inter- or intra-rater reliability scores)?</p> <p>Was there any indication that measures had been validated (e.g. validated against a gold standard measure or assessed for content validity)?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.2 Were all outcome measurements complete?</p> <p>Were all/most study participants who met the defined study outcome definitions likely to have been identified?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.3 Were all important outcomes assessed?</p> <p>Were all important benefits and harms assessed?</p> <p>Was it possible to determine the overall balance of benefits and harms of the intervention versus comparison?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

<p>3.4 Were outcomes relevant?</p> <p>Where surrogate outcome measures were used, did they measure what they set out to measure? (e.g. a study to assess impact on physical activity assesses gym membership – a potentially objective outcome measure – but is it a reliable predictor of physical activity?)</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.5 Were there similar follow-up times in exposure and comparison groups?</p> <p>If groups are followed for different lengths of time, then more events are likely to occur in the group followed-up for longer distorting the comparison.</p> <p>Analyses can be adjusted to allow for differences in length of follow-up (e.g. using person-years).</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.6 Was follow-up time meaningful?</p> <p>Was follow-up long enough to assess long-term benefits/harms?</p> <p>Was it too long, e.g. participants lost to follow-up?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>Section 4: Analyses</p>		
<p>4.1 Were exposure and comparison groups similar at baseline? If not, were these adjusted?</p> <p>Were there any differences between groups in important confounders at baseline?</p> <p>If so, were these adjusted for in the analyses (e.g. multivariate analyses or stratification)?</p> <p>Were there likely to be any residual differences of relevance?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

<p>4.2 Was Intention to treat (ITT) analysis conducted?</p> <p>Were all participants (including those that dropped out or did not fully complete the intervention course) analysed in the groups (i.e. intervention or comparison) to which they were originally allocated?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>4.3 Was the study sufficiently powered to detect an intervention effect (if one exists)?</p> <p>A power of 0.8 (i.e. it is likely to see an effect of a given size if one exists, 80% of the time) is the conventionally accepted standard.</p> <p>Is a power calculation presented? If not, what is the expected effect size? Is the sample size adequate?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>4.4 Were the estimates of effect size given or calculable?</p> <p>Were effect estimates (e.g. relative risks, absolute risks) given or possible to calculate?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>4.5 Were the analytical methods appropriate?</p> <p>Were important differences in follow-up time and likely confounders adjusted for?</p> <p>If a cluster design, were analyses of sample size (and power), and effect size performed on clusters (and not individuals)?</p> <p>Were subgroup analyses pre-specified?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>4.6 Was the precision of intervention effects given or calculable? Were they meaningful?</p> <p>Were confidence intervals (CIs) and/or p-values for effect estimates given or possible to calculate?</p> <p>Were CIs wide or were they sufficiently precise to aid decision-making? If precision is lacking, is this because the study is under-powered?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

Section 5: Summary

5.1 Are the study results internally valid (i.e. unbiased)?

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)?
Were there significant flaws in the study design?

++

+

–

Comments:

5.2 Are the findings generalisable to the source population (i.e. externally valid)?

Are there sufficient details given about the study to determine if the findings are generalisable to the source population? Consider: participants, interventions and comparisons, outcomes, resource and policy implications.

++

+

–

Comments:

Notes on the use of the quantitative intervention studies checklist

The following sections outline the checklist questions, the prompts provided as pop-up boxes in the electronic version (highlighted in boxes) and additional guidance notes to aid the reviewer in assessing the study's internal and external validity.

Section 1: population

This section seeks to assess the key population criteria for determining the study's **external validity**.

Although there are checklists for assessing external validity of RCTs (with a particular focus on clinical interventions) (see for example [Rothwell 2005]), there don't appear to be any checklists specific for public health interventions.

The questions asked in this section ask the reviewer to identify and describe the **source** population of the study (that is, those the study aims to represent), the **eligible** population (those that meet the study eligibility criteria), and the **study participants** (those that agreed to participate in the study). Where a study assesses an intervention delivered to a particular geographical setting or area (rather than delivered to individuals), the questions in this section relate to describing the source area or setting, and how the study areas or settings were chosen. For example, a study might assess the effect on health outcomes of neighbourhood renewal schemes and this section seeks to identify and describe how those neighbourhoods were chosen and whether they are representative of the neighbourhoods the study seeks to represent.

External validity is defined as the extent to which the findings of a study are generalisable beyond the confines of the study itself to the source population. So, for example, findings from a study conducted in a school setting in the USA might be generalisable to other schools within the USA (the source population of the study). An assessment of external validity will consider how representative of the source population the study population is and whether or not there are any specific population, demographic or geographic features of the selected population that might limit or support generalisability. Also important are considerations of the setting, intervention and outcomes assessed. These factors will be considered in sections 2 and 3 of the checklist.

1.1 Is the source population or source area well described?

Was the source population or area described in sufficient detail? For example, country (developed or non-developed, type of healthcare system), setting (for example, primary school, community centre), location (urban, rural) and population demographics.

This question seeks to determine the study's source population or area (that is, to whom or what the study aims to represent). The source population is usually best identified by referring to the study's original research question.

It is important to consider those population demographic characteristics such as age, sex, sexual orientation, disability, ethnicity, religion, place of residence, occupation, education, socioeconomic position and social capital¹⁷, that can help to assess the impact of interventions on health inequalities and may help guide recommendations for specific population subgroups.

1.2 Is the eligible population or area representative of the source population or area?

Was the recruitment of individuals/clusters/areas well-defined (for example, advertisement, birth register, class list, area)?

Was the eligible population or area representative of the source or were important groups under-represented?

To determine if the eligible population or area (for example, smokers responding to a media advertisement, areas of high density housing within a particular catchment area) are representative of the source population (for example, smokers or areas of high-density housing), consider the means by which the eligible population was defined or identified and the implicit or explicit inclusion and exclusion criteria used. Were important groups likely to have been missed or under-represented? For example, were recruitment strategies geared toward more affluent or motivated groups? (For example, recruitment from more affluent areas or local fitness centres.) Were significant numbers of potentially eligible participants likely to have been inadvertently excluded? (For example, through referral to practitioners not involved in the research study.)

1.3 Do the selected participants or areas represent the eligible population or area?

Was the method of selection of participants from the eligible population well described?

What percentage of selected individuals or clusters agreed to participate? Were there any sources of bias?

Were the inclusion/exclusion criteria explicit and appropriate?

Consider whether the method of selection of participants or areas from the eligible population or area was well described (for example, consecutive cases or random sampling). Were any significant sources of biases likely to have been introduced? Consider what proportion of selected individuals or clusters agreed to participate. Was there a bias toward more healthier or motivated individuals or wealthier areas?

¹⁷ Demographic criteria as outlined by the PROGRESS-Plus categorisation (Kavanagh et al. 2008).

Also consider whether the inclusion and exclusion criteria were well described and whether they were appropriate given the study objectives and the source population. Strict eligibility criteria can limit the external validity of intervention studies if the selected participants are not representative of the eligible population. This has been well-documented for RCTs where recruited participants have been found to differ from those who are eligible but not recruited, in terms of age, sex, race, severity of disease, educational status, social class and place of residence (Rothwell 2005).

Finally, consider whether sufficient detail of the demographic (for example, age, education, socioeconomic status, employment) and/or personal health-related (for example, smoking, physical activity levels) characteristics of the selected participants were presented. Are selected participants representative of the eligible population?

Section 2: method of allocation to intervention (or comparison)

This section aims to assess the likelihood of selection bias and confounding being introduced into a study.

Selection bias exists when there are systematic differences between the participants in the different intervention groups. As a result, the differences in the outcome observed may be explained by pre-existing differences between the groups, rather than because of the intervention itself. For example, if the people in one group are generally in poorer health compared with the second group, then they are more likely to have a worse outcome, regardless of the effect of the intervention. The intervention groups should be similar at the start of the study so that the only difference between the groups should be the intervention received.

2.1 Allocation to intervention (or comparison). How was selection bias minimised?

Was allocation to exposure and comparison randomised? Was it truly random (++) or pseudo-randomised (+) (for example, consecutive admissions)?

If not randomised, was significant confounding likely (–) or not (+)?

If a crossover, was order of intervention randomised?

Consider the method by which individuals were allocated to either intervention or control conditions. Random allocation of individuals (as in RCTs) to receive one or other of the interventions under investigation, is considered the most reliable means of minimising the risk of selection bias and confounding.

If an appropriate method of randomisation has been used, each participant should have an equal chance of ending up in each of the intervention groups. Examples of random allocation

sequences include random numbers generated by computer, tables of random numbers and drawing of lots or envelopes. However, if the description of randomisation is poor, or the process used is not truly random (for example, if the allocation sequence is predictable, such as date of birth or alternating between one group and another) or can otherwise be seen as flawed, this component should be given a lower quality rating.

2.2 Were the interventions (and comparisons) well-described and appropriate?

Were interventions and comparisons described in sufficient detail (that is, enough for the study to be replicated)?

Were comparisons appropriate (for example, usual practice rather than no treatment)?

2.3 Was the allocation concealed?

Could the person(s) determining the allocation of participants or the clusters to intervention or comparison groups have influenced the allocation?

Adequate allocation concealment (++) would include centralised allocation or computerised allocation systems.

If investigators are aware of the allocation group for the next individual to be enrolled in the study, there is potential for people to be enrolled in an order that results in imbalances in important characteristics. For example, a practitioner might feel that people with mild rather than severe mental health problems would be more likely to do better on a new behavioural intervention and be tempted to only enrol such individuals when they know they will be allocated to that group. This would result in the intervention group being, on average, less severe at baseline than the control group. Concealment of treatment group may not always be feasible but concealment of allocation up until the point of enrolment in the study should always be possible.

Information should be presented within the paper that provides some assurance that allocations were not known until at least the point of allocation. Centralised allocation, computerised allocation systems and the use of coded identical containers would all be regarded as adequate methods of concealment. Sealed envelopes can be considered as adequate concealment if the envelopes are serially numbered, sealed and opaque, and allocation is performed by a third party. Poor methods of allocation concealment include alternation, or the use of case record numbers, date of birth or day of the week.

If the method of allocation concealment used is regarded as poor, or relatively easy to subvert, the study should be given a lower quality rating. If a study does not report any concealment approach, this should be scored as 'not reported'.

2.4 Were participants and investigators blind to exposure and comparison?

Were participants AND investigators – those delivering and/or assessing the intervention kept blind to intervention allocation? (Triple or double-blinding score [++]).

If lack of blinding is likely to cause important bias, score (-).

Blinding refers to the process of withholding information about treatment allocation or exposure status from those involved in the study who could potentially be influenced by this information. This can include participants, investigators, those administering care and those involved in data collection and analysis.

Unblinded individuals can bias the results of studies, either intentionally or unintentionally, through the use of other effective co-interventions, decisions about withdrawal, differential reporting of symptoms, or influencing concordance with treatment.

The terms 'single blind', 'double blind' and even 'triple blind' are sometimes used in studies. Unfortunately, they are not always used consistently. Commonly, when a study is described as 'single blind', only the participants are blind to their group allocation. When both participants and investigators are blind to group allocation the study is often described as 'double blind'. It is preferable to record exactly who was blinded, if reported, to avoid misunderstanding.

It is important to note that blinding of participants and researchers is not always possible, and it is important to think about the likely size and direction of bias caused by failure to blind in making an assessment of this component.

2.5 Was the exposure to the intervention and comparison adequate?

Is reduced exposure to the intervention or control related to the intervention (for example, adverse effects leading to reduced compliance) or fidelity of implementation (for example, reduced adherence to protocol)?

Was lack of exposure sufficient to cause important bias?

2.6 Was contamination acceptably low?

Did any in the comparison group receive the intervention or vice versa?

If so, was it sufficient to cause important bias?

If a crossover trial, was there a sufficient wash-out period between interventions?

2.7 Were other interventions similar in both groups?

Did either group receive additional interventions or have services provided in a different manner?

Were the groups treated equally by researchers or other professionals?

Was this sufficient to cause important bias?

This question seeks to establish if there were any important differences between the intervention groups aside from the intervention received. If some patients received additional intervention (known as 'co-intervention'), this additional intervention is a potential confounding factor in the presence of which can make it difficult to attribute any observed effect to the intervention rather than to the other factors.

2.8 Were all participants accounted for at study conclusion?

Were those lost to follow-up (that is, dropped or lost pre-/during/post- intervention) acceptably low (that is, typically less than 20%)?

Did the proportion dropped differ by group? For example, were drop-outs related to the adverse effects of intervention?

Section 2 also aims to assess the likelihood of attrition bias being introduced into a study.

Attrition bias occurs when there are systematic differences between the comparison groups with respect to participants lost, or differences between participants lost to the study and those who remain. Attrition can occur at any point after participants have been allocated to their intervention groups. As such, it includes participants who are excluded post-allocation (and may indicate a violation of eligibility criteria), those who fail to complete the intervention and those who fail to complete outcome measurement (regardless of whether or not the intervention was completed).

It is a concern if the number of participants who were lost to follow-up (that is, dropped out) is high – typically >20%, although it is not unreasonable to expect a higher drop-out rate in studies conducted over a longer period of time.

Consideration should also be given to the reasons why participants dropped out. Participants who dropped out of a study may differ in some significant way from those who remained in the study. Drop-out rates and reasons for dropping out should be similar across all treatment groups. In good quality studies, the proportion of participants lost after allocation are reported and the possibility of attrition bias considered within the analysis.

2.9 Did the setting reflect usual UK practice?

Did the setting in which the intervention or comparison was delivered differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) condition in a hospital rather than a community-based setting?

2.10 Did the intervention or control comparison reflect usual UK practice?

Did the intervention or comparison differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) delivered by specialists rather than GPs? Were participants monitored more closely?

Section 3: outcomes

Some of the items on this checklist may need to be filled in separately for each of the different outcomes reported by the study. For example, a study may report only one outcome of interest, measured by one tool, at one point in time, in which case each of the components (for example, reliability of outcome measure, relevance, withdrawals and drop-outs) can be assessed based on that one tool. However, if a study reports multiple outcomes of interest, scored by multiple tools (for example, self-report AND biochemically validated measures), at multiple points in time (for example, 6-month follow-up AND 1-year follow-up) individual components will need to be assessed for each outcome of interest.

It is important, therefore, that the reviewer has a clear idea of what the important outcomes are and over what timeframe, before appraising a study. The important outcomes for a piece of guidance will be identified through consultation with the NICE project team, the guidance development committee (PHIAC or the PDG) and stakeholders.

3.1 Were the outcome measures reliable?

Were outcome measures subjective or objective (for example biochemically validated nicotine levels [++] versus self-reported smoking)?

How reliable were outcome measures (for example inter- or intra-rater reliability scores)?

Was there any indication that measures had been validated (for example validated against a gold standard measure or assessed for content validity)?

This question seeks to determine how reliable (that is, how consistently the method measures a particular outcome) and valid (that is, the method measures what it claims to measure) the

outcome measures were. For example, a study assessing effectiveness of a smoking cessation intervention may report on a number of outcomes using a number of different tools, including self-reported smoking rates (a subjective outcome measure that is often unreliable) and biochemically validated smoking rates (an objective outcome measure that is likely to be more reliable).

If the outcome measures were subjective, it is also important to consider if the participant and/or researcher was blinded to the intervention/exposure (see question 2.4) as blinding may rescue the reliability of some subjective outcome measures.

3.2 Were all outcome measurements complete?

Were all or most study participants who met the defined study outcome definitions likely to have been identified?

3.3 Were all important outcomes assessed?

Were all important benefits and harms assessed?

Was it possible to determine the overall balance of benefits and harms of the intervention versus comparison?

3.4 Were outcomes relevant?

Where surrogate outcome measures were used, did they measure what they set out to measure? For example, a study to assess impact on physical activity assesses gym membership – a potentially objective outcome measure – but is it a reliable predictor of physical activity?

3.5 Were there similar follow-up times in exposure and comparison groups?

If groups are followed for different lengths of time, then more events are likely to occur in the group followed up for longer distorting the comparison.

Analyses can be adjusted to allow for differences in length of follow-up (for example, using person-years).

It is possible to overcome differences in the length of follow-up between groups in the analyses, for example, by adjusting the denominator to take the time into account (by using person-years).

3.6 Was follow-up time meaningful?

Was follow-up long enough to assess long-term benefits/harms?
Was it too long, for example, participants lost to follow-up?

The duration of post-intervention follow-up of participants should be of an adequate length to identify the outcome of interest.

Section 4: analyses

4.1 Were the exposure and comparison groups similar at baseline? If not, were these adjusted?

Were there any differences between groups in important confounders at baseline?
If so, were these adjusted for in the analyses (for example, multivariate analyses or stratification)?
Were there likely to be any residual differences of relevance?

Studies may report the distributions or important differences in potential confounding factors between intervention groups. However, formal tests comparing the groups are problematic – failure to detect a difference does not mean a difference does not exist, and multiple comparisons of factors may falsely detect some differences that are not real.

It is important to assess whether all likely confounders have been considered. Confounding factors may differ by outcome, so potential confounding factors for all of the outcomes that are of interest will need to be considered.

4.2 Was intention to treat (ITT) analysis conducted?

Were all participants (including those that dropped out or did not fully complete the intervention course) analysed in the groups (that is, intervention or comparison) to which they were originally allocated?

4.3 Was the study sufficiently powered to detect an intervention effect (if one exists)?

A power of 0.8 (that is, it is likely to see an effect of a given size if one exists, 80% of the time) is the conventionally accepted standard.
Is a power calculation presented? If not, what is the expected effect size? Is the sample size adequate?

For cluster RCTs in particular, it is important to consider whether the cluster design has been appropriately taken into account in calculating required sample size for adequate power.

4.4 Were the estimates of effect size given or calculable?

Were effect estimates (for example, relative risks, absolute risks) given or possible to calculate?

4.5 Were the analytical methods appropriate?

Were important differences in follow-up time, and likely confounders, adjusted for?

If a cluster design, were analyses of sample size (and power), and effect size performed on clusters (and not individuals)?

Were subgroup analyses pre-specified?

There are a large number of considerations in deciding whether analytical methods were appropriate. For example, it is important to review the appropriateness of any subgroup analyses (and whether pre-specified or exploratory) that are presented. Although subgroup analyses can often provide valuable information on which to base further research (that is, are often exploratory), it is important that findings of subgroup analyses are not over (or under) emphasised. Meaningful results from subgroup analyses are beset by the problems of multiplicity of testing (in which the risk of a false positive result increases with the number of tests performed) and low statistical power (that is, studies generally only enrol sufficient participants to ensure that testing the primary study hypothesis is adequately powered) (Assmann et al. 2000). In a good quality paper, subgroup analyses are restricted to pre-specified subgroups and are often confined to primary outcome measures. Data are analysed using formal statistical tests of interaction (that assess whether intervention effect differs between subgroups) rather than comparison of subgroup p values. A correction for multiple testing is performed where appropriate (for example, 'Bonferroni correction' where a stricter significance level is used to define statistical significance). The results are delineated carefully, and full details of how analyses were performed are provided (Assmann et al. 2000; Guillemin 2007).

The appropriateness of some analytical methods will also depend on the study design under investigation. For example, with cluster RCTs, because participants are randomised at the group level and are not independent 'units' (as is the case with RCTs based on individuals without clustering), and outcomes are often assessed at the individual level, statistical adjustments are necessary before pooled intervention and control group outcomes can be compared.

Likewise, it is also important to consider whether the degree of similarity or difference within clusters has been considered in analyses of cluster RCTs. Good quality cluster RCTs will determine the intra-class correlation coefficient of their study (a statistical measure of the interdependence within each cluster that is calculated by taking the ratio of the variance between groups compared with variance within groups).

Studies may also report other forms of statistical analysis such as regression, time series, factor analysis and discriminant analysis, as well as epidemiological or economic modelling. Economic modelling is covered in more detail in chapter 6. The other topics are specialised and advice should be sought from the NICE project team before attempting to assess such studies.

4.6 Was the precision of intervention effects given or calculable? Were they meaningful?

Were confidence intervals and/or p-values for effect estimates given or possible to calculate?

Were confidence intervals wide or were they sufficiently precise to aid decision-making? If precision is lacking, is this because the study is under-powered?

Section 5: summary

5.1 Are the study results internally valid (that is, unbiased)?

How well did the study minimise sources of bias (that is, any factor that skews the data in one particular direction) and confounding?

Were there significant flaws in the study design?

5.2 Are the findings generalisable to the source population (that is, externally valid)?

Are there sufficient details given about the study to determine if the findings are generalisable to the source population? Consider participants, interventions and comparisons, outcomes, resource and policy implications.

References

- Assmann SF, Pocock SJ, Enos LE et al. (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet* 355 (9209):1064–9
- Effective Public Health Practice Project 'Quality assessment tool for quantitative studies' [online]. Available from: www.myhamilton.ca/myhamilton/CityandGovernment/HealthandSocialServices/Research/EPHPP/ [Accessed 18 March 2008]
- Guillemin F (2007) Primer: the fallacy of subgroup analysis. *Nature Clinical Practice Rheumatology* 3 (7): 407–13
- Heller RF, Verma A, Gemmell I et al. (2008) Critical appraisal for public health: a new checklist. *Public Health* 122 (1): 92–8
- Jackson R, Ameratunga S, Broad J et al. (2006) The GATE frame: critical appraisal with pictures. *Evidence Based Medicine* 11 (2): 35–8
- Kavanagh J, Oliver S, Lorenc T (2008) Reflections on developing and using PROGRESS-Plus. *Equity Update* 2: 1–3 [online]. Available from http://equity.cochrane.org/Files/Equity_Update_Vol2_Issue1.pdf [accessed 13 February 2009]
- Rothwell PM (2005) External validity of randomised controlled trials: to whom do the results of this trial apply? *Lancet* 365 (9453): 82–93

Appendix G Quality appraisal checklist – quantitative studies reporting correlations and associations¹⁸

A correlates review (see section 3.3.4) attempts to establish the factors that are associated or correlated with positive or negative health behaviours or outcomes. Evidence for correlate reviews will come both from specifically designed correlation studies and other study designs which also report on correlations.

A critical appraisal form specific for assessing the validity of studies reporting correlations has been developed based on the appraisal step of the 'Graphical appraisal tool for epidemiological studies (GATE)', developed by Jackson et al. (2006).

The critical appraisal form enables a reviewer to appraise a study's internal and external validity after addressing the following key aspects of study design: characteristics of study participants; definition of independent variables; outcomes assessed and method(s) of analyses.

Like GATE, the critical appraisal tool is intended to be used in an electronic (Excel) format which will facilitate both the sharing and storage of data, and through linkage with other documents, the compilation of research reports. Much of the guidance to support the completion of the critical appraisal form appears in 'pop-up' windows in the electronic version¹⁹.

There are five sections of the revised GATE. Section 1 seeks to assess the key population criteria for determining the study's **external validity** – that is, the extent to which the findings of a study are generalisable beyond the confines of the study to the study's source population.

Sections 2 to 4 assess the key criteria for determining the study's **internal validity** – that is, making sure that the study has been carried out carefully, and that the identified associations are valid and are not due to some other (often unidentified) factor.

¹⁸ Appraisal form derived from: Jackson R, Ameratunga S, Broad J et al. (2006) The GATE frame: critical appraisal with pictures. *Evidence Based Medicine* 11 (2): 35–8.

¹⁹ Available from CPHE on request.

Checklist items are worded so that one of five responses is possible:

- ++ Indicates that for that particular aspect of study design, the study has been designed/conducted in such a way as to minimise the risk of bias.
- + Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias for that particular aspect of study design.
- Should be reserved for those aspects of the study design in which significant sources of bias may persist.
- Not reported (nr) Should be reserved for those aspects in which the study under review fails to report how they have/might have been considered.
- Not applicable (na) Should be reserved for those study design aspects which are not applicable given the study design under review (for example, allocation concealment would not be applicable for case–control studies).

In addition, the reviewer is requested to complete in detail the comments section of the quality appraisal form so that the grade awarded for each study aspect is as transparent as possible.

Each study is then awarded an overall study quality grading for internal validity (IV) and a separate one for external validity (EV):

- ++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
- + Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Checklist

Study identification (Include full citation details)		
Study design: Refer to the the glossary of study designs (appendix D) and the algorithm for classifying experimental and observational study designs (appendix E) to best describe the paper's underpinning study design		
Guidance topic:		
Assessed by:		
Section 1: Population		
<p>1.1 Is the source population or source area well described?</p> <p>Was the country (e.g. developed or non-developed, type of healthcare system), setting (primary schools, community centres etc.), location (urban, rural), population demographics etc. adequately described?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>1.2 Is the eligible population or area representative of the source population or area?</p> <p>Was the recruitment of individuals/clusters/areas well defined (e.g. advertisement, birth register)? Was the eligible population representative of the source? Were important groups underrepresented?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>1.3 Do the selected participants or areas represent the eligible population or area?</p> <p>Was the method of selection of participants from the eligible population well described? What % of selected individuals/clusters agreed to participate? Were there any sources of bias? Were the inclusion/exclusion criteria explicit and appropriate?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:

Section 2: Method of selection of exposure (or comparison) group		
<p>2.1 Selection of exposure (and comparison) group. How was selection bias minimised?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>2.2 Was the selection of explanatory variables based on a sound theoretical basis? How sound was the theoretical basis for selecting the explanatory variables?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>2.3 Was the contamination acceptably low? Did any in the comparison group receive the exposure? If so, was it sufficient to cause important bias?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause important bias?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>2.5 Is the setting applicable to the UK? Did the setting differ significantly from the UK?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:

Section 3: Outcomes		
<p>3.1 Were the outcome measures and procedures reliable?</p> <p>Were outcome measures subjective or objective (e.g. biochemically validated nicotine levels [++] vs self-reported smoking [-]).</p> <p>How reliable were outcome measures (e.g. inter- or intra-rater reliability scores)?</p> <p>Was there any indication that measures had been validated (e.g. against a gold standard measure or assessed for content validity)?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.2 Were the outcome measurements complete?</p> <p>Were all or most of the study participants who met the defined study outcome definitions likely to have been identified?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.3 Were all the important outcomes assessed?</p> <p>Were all the important benefits and harms assessed?</p> <p>Was it possible to determine the overall balance of benefits and harms of the intervention versus comparison?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
<p>3.4 Was there a similar follow-up time in exposure and comparison groups?</p> <p>If groups are followed for different lengths of time, then more events are likely to occur in the group followed-up for longer distorting the comparison.</p> <p>Analyses can be adjusted to allow for differences in length of follow-up (e.g. using person-years).</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>

<p>3.5 Was follow-up time meaningful? Was follow-up long enough to assess long-term benefits and harms? Was it too long, e.g. participants lost to follow-up?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
Section 4: Analyses		
<p>4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 (i.e. it is likely to see an effect of a given size if one exists, 80% of the time) is the conventionally accepted standard. Is a power calculation presented? If not, what is the expected effect size? Is the sample size adequate?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>4.2 Were multiple explanatory variables considered in the analyses? Were there sufficient explanatory variables considered in the analysis?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:
<p>4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	Comments:

<p>4.4 Was the precision of association given or calculable? Is association meaningful?</p> <p>Were confidence intervals (CIs) and/or p-values for effect estimates given or possible to calculate?</p> <p>Were CIs wide or were they sufficiently precise to aid decision-making? If precision is lacking, is this because the study is under-powered?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> – <input type="checkbox"/> NR <input type="checkbox"/> NA	<p>Comments:</p>
--	---	------------------

Section 5: Summary

<p>5.1 Are the study results internally valid (i.e. unbiased)?</p> <p>How well did the study minimise sources of bias (i.e. adjusting for potential confounders)?</p> <p>Were there significant flaws in the study design?</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> –	<p>Comments:</p>
<p>5.2 Are the findings generalisable to the source population (i.e. externally valid)?</p> <p>Are there sufficient details given about the study to determine if the findings are generalisable to the source population? Consider: participants, interventions and comparisons, outcomes, resource and policy implications.</p>	<input type="checkbox"/> ++ <input type="checkbox"/> + <input type="checkbox"/> –	<p>Comments:</p>

Appendix H Quality appraisal checklist – qualitative studies²⁰

There is considerable debate over what quality criteria should be used to assess qualitative studies. Quality in qualitative research can be assessed using the same broad concepts of validity (or trustworthiness) used for quantitative research, but these need to be put in a different contextual framework to take into account the aims of qualitative research.

This qualitative checklist is designed for people with a basic understanding of qualitative research methodology, and is based on the broadly accepted principles that characterise qualitative research and which may affect its validity. The following notes provide suggestions for completing the checklist. A list of publications on qualitative research is provided at the end of these notes for further reading on this topic.

The studies covered by this checklist are studies which collect and analyse qualitative data, usually (but not exclusively) textual (written), spoken or observational data. Qualitative data are occasionally collected by structured questionnaires (for example, as thematically organised free text comments), but such data needs to be carefully scrutinised as it may not meet acceptable quality criteria for consideration as a qualitative study.

The checklist's questions are framed in such a way so that it can encompass the variety of ways qualitative research is conducted. Care must be taken to apply the checklist in a way which matches the research methodology.

Please note that the subquestions given as examples under each question are intended to highlight some of the key issues to be considered for that question. They are not intended to be exhaustive. Please add any additional considerations in the comments box.

Notes on the completion of the separate sections of the checklist are appended to it.

In some circumstances it may be necessary to analyse qualitative material using a different approach, where the goal will be to seek to extract underlying theories, propositions and principles from the data, rather than focusing on the quality of the study per se. This may be appropriate where the aim is to gain particular insights into social processes. Where developments of the processes of appraisal are required these will be discussed with the CPHE team.

²⁰ This checklist is based on checklists in:

Spencer L, Ritchie J, Lewis J et al. (2003) Quality in qualitative evaluation: a framework for assessing research evidence. London: Government Chief Social Researcher's Office

Public Health Resource Unit (2006) 10 questions to help you make sense of qualitative research [online]. Available at www.phru.nhs.uk/Doc_Links/Qualitative%20Appraisal%20Tool.pdf [accessed 7 July 2008]

National Training and Research Appraisal Group (NTRAG) at www.ntrag.co.uk

British Sociological Association (BSA) at www.britsoc.co.uk

Checklist

Study identification Include author, title, reference, year of publication		
Guidance topic:	Key research question/aim:	
Checklist completed by:		
Theoretical approach		
<p>1. Is a qualitative approach appropriate? For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question?</p>	<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? For example: Is the purpose of the study discussed – aims/objectives/research question/s? Is there adequate/appropriate reference to the literature? Are underpinning values/assumptions/theory discussed?</p>	<input type="checkbox"/> Clear <input type="checkbox"/> Unclear <input type="checkbox"/> Mixed	Comments:

Study design		
<p>3. How defensible/rigorous is the research design/methodology?</p> <p>For example:</p> <p>Is the design appropriate to the research question?</p> <p>Is a rationale given for using a qualitative approach?</p> <p>Are there clear accounts of the rationale/justification for the sampling, data collection and data analysis techniques used?</p> <p>Is the selection of cases/sampling strategy theoretically justified?</p>	<p><input type="checkbox"/> Defensible</p> <p><input type="checkbox"/> Indefensible</p> <p><input type="checkbox"/> Not sure</p>	<p>Comments:</p>
Data collection		
<p>4. How well was the data collection carried out?</p> <p>For example:</p> <p>Are the data collection methods clearly described?</p> <p>Were the appropriate data collected to address the research question?</p> <p>Was the data collection and record keeping systematic?</p>	<p><input type="checkbox"/> Appropriately</p> <p><input type="checkbox"/> Inappropriately</p> <p><input type="checkbox"/> Not sure/ inadequately reported</p>	<p>Comments:</p>
Trustworthiness		
<p>5. Is the role of the researcher clearly described?</p> <p>For example:</p> <p>Has the relationship between the researcher and the participants been adequately considered?</p> <p>Does the paper describe how the research was explained and presented to the participants?</p>	<p><input type="checkbox"/> Clearly described</p> <p><input type="checkbox"/> Unclear</p> <p><input type="checkbox"/> Not described</p>	<p>Comments:</p>

<p>6. Is the context clearly described? For example: Are the characteristics of the participants and settings clearly defined? Were observations made in a sufficient variety of circumstances? Was context bias considered?</p>	<p><input type="checkbox"/> Clear <input type="checkbox"/> Unclear <input type="checkbox"/> Not sure</p>	<p>Comments:</p>
<p>7. Were the methods reliable? For example: Was data collected by more than one method? Is there justification for triangulation, or for not triangulating? Do the methods investigate what they claim to?</p>	<p><input type="checkbox"/> Reliable <input type="checkbox"/> Unreliable <input type="checkbox"/> Not sure</p>	<p>Comments:</p>
<p>Analysis</p>		
<p>8. Is the data analysis sufficiently rigorous? For example: Is the procedure explicit – i.e. is it clear how the data was analysed 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?</p>	<p><input type="checkbox"/> Rigorous <input type="checkbox"/> Not rigorous <input type="checkbox"/> Not sure/ not reported</p>	<p>Comments:</p>
<p>9. Are the data 'rich'? For example: How well are the contexts of the data described? Has the diversity of perspective and content been explored? How well has the detail and depth been demonstrated? Are responses compared and contrasted across groups/sites?</p>	<p><input type="checkbox"/> Rich <input type="checkbox"/> Poor <input type="checkbox"/> Not sure/ not reported</p>	<p>Comments:</p>

<p>10. Is the analysis reliable?</p> <p>For example: Did more than one researcher theme and code transcripts/data? If so, how were differences resolved? Did participants feed back on the transcripts/data if possible and relevant? Were negative/discrepant results addressed or ignored?</p>	<p><input type="checkbox"/> Reliable</p> <p><input type="checkbox"/> Unreliable</p> <p><input type="checkbox"/> Not sure/ not reported</p>	<p>Comments:</p>
<p>11. Are the findings convincing?</p> <p>For example: Are the findings clearly presented? Are the findings internally coherent? Are extracts from the original data included? Are the data appropriately referenced? Is the reporting clear and coherent?</p>	<p><input type="checkbox"/> Convincing</p> <p><input type="checkbox"/> Not convincing</p> <p><input type="checkbox"/> Not sure</p>	<p>Comments:</p>
<p>12. Are the findings relevant to the aims of the study?</p>	<p><input type="checkbox"/> Relevant</p> <p><input type="checkbox"/> Irrelevant</p> <p><input type="checkbox"/> Partially relevant</p>	<p>Comments:</p>

<p>13. Conclusions</p> <p>For example:</p> <p>How clear are the links between data, interpretation and conclusions?</p> <p>Are the conclusions plausible and coherent?</p> <p>Have alternative explanations been explored and discounted?</p> <p>Does this enhance understanding of the research topic?</p> <p>Are the implications of the research clearly defined?</p> <p>Is there adequate discussion of any limitations encountered?</p>	<p><input type="checkbox"/> Adequate</p> <p><input type="checkbox"/> Inadequate</p> <p><input type="checkbox"/> Not sure</p>	<p>Comments:</p>
<p>Ethics</p>		
<p>14. How clear and coherent is the reporting of ethics?</p> <p>For example:</p> <p>Have ethical issues been taken into consideration?</p> <p>Are they adequately discussed e.g. do they address consent and anonymity?</p> <p>Have the consequences of the research been considered i.e. raising expectations, changing behaviour?</p> <p>Was the study approved by an ethics committee?</p>	<p><input type="checkbox"/> Appropriate</p> <p><input type="checkbox"/> Inappropriate</p> <p><input type="checkbox"/> Not sure/ not reported</p>	<p>Comments:</p>
<p>Overall assessment</p>		
<p>As far as can be ascertained from the paper, how well was the study conducted (see guidance notes)?</p>	<p><input type="checkbox"/> ++</p> <p><input type="checkbox"/> +</p> <p><input type="checkbox"/> -</p>	<p>Comments:</p>

Notes on the use of the qualitative studies checklist

Section 1: theoretical approach

This section deals with the underlying theory and principles applied to the research.

1. Is a qualitative approach appropriate?

A qualitative approach can be judged to be appropriate when the research sets out to investigate phenomena which are not easy to accurately quantify or measure, or where such measurement would be arbitrary and inexact. If clear numerical measures could reasonably have been put in place then consider whether a quantitative approach may have been more appropriate. This is because most qualitative research seeks to explain the meanings which social actors use in their everyday lives rather than the meanings which the researchers bring to the situation.

Qualitative research in public health commonly measures:

- personal/lives experiences (for example, of a condition, treatment, situation)
- processes (for example, action research, practitioner/patient views on the acceptability of using new technology)
- personal meanings (for example, about death, birth, disability)
- interactions/relationships (for example, the quality of the GP/patient relationship, the openness of a psychotherapeutic relationship)
- service evaluations (for example, what was good/bad about patients' experiences of a smoking cessation group).

2. Is the study clear in what it seeks to do?

Qualitative research designs tend to be theory generative rather than theory testing; therefore it is unlikely that a research question will be found in the form of a hypothesis or null hypothesis in the way that you would expect in conventional quantitative research. This does not mean however that the paper should not set out early and clearly what it is that the study is investigating and what the parameters are for that. The research question should be set in context by the provision of an adequate summary of the background literature and of the study's underpinning values and assumptions.

Section 2: study design

Considers the robustness of the design of the research project.

3. How defensible is the research design?

There are a large number of qualitative methodologies, and a tendency in health to 'mix' aspects of different methodologies or to use a generic qualitative method. From a qualitative perspective, none of this compromises the quality of a study as long as:

- The research design captures appropriate data and has an appropriate plan of analysis for the subject under investigation. There should be a clear and reasonable justification for the methods chosen.
- The choice of sample and sampling method should be clearly set out, (ideally including any shortcomings of the sample) and should be reasonable. It is important to remember that sampling in qualitative research can be purposive and should not be random. Qualitative research is not experimental, does not purport to be generalisable, and therefore does not require a large or random sample. People are usually 'chosen' for qualitative research based on being key informers.

Section 3: data collection

4. How well was the data collection carried out?

Were the method of data collection the most appropriate given the aims of the research? Was the data collection robust? Are there details of:

- how the data were collected?
- how the data were recorded and transcribed (if verbal data)?
- how the data were stored?
- what records were kept of the data collection?

Section 4: trustworthiness

Assessing the validity of qualitative research is very different from quantitative research. Qualitative research is much more focused on demonstrating the causes of bias rather than eliminating them, as a result it is good practice to include sections in the report about the reflexive position of the researcher (what was their 'part' in the research?), about the context in which the research was conducted, and about the reliability of the data themselves.

5. Is the role of the researcher clearly described?

The researcher should have considered their role in the research either as reader, interviewer, or observer, for example. This is often referred to as 'reflexivity'. It is important that we can determine: a clear audit trail from respondent all the way through to reporting, why the author reported what they did report, and that we can follow the reasoning from the data to the final analysis or theory.

The 'status' of the researcher can profoundly affect the data, for example, a middle-aged woman and a teenage man are likely to get different responses to questions about sexual activity if they interview a group of teenage boys. It is important to consider age, gender, ethnicity, 'insider' status (where the interviewer/researcher is part of the group being researched or has the same condition/illness, for example). The researcher can also profoundly influence the data by use of questions, opinions and judgments, so it is important to know what the researcher's position is in that regard and how the researcher introduced and talked about the research with the participants.

6. Is the context clearly described?

It is important when gauging the validity of qualitative data to engage with the data in a meaningful way, and to consider whether the data are plausible/realistic. To make an accurate assessment of this it is important to have information about the context of the research, not only in terms of the physical context – for example, youth club, GP surgery, gang headquarters, who else was there (discussion with parents present or discussion with peers present are likely to cause the participant to position himself very differently and thus to respond very differently), but also in terms of feeling that the participants are described in enough detail that the reader can have some sort of insight into their life/situation. Any potential context bias should be considered.

7. Were the methods reliable?

It is important that the method used to collect the data is appropriate for the research question, and that the data generated map well onto the aims of the study. Ideally, more than one method should have been used to collect data, or there should be some other kind of system of comparison which allows the data to be compared. This is referred to as triangulation.

Section 5: analysis

Qualitative data analysis is very different from quantitative analysis. This does not mean that it should not be systematic and rigorous but systematicity and rigour require different methods of assessment.

8. Is the data analysis sufficiently rigorous?

The main way to assess this is by how clearly the analysis is reported and whether the analysis is approached systematically. There should be a clear and consistent method for coding and analysing data, and it should be clear how the coding and analytic strategies were derived. Above all, these must be reasonable in light of the evidence and the aims of the study. Transparency is the key to addressing the rigour of the analysis.

9. Are the data rich?

Qualitative researchers use the adjective 'rich' to describe data which is in-depth, convincing, compelling and detailed enough that the reader feels that they have achieved some level of insight into the research participants' experience. It is also important to know the 'context' of the data, that is, where it came from, what prompted it and what it pertains to.

10. Is the analysis reliable?

The analysis of data can be made more reliable by setting checks in place. It is good practice to have sections of data coded by another researcher, or at least have a second researcher check the coding for consistency. Participants may also be allowed to verify the transcripts of their interview (or other data collection, if appropriate). Negative/discrepant results should always be highlighted and discussed.

11. Are the findings convincing?

In qualitative research, the reader should find the results of the research convincing, or credible. This means that the findings should be clearly presented and logically organised, that they should not contradict themselves without explanation or consideration and that they should be clear and coherent.

Extracts from original data should be included where possible to give a fuller sense of the findings, and these data should be appropriately referenced – although you would expect data to be anonymised, it still needs to be referenced in relevant ways, for example if gender differences were important then you would expect extracts to be marked male/female.

12–13. Relevance of findings and conclusions

These sections are self explanatory.

Section 6: ethics

14. How clear and coherent is the reporting of ethics?

All qualitative research has ethical considerations and these should be considered within any research report. Ideally there should be a full discussion of ethics, although this is rare because of space limitations in peer-reviewed journals. If there are particularly fraught ethical issues raised by a particularly sensitive piece of research, then these should be discussed in enough detail that the reader is convinced that every care was taken to protect research participants.

Any research with human participants should be approved by a research ethics committee and this should be reported.

Section 7: overall assessment

15. Is the study relevant?

Does the study cast light on the review being undertaken?

16. How well was the study conducted?

Grade the study according to the list below:

++	All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
+	Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
–	Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Appendix I Quality appraisal checklist – economic evaluations

This checklist is designed to determine whether an economic evaluation provides evidence that is useful to inform the decision-making of the Public Health Interventions Advisory Committee (PHIAC) and the Programme Development Groups (PDGs). It is not intended to judge the quality of the study or the quality of reporting.

Study identification		
Including author, title, reference, year of publication		
Guideline topic:	Question no:	
Checklist completed by:		
Section 1: Applicability (relevance to specific topic review question[s] and the NICE reference case)²¹		Yes/ partly/no/ unclear/not applicable
This checklist should be used first to filter out irrelevant studies		Comments
1.1	Is the study population appropriate for the topic being evaluated?	
1.2	Are the interventions appropriate for the topic being evaluated?	
1.3	Is the system in which the study was conducted sufficiently similar to the current UK context?	
1.4	Was/were the perspective(s) clearly stated and what were they?	
1.5	Are all direct health effects on individuals included, and are all other effects included where they are material?	
1.6	Are all future costs and outcomes discounted appropriately?	
1.7	Is the value of health effects expressed in terms of quality-adjusted life years (QALYs)?	
1.8	Are costs and outcomes from other sectors fully and appropriately measured and valued?	
Overall judgement: directly applicable/partially applicable/not applicable There is no need to complete section 2 of the checklist if the study is considered 'not applicable'.		
Other comments:		
²¹ As detailed in the 'Guide to the methods of technology appraisal' (June 2008), table 5.1 (page 30). Section 5.2.3 of the guide states: 'There may be important barriers to applying reference-case methods. In these cases, the reasons for a failure to meet the reference case should be clearly specified and justified, and the likely implications should, as far as possible, be quantified.'		

Section 2: Study limitations (the level of methodological quality)		Yes/ partly/no/ unclear/NA	Comments
This checklist should be used once it has been decided that the study is sufficiently applicable to the context of the clinical guideline²²			
2.1	Does the model structure adequately reflect the nature of the topic under evaluation?		
2.2	Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?		
2.3	Are all important and relevant outcomes included?		
2.4	Are the estimates of baseline outcomes from the best available source?		
2.5	Are the estimates of relative 'treatment' effects from the best available source?		
2.6	Are all important and relevant costs included?		
2.7	Are the estimates of resource use from the best available source?		
2.8	Are the unit costs of resources from the best available source?		
2.9	Is an appropriate incremental analysis presented or can it be calculated from the data?		
2.10	Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?		
2.11	Is there any potential conflict of interest?		
2.12	Overall assessment: minor limitations/potentially serious limitations/very serious limitations		
Other comments:			
²² The items and notes in this checklist have been developed from: 'Guidelines manual' (appendix H) (NICE 2009); 'Guide to the methods of technology assessment' (NICE 2008); Evers S, Goossens M, de Vet H et al. (2005) Criteria list for assessment of methodological quality of economic evaluations – CHEC. International Journal of Technology Assessment in Health Care 21: 240–5; Philips Z, Ginnelly L, Sculpher M et al. (2004) Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technology Assessment 8. Available from www.hta.ac.uk/fullmono/mon836.pdf			

Notes on the use of the economic evaluations checklist

For all questions:

- answer 'yes' if the study fully meets the criterion
- answer 'partly' if the study largely meets the criterion but differs in some important respect
- answer 'no' if the study deviates substantively from the criterion
- answer 'unclear' if the report provides insufficient information to judge whether the study complies with the criterion
- answer 'NA (not applicable)' if the criterion is not relevant in a particular instance.

For 'partly' or 'no' responses, use the comments column to explain how the study deviates from the criterion.

Section 1: applicability

1.1 Is the study population appropriate for the guideline?

The study population should be defined as precisely as possible and should be in line with that specified in the topic scope and any related review protocols.

This includes consideration of appropriate subgroups that require special attention. For many interventions, the capacity to benefit will differ for participants or communities with differing characteristics. This should be explored separately for each relevant subgroup as part of the base-case analysis by the provision of estimates of effectiveness and cost effectiveness. The characteristics of participants or communities in each subgroup should be clearly defined and, ideally, should be identified on the basis of an assumed expectation of differential effectiveness or cost effectiveness as a result of biologically, economic or sociologically plausible known mechanisms, social characteristics or other clearly justified factors.

Answer 'yes' if the study population is fully in line with that in the topic question(s) and if the study differentiates appropriately between important subgroups. Answer 'partly' if the study population is similar to that in the topic question(s) but: (i) it differs in some important respects; or (ii) the study fails to differentiate between important subgroups. Answer 'no' if the study population is substantively different from that in the topic question(s).

1.2 Are the interventions appropriate for the guideline?

All relevant alternatives should be included, as specified in the topic scope and any related review protocols. These should include routine and best practice in the NHS, existing NICE guidance and other feasible options.

Answer 'yes' if the analysis includes all options considered relevant for the topic, even if it also includes other options that are not relevant. Answer 'partly' if the analysis omits one or more relevant options but still contains comparisons likely to be useful for the guidance. Answer 'no' if the analysis does not contain any relevant comparisons.

1.3 Is the system in which the study was conducted sufficiently similar to the current UK NHS context?

This relates to the overall institutional structure within which the interventions were delivered. For example, an intervention might be delivered on an inpatient basis in one country whereas in the UK it would be provided in the community. This might significantly influence the use of healthcare resources and costs, thus limiting the applicability of the results to a UK setting. In addition, old UK studies may be severely limited in terms of their relevance to current NHS practice. Other institutional structures include the differing legislative and legal systems, different working and sick-leave provisions, different social arrangements.

Answer 'yes' if the study was conducted within the UK and is sufficiently recent to reflect current NHS and other relevant public, voluntary and private sector practice. For non-UK or older UK studies, answer 'partly' if differences in the setting are unlikely to substantively change the cost-effectiveness estimates. Answer 'no' if the setting is so different that the results are unlikely to be applicable in the current NHS or in other public, voluntary or private sectors.

1.4 Was/were the perspective(s) clearly stated?

The decision-making perspective of an economic evaluation determines the range of costs that should be included in the analysis. For public health guidance, one perspective that will usually be used is that of the NHS and personal social services (PSS). Often in public health, costs and benefits will be borne outside the NHS. When they are borne predominantly by other public sectors, it may also be appropriate to use a public sector perspective. In topics where interventions have a material effect on employment, the perspective may also need to reflect that. Where the cost effectiveness using a narrower perspective is clearly established, however, the requirement to embrace a wider perspective is much reduced.

Answer 'yes' if the study clearly and correctly states the perspective used, and that perspective is appropriate. Answer 'partly' if the perspective stated is not the perspective used. Answer 'no' if the study does not state the perspective or the perspective is not appropriate.

1.5 Are all direct health effects on individuals included, and are all other effects included where they are material?

For an NHS/PSS perspective, outcomes should include all direct health effects, whether for patients or, when relevant, other people (principally carers). This is consistent with an objective of maximising health gain from available healthcare resources. Some features of healthcare delivery that are often referred to as 'process characteristics' may ultimately have health consequences; for example, the mode of treatment delivery may have health consequences through its impact on concordance with treatment. Any significant characteristics of healthcare technologies that have a value to people that is independent of any direct effect on health should be noted. These characteristics include the convenience with which healthcare is provided and the level of information available for patients.

Answer 'yes' if the measure of health outcome used for an NHS/PSS perspective excludes non-health effects (or if such effects can be excluded from the results) **and** if non-health effects are included in a wider perspective. Answer 'partly' if the analysis from an NHS/PSS perspective includes some non-health effects but these are small and unlikely to change the cost-effectiveness results. Answer 'no' if the analysis incorrectly includes or excludes significant non-health effects that are likely to change the cost-effectiveness results for a particular perspective.

1.6 Are all future costs and effects discounted appropriately?

The need to discount to a present value is widely accepted in economic evaluation, although the specific rate varies across jurisdictions and over time. There is currently some debate about the most appropriate discount rate for health benefits. The annual rate of 3.5% should be applied to both costs and effects, but an annual rate of 1.5% for health benefits and 3.5% for costs should routinely be used in sensitivity analysis.

Answer 'yes' if both costs and health effects (for example, quality-adjusted life years [QALYs]) are discounted at 3.5% per year, or if the health benefits are discounted at an annual rate of 1.5%. Answer 'partly' if costs are discounted at a rate similar to 3.5% and if effects are discounted at a rate between 1.5% to 3.5% per year (for example, costs and effects are both discounted at 3% per year). Answer 'no' if costs and/or health effects are not discounted, or if they are discounted at a rate (or rates) different from 3.5% (for example, 5% for both costs and effects). Note in the comments column what discount rates have been used. If all costs and health effects accrue within a short time (roughly a year), answer 'not applicable'.

1.7 Is the value of health effects expressed in terms of QALYs?

The QALY is a measure of a person's length of life weighted by a valuation of their health-related quality of life (HRQL) over that period.

Given its widespread use, the QALY is considered by NICE to be the most appropriate generic measure of health benefit that reflects both mortality and effects on HRQL. It is recognised that alternative measures exist (such as the healthy-year equivalent), but few economic evaluations have used these methods and their strengths and weaknesses are not fully established.

NICE's position is that an additional QALY should be given the same weight regardless of the other characteristics of the patients receiving the health benefit.

Answer 'yes' if the health benefits of the intervention are measured using QALYs; answer 'no' if not. There may be circumstances when a QALY cannot be obtained or where the assumptions underlying QALYs are considered inappropriate. In such situations answer 'no', but consider retaining the study for appraisal. Similarly, answer 'no' but retain the study for appraisal if it does not include QALYs but it is still thought to be useful for PHIAc/the PDG decision-making: for example, if the evidence indicates that an intervention might be dominant, and estimates of the relative costs of the interventions from a cost-minimisation study are likely to be useful. When economic evaluations not using QALYs are retained for full critical appraisal, use the comments column to note why. Retention of a study for appraisal will also depend on the extent and reliability of other evidence. In the absence of good alternative sources of evidence, it might be useful to retain a study.

1.8 Are costs and outcomes from other sectors fully and appropriately measured and valued?

Studies in public health will often include costs accruing to other sectors of the economy and/or benefits gained by these sectors. Not all of these benefits will be amenable to translation into QALYs (for example, a reduction in school class size, the reduction in unwanted pregnancies or the ability to return to work earlier). Answer 'yes' if all the costs and all the benefits have been included, if they are appropriately measured and if they are appropriately valued. Answer 'partly' if omissions are not material and answer 'no' if some major cost or benefit is omitted, is improperly measured or improperly valued.

1.9 Overall judgement

Classify the applicability of the economic evaluation to the public health guidance, the current public sector situation and the context for NICE guidance as one of the following:

- **Not applicable** The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. Such studies would be excluded from further consideration and there is no need to continue with the rest of the checklist. The outcome to exclude the study will depend on the number of other cost-effectiveness studies that relate to the topic, and their applicability.

- **Partially applicable** The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness.
- **Directly applicable** The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness.

Section 2: study limitations

2.1 Does the model structure adequately reflect the nature of the topic under evaluation?

This relates to the choice of model and its structural elements (including cycle length in discrete time models, if appropriate). Model type and its structural aspects should be consistent with a coherent theory of the health condition under evaluation. The selection of treatment pathways, whether health states or branches in a decision tree, should be based on the underlying biological, economic or sociological processes of the topic under study and the potential impact (benefits and adverse consequences) of the intervention(s) of interest.

Answer 'yes' if the model design and assumptions appropriately reflect the topic and intervention(s) of interest. Answer 'partly' if there are aspects of the model design or assumptions that do not fully reflect the topic or intervention(s) but that are unlikely to change the cost-effectiveness results. Answer 'no' if the model omits some important aspect of the topic or intervention(s) and this is likely to change the cost-effectiveness results. Answer 'NA' for economic evaluations based on data from a clinical study which do not extrapolate treatment outcomes or costs beyond the study context or follow-up period.

2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?

The time horizon is the period of analysis of the study: the length of follow-up for participants in a trial-based evaluation, or the period of time over which the costs and outcomes for a cohort are tracked in a modelling study. This time horizon should always be the same for costs and outcomes, and should be long enough to include all relevant costs and outcomes relating to the intervention. A time horizon shorter than lifetime could be justified if there is no differential mortality effect between options, and the differences in costs and HRQL relate to a relatively short period (for example, in the case of an acute infection).

Answer 'yes' if the time horizon is sufficient to include all relevant costs and outcomes. Answer 'partly' if the time horizon may omit some relevant costs and outcomes but these are unlikely to change the cost-effectiveness results. Answer 'no' if the time horizon omits important costs and outcomes and this is likely to change the cost-effectiveness results.

2.3 Are all important and relevant outcomes included?

All relevant outcomes should include direct health and/or other effects relating to harms from the intervention (adverse effects) as well as any potential benefits.

Answer 'yes' if the analysis includes all relevant and important harms and benefits. Answer 'partly' if the analysis omits some harms or benefits but these would be unlikely to change the cost-effectiveness results. Answer 'no' if the analysis omits important harms and/or benefits that would be likely to change the cost-effectiveness results.

2.4 Are the estimates of baseline outcomes from the best available source?

The sources and methods for eliciting baseline probabilities should be described clearly. These data might be based on 'natural history' (patient outcomes in the absence of treatment or with routine care), sourced from cohort studies. Baseline probabilities may also be derived from the control arms of experimental studies. Sometimes it may be necessary to rely on expert opinion for particular parameters.

Answer 'yes' if the estimates of baseline outcomes reflect the best available evidence as identified from a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates are not derived from a systematic review but are likely to reflect outcomes for the relevant group of people in England (for example, if they are derived from a large UK-relevant cohort study). Answer 'no' if the estimates are unlikely to reflect outcomes for the relevant group in England.

2.5 Are the estimates of relative treatment effects from the best available source?

Evidence on outcomes should be obtained from a systematic review, defined as the systematic location, inclusion, appraisal and synthesis of evidence to obtain a reliable and valid overview of the data relating to a clearly formulated question.

Synthesis of outcome data through meta-analysis is appropriate provided that there are sufficient relevant and valid data obtained using comparable measures of outcome.

Head-to-head randomised controlled trials (RCTs) provide the most valid evidence of relative treatment effect within the medical paradigm, but might not be as valid as other well-established forms of evidence from other paradigms (such as the observational evidence for the need for a parachute when jumping from a plane or the observation that raising a price will reduce a commodity's consumption). Therefore, data from non-randomised studies may supplement RCT data. Any potential bias arising from the design of the studies used in the assessment should be explored and documented.

When multiple interventions are being assessed that have not been compared within a single RCT, data from a series of pairwise head-to-head RCTs should be presented. Consideration should also be given to presenting a combined analysis using a mixed treatment comparison framework if it is considered to add information that is not available from the head-to-head comparison.

The principles of good practice for standard meta-analyses should be followed in mixed and indirect treatment comparisons.

The methods and assumptions that are used to extrapolate short-term results to final outcomes should be clearly presented.

Evidence for the evaluation of diagnostic technologies should normally incorporate evidence on diagnostic accuracy. As for other technologies, RCTs have the potential to capture the pathway of care involving diagnostic technologies, but their feasibility and availability may be limited. Other study designs should be assessed on the basis of their fitness for purpose, taking into consideration the aim of the study (for example, to evaluate outcomes, or to evaluate sensitivity and specificity) and the purpose of the diagnostic technology.

Answer 'yes' if the estimates of treatment effect appropriately reflect all relevant studies of the best available quality, as identified through a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates of treatment effect are not derived from a systematic review but are similar in magnitude to the best available estimates (for example, if the economic evaluation is based on a single large study with treatment effects similar to pooled estimates from all relevant studies). Answer 'no' if the estimates of treatment effect are likely to differ substantively from the best available estimates.

2.6 Are all important and relevant costs included?

Costs related to the condition of interest and incurred in additional years of life gained as a result of treatment should be included in the base-case analysis. This should include the costs of handling non-adherence to treatment and treating side effects. Costs that are considered to be unrelated to the topic or intervention of interest should be excluded. If introduction of the intervention requires additional infrastructure to be put in place, consideration should be given to including such costs in the analysis.

Answer 'yes' if all important and relevant resource use and costs are included given the perspective and the research question under consideration. Answer 'partly' if some relevant resource items are omitted but these are unlikely to affect the cost-effectiveness results. Answer 'no' if important resource items are omitted and these are likely to affect the cost-effectiveness results.

2.7 Are the estimates of resource use from the best available source?

It is important to quantify the effect of the interventions on resource use in terms of physical units (for example, days in hospital, visits to a GP or hours spent in class strengthening behavioural resolve) and valuing those effects in monetary terms using appropriate prices and unit costs. Evidence on resource use should be identified systematically. When expert opinion is used as a source of information, any formal methods used to elicit these data should be clearly reported.

Answer 'yes' if the estimates of resource use appropriately reflect all relevant evidence sources of the best available quality, as identified through a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates of resource use are not derived from a systematic review but are similar in magnitude to the best available estimates. Answer 'no' if the estimates of resource use are likely to differ substantially from the best available estimates.

2.8 Are the unit costs of resources from the best available source?

Resources should be valued using the prices relevant to the NHS and PSS for health costs and in prices relevant to the respective sectors responsible for other costs. For the NHS/PSS perspective, it is appropriate for the financial costs relevant to the NHS/PSS to be used as the basis of costing, although these may not always reflect the full social opportunity cost of a given resource. A first point of reference in identifying costs and prices should be any current official listing published by the Department of Health and/or the Welsh Assembly Government.

When the acquisition price paid for a resource differs from the public list price (for example, pharmaceuticals and medical devices sold at reduced prices to NHS institutions), the public list price should be used in the base-case analysis. Sensitivity analysis should assess the implications of variations from this price.

National data based on healthcare resource groups (HRGs), such as the Payment by Results tariff, can be used when they are appropriate and available. However, data based on HRGs may not be appropriate in all circumstances (for example, when the definition of the HRG is broad, or the mean cost probably does not reflect resource use in relation to the intervention(s) under consideration). In such cases, other sources of evidence, such as micro-costing studies, may be more appropriate. When cost data are taken from the literature, the methods used to identify the sources should be defined. When several alternative sources are available, a justification for the costs chosen should be provided and discrepancies between the sources explained. When appropriate, sensitivity analysis should have been undertaken to assess the implications for results of using alternative data sources.

Similar rules apply to the costs of other public sectors.

Answer 'yes' if resources are valued using up-to-date prices relevant to the NHS/PSS, and to other sectors where the perspective allows this. Answer 'partly' if the valuations of some resource items differ from current NHS/PSS unit costs, or those of other sectors where the perspective allows this, but this is unlikely to change the cost-effectiveness results. Answer 'no' if the valuations of some resource items differ substantively from current NHS/PSS unit costs, or those of other sectors where the perspective allows this, and this is likely to change the cost-effectiveness results.

2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?

An appropriate incremental analysis is one that compares the expected costs and health outcomes of one intervention with the expected costs and health outcomes of the next-best non-dominated alternative.

Standard decision rules should be followed when combining costs and effects, and should reflect any situation where there is dominance or extended dominance. When there is a trade-off between costs and effects, the results should be presented as an incremental cost-effectiveness ratio (ICER): the ratio of the difference in mean costs to the difference in mean outcomes of a technology compared with the next best alternative. In addition to ICERs, the range of expected net monetary or health benefits can be presented using values placed on a QALY gained of £20,000 and £30,000.

For cost-consequence analyses, appropriate incremental analysis can only be done by selecting one of the consequences as the primary measure of effectiveness, providing the consequences are independent of one another.

Answer 'yes' if appropriate incremental results are presented, or if data are presented that allow the reader to calculate the incremental results. Answer 'no' if: (i) simple ratios of costs to effects are presented for each alternative compared with a standard intervention; or (ii) options subject to simple or extended dominance are not excluded from the incremental analyses.

2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?

There are a number of potential selection biases and uncertainties in any evaluation (trial- or model-based) and these should be identified and quantified where possible. There are three types of bias or uncertainty to consider:

- Structural uncertainty – for example in relation to the categorisation of different states of health and the representation of different pathways of care. These structural assumptions should be clearly documented and the evidence and rationale to support them provided. The impact of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of plausible scenarios.

- Source of values to inform parameter estimates – the implications of different estimates of key parameters (such as estimates of relative effectiveness) must be reflected in sensitivity analyses (for example, through the inclusion of alternative scenarios). Inputs must be fully justified, and uncertainty explored by sensitivity analysis using alternative input values.
- Parameter precision – uncertainty around the mean health and cost inputs in the model. Distributions should be assigned to characterise the uncertainty associated with the (precision of) mean parameter values. Probabilistic sensitivity analysis is preferred, as this enables the uncertainty associated with parameters to be simultaneously reflected in the results of the model. In non-linear decision models – when there is not a straight-line relationship between inputs and outputs of a model (such as Markov models) – probabilistic methods provide the best estimates of mean costs and outcomes. Simple decision trees are usually linear.
- The mean value, distribution around the mean, and the source and rationale for the supporting evidence should be clearly described for each parameter included in the model.
- Evidence about the extent of correlation between individual parameters should be considered carefully and reflected in the probabilistic analysis. Assumptions made about the correlations should be clearly presented.

Answer 'yes' if an extensive sensitivity analysis was undertaken that explored all key uncertainties in the economic evaluation. Answer 'partly' if the sensitivity analysis failed to explore some important uncertainties in the economic evaluation. Answer 'no' if the sensitivity analysis was very limited and omitted consideration of a number of important uncertainties, or if the range of values or distributions around parameters considered in the sensitivity analysis were not reported.

2.11 Is there any potential conflict of interest?

The BMJ defines competing interests for its authors as follows: "A competing interest exists when professional judgment concerning a primary interest (such as patients' welfare or the validity of research) may be influenced by a secondary interest (such as financial gain or personal rivalry). It may arise for the authors of a BMJ article when they have a financial interest that may influence, probably without their knowing, their interpretation of their results or those of others."

Whenever a financial conflict of interest is possible, this should be declared.

Answer 'yes' if the authors declare that they have no financial conflicts of interest. Answer 'no' if clear financial conflicts of interest are declared or apparent (for example, from the stated affiliation of the authors). Answer 'unclear' if the article does not indicate whether or not there are financial conflicts of interest.

2.12 Overall assessment

The overall methodological study quality of the economic evaluation should be classified as one of the following:

- **Very serious limitations** The study fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration.
- **Potentially serious limitations** The study fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness.
- **Minor limitations** The study meets all quality criteria, or the study fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness.

References and further reading

Evers, S, Goossens M, de Vet H et al. (2005) Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *International Journal of Technology Assessment in Health Care* 21: 240–5. Available from <http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=292675>

*NICE (2008) Guide to the methods of technology appraisal. Available from www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

NICE (2008) Social value judgements: principles for the development of NICE guidance. Available from www.nice.org.uk/aboutnice/howwework/socialvaluejudgements/socialvaluejudgements.jsp

NICE (2009) Guidelines manual (appendix H). Available from www.nice.org.uk/media/633/71/The_guidelines_manual_2009_-_Appendix_H_Methodology_checklist_-_economic_evaluations.pdf

*Six workshops were held to help NICE explore and capture different perspectives on specific questions as part of the 2007 review of the 'Guide to the methods of technology appraisal'. Documents listed below include briefing papers that were produced to facilitate discussion at each of the workshops and working party meetings:

- costs
- diagnostic technologies
- evidence synthesis (indirect and mixed treatment comparisons)
- identifying subgroups and exploring heterogeneity
- threshold
- exploring uncertainty
- health-related utility measurement.

These documents are available from www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/selectedfurtherreadingguidetothemethodsoftechnologyappraisal.jsp

Philips Z, Ginnelly L, Sculpher M et al. (2004) Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 8 (36). Available from: www.nchta.org/project/1342.asp

Appendix J Process for using review-level material in exceptional circumstances

If it has been agreed with the CPHE project team that review-level material will be quality assessed, data extracted and integrated into the evidence reviews, the following process will be undertaken by the contractors/review team:

- Literature search via agreed sources for review-level material (for example, database and website searching, contacting of experts, reference checking of primary studies and citation searching).
- Title and abstract screening to determine if the review **is likely to be relevant** to the guidance topic under consideration. Please see chapter 5 for further details of the title/abstract screening process.
- Full paper retrieval of any reviews assessed to be potentially relevant at title/abstract screening stage.
- Full review paper screening (using the review screening form on page 219) to determine if the review **is relevant** to the guidance topic. Adaptations to the form to be jointly agreed with the CPHE project team. Please see page 219 for points to consider when adapting and using this form.
- If the review does not meet all of the criteria in the full review screening form, it may still be useful as a source of references (please see chapter 4), but it should not be relied upon on its own to address a research question.
- If the review passes all of the criteria in the full **review screening form** the following should be undertaken:
 - Quality assessment using a review quality assessment form. This form is to be developed by the review team and will need adapting to meet the requirements of the specific topic's research question(s) (the final review form needs to be agreed by the CPHE project team prior to use).
 - Data extraction using the review evidence table templates (see appendix K). Any adaptations to the review evidence table template to be jointly agreed with the CPHE project team.

Study identification Include author, title, reference, year of publication				
Programme/intervention topic:		Key question no:		
Checklist completed by:				
SCREENING QUESTIONS				
In a well-conducted systematic review:		In this review this criterion is met: (Circle one option for each question)		
1	Does the review address an appropriate and clearly-focused question that is relevant to one or more of the guidance topic's key research question/s?	Yes	No	Unclear
2	Does the review include the types of study/ies relevant to the key research question/s?	Yes	No	Unclear
3	Is the literature search sufficiently rigorous to identify all the relevant studies?	Yes	No	Unclear
4	Is the study quality of included studies appropriately assessed and reported?	Yes	No	Unclear
5	Is an adequate description of the analytical methodology used included, and are the methods used appropriate to the question?	Yes	No	Unclear

Notes on the use of the review screening form

This form is intended for use with review-level material including: systematic reviews, meta-analyses, and literature reviews (see appendix D for definitions of these terms).

This form aims to consider the suitability of the review to answer a guidance topic's research question(s). There are two aspects to this assessment:

- Are the questions addressed by the review (in terms of the populations, interventions, comparisons and outcomes considered) appropriate to answer the research question(s) addressed by the guidance?
- Is the methodology employed by the review sufficiently robust to permit a valid conclusion to be reached?

For each questions the reviewer should use one of the following to indicate whether it has been addressed in the review.

- Yes
- No
- Unclear (indicates that this aspect of the review process was ignored, or is not described in the report).

Below are notes which should help the reviewer answer each of the screening questions in the form.

Screening questions

1 Does the review address an appropriate and clearly-focused question that is relevant to the guidance topic's research question/s?

If the research question/s addressed by the review is not clearly stated it will be difficult to determine whether the review is adequate to answer the question addressed by the guidance topic. If the question is not clear, it is unlikely to be a good review as it is difficult to be systematic in addressing an unclear question. The review should give a clear description of the population considered; the interventions included; comparators; and outcomes evaluated. Inclusion and exclusion criteria should be clearly described. Outcomes considered should be clearly described within the methodology, including a precise definition and how validated. The appropriateness of the question addressed in the review for answering the research question/s considered within the guidance can be determined by comparing these components. If the review does not consider all of the outcomes that are judged to be important to the evidence review, the outcome data presented may be able to be used, although the individual studies may also need to be reviewed in order to obtain other outcome data.

2 Does the review include the types of study/ies relevant to the key research question/s?

You should be clear about the characteristics of studies that you consider will adequately address your guidance question/s. This may relate to minimum design or quality characteristics, for example, randomised trials only. Reviews should report the types of studies they sought, including any inclusion/exclusion criteria used, and you can use this to quickly assess the review's suitability for your purposes.

3 Is the literature search sufficiently rigorous to identify all the relevant studies?

Systematic and rigorous searches to identify as much relevant data from as many sources as possible can help to minimise publication biases. Exact search terms depend on the question,

but there may be core databases that should be searched for every question, these databases are likely to be different for each guidance topic. Hand-searching of key journals may suggest a good quality review, and good quality reviews should examine reference lists of retrieved studies for further references. If the methods used to locate studies are not clearly reported, it will be difficult to determine whether the review is likely to have missed important, relevant studies. Ideally, the search strategy used should be reported in sufficient detail so that the process could be replicated.

Any restrictions applied to the search (for example, language, or year of publication) should also be reported. You should consider how these might influence the findings of the review.

Advice from an information scientist working on the guidance may be useful to decide whether any important search terms have been omitted.

If the search is judged to be inadequate to identify all relevant studies, it may be possible to expand the search by including additional databases or extra search terms within the search strategy, or by updating the search to identify more recently published studies. Any additional studies identified by this expanded search should then follow the assessment process for primary studies.

4 Is the study quality appropriately assessed and reported?

The inclusion of low quality studies within a review can result in biased estimates of effect. A well-conducted systematic review should have used appropriate and clear criteria to assess whether individual studies had been appropriately designed and conducted before deciding whether to include or exclude them. These criteria should be clearly described and should be reported for each included study. The other quality appraisal checklists in the appendices can be used as a guide to the types of quality criteria that should be considered.

If there is no indication of such an assessment, the review is unlikely to be reliable enough to be used in formulating recommendations. In such circumstances the relevant primary studies from the review should be obtained and quality appraised as part of the evidence review.

5 Is an adequate description of the analytical methodology used included, and are the methods used appropriate to the question?

In common with primary research, the approach used to analyse the data should be described and justified where appropriate. This may include the choice of statistic used to analyse the outcome data, meta-analytic techniques, and approaches to dealing with heterogeneity including the specification of any subgroup analyses and sensitivity analyses.

Possible issues to consider when using review-level material in exceptional circumstances and/or developing screening tools

Year of publication

Typically the CPHE advocate including primary studies from a specified date onwards. However, a review published since that date may include older primary studies. A decision will need to be made as to how to process such reviews. An example of an approach that has been used is to accept the review if >80% of the primary studies covered by the review meet the specified topic inclusion criteria. An alternative approach could be to specify a different year of publication for review-level material compared to year of publication for primary studies.

Inclusion/exclusion criteria

Reviews may well cover criteria that accord both with the guidance topic/scope inclusion criteria and its exclusion criteria. For example, the guidance topic referral and scope are explicit that the inclusion criteria are males and females under 16 years of age and will only focus on non-treatment-related outcomes. However, a review paper's criteria may cover adolescents aged up to 25 years and include treatment as a component of the interventions delivered. In such situations a decision will need to be made as to how to process such review papers. Is the review included but it is clearly documented that it also includes studies which cover one or more of the topic/scope's exclusion criteria and therefore any findings associated with this review need to be treated with caution? Or is the review automatically excluded if the data relevant to the guidance topic is not sufficiently disaggregated? Or is the review only used as a source to identify additional primary studies which meet the specified inclusion criteria?

Double counting of studies between reviews and primary studies

If review-level material and primary studies are quality assessed and data extracted it is likely that a number of the primary studies included in the evidence review will also be reported in the review-level material. In such circumstances it is important that a clear description of the number of overlapping studies and an outline of any limitations associated with the overlap should be provided (for example, the potential for over-emphasising study findings through 'double counting', or the possibility that different conclusions or interpretations of the primary study's findings can be made).

Double counting of studies across review papers

Where review-level material has been quality assessed and data extracted there may be a significant overlap between the primary studies included across different review-level material. In

such circumstances it is important that a clear description of the number of overlapping studies and an outline of any limitations associated with the overlap should be provided (for example, where different reviews which include in the main the same primary studies, reaching different conclusions about the evidence).

Whatever approaches are undertaken to address the above, these need to be clearly described within the methods section of the evidence review and any limitations associated with the selected approach clearly documented.

Appendix K Evidence tables

Evidence table for quantitative studies ²³					
Study details	Population and setting	Method of allocation to intervention/control	Outcomes ²⁴ and methods of analysis	Results ²⁵	Notes
<p>Authors:</p> <p>Year:</p> <p>Citation:</p> <p>Aim of study:</p> <p>Study design:</p>	<p>Source population/s: Where available describe the following details: country (developed or non-developed, public or private healthcare system etc.); setting (primary schools, community centres etc.); location (urban, rural), and population demographics²⁶ (age, sex, sexual orientation, disability, ethnicity, religion, place of residence, occupation, education, socioeconomic position and social capital).</p>	<p>Method of allocation: Describe how the selected individuals/clusters were allocated to receive either intervention or control. How was confounding minimised?</p> <p>Intervention/s description: Describe intervention in detail including:</p> <ul style="list-style-type: none"> ● what delivered ● by whom ● when/where ● how often ● how long for etc. 	<p>Primary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p>	<p>Primary outcomes:</p> <p>Secondary outcomes:</p>	<p>Limitations identified by author:</p> <p>Limitations identified by review team:</p>

²³ Please complete for all headings and note where data is 'Not reported' or 'Not applicable'.

²⁴ Only report the outcomes that are relevant to the primary and secondary outcomes for the guidance topic under consideration. (NB the study may cover additional outcomes that need not be included here).

²⁵ Only report the results that are relevant for the guidance topic under consideration.

²⁶ Demographic criteria as outlined by the PROGRESS-Plus categorisation (Kavanagh J, Oliver S, Lorenc T (2008) Reflections on developing and using PROGRESS-Plus. Equity Update 2: 1–3. Available from http://equity.cochrane.org/Files/Equity_Update_Vol2_Issue1.pdf [accessed 13 February 2009].

Evidence table for quantitative studies continued

Study details	Population and setting	Method of allocation to intervention/control	Outcomes ²⁴ and methods of analysis	Results ²⁵	Notes
<p>Quality score²⁷: (++, + or –)</p> <p>External validity score: (++, + or –)</p>	<p>Eligible population: Describe how individuals or clusters were recruited (e.g. media advertisement, birth register, class list, area). Was the eligible population representative of the source population?</p> <p>Selected population: Describe the inclusion criteria used to select participants from the eligible population. What % of selected individuals/clusters agreed to participate? Indicate potential sources of bias.</p> <p>Excluded population/s: (as above)</p> <p>Setting: (provide details on country and other setting features e.g. workplace, schools, urban or rural)</p>	<p>Control/comparison/s description: (as above)</p> <p>Sample sizes: Total n = Intervention n = Control n =</p> <p>Baseline comparisons: Indicate if there were any baseline differences between groups in important confounders.</p> <p>Study sufficiently powered? Indicate if a power calculation was presented and provide details. Indicate whether the study was likely to be sufficiently powered</p>	<p>Secondary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p> <p>Follow-up periods:</p> <p>Method of analysis: Indicate if ITT or completer analysis was used. Were adjustments made for any baseline differences in important confounders?</p>	<p>Attrition details: Indicate the number lost to follow-up and whether the proportion lost to follow-up differed by group (i.e. intervention v control)</p>	<p>Evidence gaps and/or recommendations for future research:</p> <p>Source of funding: For example, government (NHS), voluntary/charity, pharmaceutical company and the role of funding organisations</p>

²⁷ The internal validity score of a study may vary depending on the reliability and validity of the outcome measures of interest.

Study details	Research parameters	Population and sample selection	Outcomes and methods of analysis Results	Notes
<p>Authors:</p> <p>Year:</p> <p>Citation:</p> <p>Quality score: (++, + or -)</p>	<p>What was/were the research questions:</p> <p>What theoretical approach (e.g. grounded theory, IPA) does the study take (if specified):</p> <p>How were the data collected:</p> <ul style="list-style-type: none"> ● What method (s): ● By whom: ● What setting(s): ● When: 	<p>What population were the sample recruited from:</p> <p>How were they recruited:</p> <p>How many participants were recruited:</p> <p>Were there specific exclusion criteria:</p> <p>Were there specific inclusion criteria:</p>	<p>Brief description of method and process of analysis:</p> <p>Key themes (with illustrative quotes if available) relevant to this review:</p>	<p>Limitations identified by author:</p> <p>Limitations identified by review team:</p> <p>Evidence gaps and/or recommendations for future research:</p> <p>Source of funding:</p>

Evidence table for economic evaluation studies²⁸

Study details	Population and setting	Intervention/comparator	Outcomes ²⁹ and methods of analysis	Results ³⁰	Notes
<p>Authors:</p> <p>Year:</p> <p>Aim of study:</p> <p>Type of economic analysis:</p> <p>Economic perspective:</p> <p>Quality score³¹: (++, +, -)</p> <p>Applicability³²:</p>	<p>Source population/s: Describe country details (including developed or non-developed, public or private healthcare system), setting (primary schools, community centres etc.), location (urban, rural), and population demographics (age, sex, ethnicity and other socioeconomic variables where available).</p> <p>Setting: (provide details on country and other setting features e.g. workplace, schools, urban or rural).</p> <p>Data sources: whether through primary research, published studies, meta-analyses, published sources, or decision-analytic techniques</p>	<p>Intervention/s description: Describe intervention in detail including:</p> <ul style="list-style-type: none"> ● what delivered ● by whom ● when/where ● how often ● how long for etc. <p>Comparator/control/s description: (as above)</p> <p>Sample sizes: Total n = Intervention n = Control n =</p>	<p>Primary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p> <p>Secondary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p> <p>Time horizon:</p> <p>Discount rates:</p> <ul style="list-style-type: none"> ● Benefits ● Costs <p>Modelling method:</p>	<p>Primary analysis:</p> <ul style="list-style-type: none"> ● Benefits ● Costs ● ICERs <p>Secondary analysis:</p>	<p>Limitations identified by author:</p> <p>Limitations identified by review team:</p> <p>Evidence gaps and/or recommendations for future research:</p> <p>Source of funding: For example, government (NHS), voluntary/charity, pharmaceutical company and the role of funding organisations</p>

²⁸ Please complete for all headings and note where data is 'Not reported' or 'Not applicable'.

²⁹ Only report the outcomes that are relevant to the primary and secondary outcomes for the guidance topic under consideration. (NB the study may cover additional outcomes that need not be included here).

³⁰ Only report the results that are relevant for the guidance topic under consideration.

³¹ Taken from economic evaluation checklist (appendix I).

³² Taken from economic evaluation checklist (appendix I).

Evidence table for review-level material³³

Review details	Review search parameters	Review population and setting	Intervention/s	Outcomes ³⁴ and method of analysis	Results ³⁵	Notes
<p>Authors:</p> <p>Year:</p> <p>Citation:</p> <p>Aim of review:</p> <p>Review design:</p> <p>Quality score: (++, + or -)</p>	<p>Databases and websites searched:</p> <p>Other search methods undertaken (e.g. reference checking):</p> <p>Years searched:</p> <p>Study type inclusion criteria:</p> <p>Study type exclusion criteria:</p>	<p>Included population/s: Where available describe the following population demographics³⁶ (age, sex, sexual orientation, disability, ethnicity, religion, place of residence, occupation, education, socioeconomic position and social capital variables).</p>	<p>Intervention/s description: Describe intervention/s in detail including:</p> <ul style="list-style-type: none"> ● what delivered ● by whom ● when/where ● how often ● how long for etc. <p>Control/comparison/s description: (as above)</p>	<p>Primary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p> <p>Secondary outcomes: Include details of all relevant outcome measures and whether measures are objective or subjective or otherwise validated.</p>	<p>Primary outcomes:</p> <p>Secondary outcomes:</p> <p>Attrition details: Indicate the number lost to follow-up and whether the proportion lost to follow-up differed by group (i.e. intervention v control)</p>	<p>Limitations identified by author:</p> <p>Limitations identified by review team:</p> <p>Evidence gaps and/or recommendations for future research:</p>

³³ Please complete for all headings and note where data is 'Not reported' or 'Not applicable'.

³⁴ Only report outcomes that are relevant to the primary and secondary outcomes for the guidance topic under consideration (NB the review may cover additional outcomes that need not be included here).

³⁵ Only report the results that are relevant for the guidance topic under consideration.

³⁶ Demographic criteria as outlined by the PROGRESS-Plus categorisation. Kavanagh et al. (2008).

Review details	Review search parameters	Review population and setting	Intervention/s	Outcomes ³⁴ and method of analysis	Results ³⁵	Notes
	<p>Number of studies included:</p> <p>Method of synthesis: for example meta-analysis</p>	<p>Excluded population/s: (as above)</p> <p>Setting of included studies: Describe country details (including developed or non-developed, public or private health care system), setting features (e.g. workplace, schools) and location (e.g. urban or rural).</p> <p>External validity score: (++, + or -)</p>		<p>Follow-up periods:</p> <p>Methods of analysis: Indicate how the studies were combined and analysed</p>		<p>Source of funding: For example, government (NHS), voluntary/charity, pharmaceutical company and the role of funding organisations</p>

Appendix L NICE review format

While there is no strict guidance for the way an evidence review is structured, it is important that it sets out as clearly as possible the information that PHIAc/the PDG will need to use to inform its deliberations and recommendations.

The exact structure of the review should be agreed with the CPHE project team on a review by review basis, however, in general we would expect a review to report the following:

Executive summary

- Brief summary of the aims and objectives, methods, main findings and conclusions.
- It should include all of the evidence statements and the related references.

Contents/structure for main report

1. Introduction

- Context in which the review is set, this may include:
 - reference to the scope
 - epidemiological background
 - policy context
 - organisational context
 - theoretical perspectives
 - summary of effectiveness review.

All to be supported by current literature.

- Aims and objectives of the review.
- Research questions.
- Operational definitions.
- Identification of possible equality and equity issues.

- Review team:
 - expertise (both in reviewing and subject area) and perspective brought to the review, for example:
 - ◆ researcher
 - ◆ professional/end user of guidance – clinician/practitioner from health/social/local authority/private sector
 - ◆ target population – general public/patient, carer
 - roles in the review process
 - conflicts of interest.

2. Methodology

Identification of evidence – for example, databases, websites, search strategies, hand-searching, contacts with experts in the field, author contacting.

- Inclusion/exclusion criteria for review – type of studies, years, country, population, implementation process, moderation process.
- Flow chart of number of studies identified from different sources and numbers excluded at different stages of process and reasons for exclusion.
- Quality appraisal processes including consistency checking within and between appraisers, moderation at data extraction and analysis stages.
- Software used for screening and coding of studies, data extraction, analysis and synthesis, managing the bibliography.
- Criteria for appraising for applicability. Sample characteristics, context, conceptual and theoretical focus.
- Methods of synthesis and data presentation.

3. Findings

- Overview of the studies for each research question, such as, sub-question, population and outcome.
- Narrative summary and evidence statements for each question, such as, sub-question, population, outcome:
 - quality, quantity and consistency of evidence
 - applicability of the evidence.
- Meta-analyses, if applicable.

4. Discussion

- Findings into context.
- Implications of findings.
- Limitations of the evidence, gaps.
- Limitations of the review and potential impact on findings.

5. Conclusion and recommendations

6. Appendices

- Sample search strategies.
- Bibliography of included studies.
- Bibliography of excluded studies with reasons for each study.
- Evidence tables.
- Examples of methodology checklists used.

**National Institute for
Health and Clinical Excellence**

MidCity Place
71 High Holborn
London
WC1V 6NA

www.nice.org.uk

N1894 0.6k April 2009

ISBN 1-84629-986-1