NICE National Institute for Health and Care Excellence

Interim methods guide for developing service guidance 2014

NICE process and methods
Published: 31 October 2014

www.nice.org.uk/process/pmg8

Contents

This corporate should be read in conjunction with PMG20.

1 Introduction

This guide should be read alongside Developing NICE guidelines: the manual.

The Centre for Clinical Practice at NICE is responsible for developing guidance that is solely or mostly focused on the organisation and delivery of healthcare services ('service guidance'). The guidelines manual provides limited information about methods for developing evidence-informed recommendations in this area, so specific methods for developing service guidance are needed.

NICE clinical guidelines deal mainly with aspects of the process of care, and in particular the interventions that should be delivered. Some clinical guidelines have also considered the questions of by whom, where and when interventions should be delivered. What service guidance attempts to do is link these issues with the broader health service – in particular, the interaction between structures and processes. For example, to deliver effective care it is necessary to ensure that there is enough appropriate equipment to deliver the required service.

Although the term 'service guidance' in the context of clinical guidelines has no clear or agreed definition, a working definition is that it comprises recommendations on what resources need to be available, how services should be organised and configured, and the processes that need to be followed to ensure the efficient provision of healthcare interventions of proven clinical and cost effectiveness.

NICE guidance on service configuration has been based on a set of core principles: multidisciplinary teams make better decisions than individuals; the configuration of services should optimise a clinician's ability to specialise by providing sufficient volume of procedures. These principles have guided the reorganisation of cancer and stroke care in England. Future service guidance will in part evaluate the transferability of these principles to other clinical areas. As with other NICE guidance, NICE service guidance will be developed with public involvement as an integral part of the process.

The purpose of this methods guide is to provide information additional to that in The guidelines manual to guide developers on how to approach developing service guidance,

and also to inform stakeholders about the steps that NICE will take in developing this guidance.

To ensure consistency with the purpose of NICE and its other guidance programmes, the essential criteria for NICE service guidance are that it is:

- designed to promote good health and prevent ill health
- produced by the people affected by our work, including health and social care professionals, patients and the public
- based on the best available evidence
- transparent in its development, consistent, reliable and based on a rigorous development process
- good value for money, weighing up the cost and benefits of the service.

It is likely that a variety of methods will need to be used to develop service guidance based on different types of questions, and this methods guide outlines possible approaches rather than being prescriptive. Developers should plan the development of service guidance in collaboration with NICE to ensure that there is sufficient time to identify and review the evidence base, and to develop and test the assumptions underpinning the conceptual model.

2 Approaches to developing service guidance

There are 3 main routes that NICE service guidance will be commissioned:

- A specific referral for the development of service guidance without any corresponding clinical guideline development.
- Service guidance that is developed simultaneously with 1 or more relevant clinical guidelines.
- A clinical guideline that includes questions and recommendations about service delivery.

These 3 approaches will affect mainly the scoping process and the composition of the Committee. Some other aspects of the guidance development process might also be affected (for example, different scheduling for submission for pre-consultation checks and for consultation).

For some service delivery questions the current guideline methods maybe appropriate, and so the methods set out in The guidelines manual should be followed.

3 The Committee

The particular Committee composition and the best time for recruiting the Committee for service guidance will be determined by the topic, so no specific guidance is possible. Consideration should be given to recruiting Committee members early in the process in order to support the development of the scope and early conceptual model development. Standard recruitment methods should be followed to ensure that the process is fair and transparent (see section 3.2 of 'The guidelines manual'), and attendees at the stakeholder scoping workshop should be encouraged to apply.

The Committee should include representatives from:

- all clinical specialities involved in the delivery of care
- providers and commissioners of services
- the main regional areas of England
- national clinical organisations
- patients and carers and their organisations
- methodologists such as health economists, statisticians and clinical epidemiologists (especially if they have topic-specific knowledge).

In view of the breadth of clinical areas that may be covered and the various levels of NHS management, it will be necessary to have wider representation on the Committee than is the case for clinical guidelines. To ensure that all relevant groups are all represented, a larger Committee than for standard clinical guidelines may be required. Alternatively, to maintain good group dynamics, additional groups could be represented by co-opted expert advisers (see section 3.1.7.1 of 'The guidelines manual'). The approach used will vary by topic and should be decided in consultation with NICE. The quorum of the Committee will be 50% of appointed members, as specified in The guidelines manual.

Committee meetings will follow the procedures outlined in section 3.8 of 'The guidelines manual'.

4 Developing the decision problem (scoping)

The general principles outlined in chapter 2 of 'The guidelines manual' should be applied when developing the scope.

Assessment of the effectiveness and cost-effectiveness of service delivery interventions will consider the differences in outcomes and cost between a current model of service delivery (the intervention) with alternative current models of service delivery, new models of service delivery, and/or aspirational models of service delivery (the comparators). Ultimately the aim of a service delivery intervention is to maximise health gain, while taking into account equity considerations. Health gain may be measured directly or indirectly, through intermediate or process outcomes, often classed as performance measures. Performance measures can be categorised as follows:

- Registers (lists of included 'cases').
- Care processes indirectly linked to outcomes (for example, blood pressure measurement).
- Care processes directly linked to outcomes (for example, proportion of people with hypertension on therapy to lower blood pressure).
- Intermediate outcomes (for example, proportion of people with hypertension whose blood pressure is within target range).
- Outcomes (for example, subsequent cardiovascular event).

However, there are additional considerations for service guidance – in particular, the context in which it is being developed. The information in the following sections should be considered in addition to the information identified during standard scoping searches for effectiveness and cost-effectiveness evidence.

4.1 Policy context and constraints

Service guidance is likely to have considerable overlap with policy considerations, and a review of existing and planned policies is needed. There should be discussions with

relevant policy/domain leads at NHS England, through NICE, to identify any existing policy initiatives. This may also include a review of existing government policy and other guidance, for example from the Royal Colleges.

4.2 Understanding the current service and variations

Information on current service configuration and provision is needed to inform the development of service guidance and to identify important variations in the provision and quality of services. This can usually be obtained from published sources such as reports and audits. Sources that could be considered include:

- National data sets that can identify factors such as demand, potential variation in practice, current resource use and staffing.
- National organisations.
- Royal Colleges and other professional bodies.
- NICE's QIPP (Quality, Innovation, Productivity and Prevention) collection and local practice collection.
- Specific audit data may be identifiable as useful at this point and the relevant contacts should be approached.
- Patient safety organisations, such as the Patient Safety Division of NHS England.

See appendix 1 for a list of possible sources.

Any relevant recommendations from published NICE guidance should be identified. These may include service recommendations, such as when interventions and services should be provided. NICE guidance will also include recommendations on interventions that the service(s) under consideration will be providing.

It is important that contacts within relevant national organisations, such as NHS England and the Health and Social Care Information Centre, are identified and engaged with as early as possible, because getting access to data sources may be time consuming.

4.3 Clinical and NHS engagement

There will need to be engagement with the service commissioning community to put the information gathered into clinical context. Stakeholders may also be aware of other useful data sources.

The process of identifying key areas for improvement and guidance should be informed by a combination of stakeholder input and the earlier intelligence gathering, especially if there are policy targets. For example, key areas could include addressing variation in practice or ensuring timely access to treatments.

A meeting with stakeholders could be arranged with the aims of producing an initial list of potential areas to be covered and/or developing potential hypotheses about the service. The feedback obtained could be combined with information already gathered to develop a list of priorities and initial searches to be conducted.

The areas identified should be assessed for the type of evidence that would be needed in order to develop recommendations and whether this is likely to be available: for example, a systematic review of when an intervention should be delivered, or a qualitative review to inform recommendations about behaviour change. These can then be confirmed with focused scoping searches if needed.

A key objective of this process is to develop a list of key stakeholder organisations who would be willing to provide information to support guidance development. This group of stakeholders can then be approached throughout development and asked to provide particular types of information as needed.

4.4 Involving patients, service users and carers

The views of patients, service users and carers should be actively sought. This is particularly useful for identifying the key outcomes of the service that are important to them and the relative importance of outcomes. Patient and carer organisations may also have valuable information about patients' experiences of services.

4.5 Approach to scoping

Because a number of topics will cut across clinical areas, it may be difficult to identify key

areas for the guidance to focus on. Therefore various approaches to scoping may be used. In particular, consideration should be given to recruiting not only a Committee chair and topic adviser to help in developing the scope, but also additional members of the scoping group who will become Committee members. When several relevant clinical guidelines are being developed simultaneously with service guidance, there should be representation from each guideline scoping group (or Committee if established) on the service guidance scoping group.

Various methods can be used to run the stakeholder scoping workshop. It is likely that more stakeholders will need to be engaged, so the workshop may have more participants than for a clinical guideline. The workshop could be split over 2 days; for example:

- day 1: identifying the issues and developing a problem-oriented conceptual model (see section 4)
- day 2: identifying which topics should be considered for a full evidence review and which areas could be based on expert opinion alone.

However, when service delivery forms part of a clinical guideline, the stakeholder scoping workshop for the clinical aspects, as described in stage 5 in section 2.3 of 'The guidelines manual', could include service delivery considerations. In this case, specific time should be set aside to discuss the service delivery sections and identify the key issues to be addressed in the guideline.

4.6 The scoping search

A scoping search is undertaken by the Developer and is important to identify:

- related guidance from NICE and other accredited developers
- policy and legislation
- key systematic reviews
- economic evaluations
- information on current practice, including costs and resource use and variations
- types of interventions that may be appropriate

- statistics (for example, on service configuration or staffing levels)
- information on the views and experiences of people using services, their family members or carers, or the public.

The search should not aim to be exhaustive. It should be based on the need to inform the development of the draft scope and the issues to be discussed at the scoping workshop. See appendix 1 for a list of suggested sources.

5 Developing a conceptual model

A conceptual model is defined by Tappenden as 'the abstraction and representation of complex phenomena of interest in some readily expressible form, such that the individual stakeholders' understanding of the parts of the system, and the mathematical representation of that system, can be shared, questioned, tested and ultimately agreed'[1]. Two different, but interlinked, conceptual model forms can be used: problem-oriented and design-oriented.

It is recommended that each review question should have a linked conceptual model. The 'problem-oriented conceptual' model is a simplified, diagrammatical representation of the care/service pathway that describes the resources, processes and interactions in the delivery of healthcare interventions. The main aims of using a problem-oriented conceptual model are to assist in the understanding of a service in order to identify areas for improvement and to describe existing or planned services to allow a shared understanding of them.

The use of a problem-oriented conceptual model is a method for exploring the interrelationship between processes and structures, and to ensure that key areas are approached in the most logical and efficient way. It also should help focus questions and reviews to address key areas, as well as providing a useful tool to assess how the various discrete questions are interlinked and how much of the service will be covered. The problem-oriented conceptual model should not be limited by what is feasible. This problem-oriented conceptual model can be developed as part of the scoping process (for example, at the stakeholder scoping workshop.

The problem-oriented conceptual model should be able to contextualise and describe the service in terms of the following areas:

- who is using the service
- interventions being delivered
- current service models being used
- regional and/or national variations
- key decision makers

- key outcomes for the service
- assumed strengths of the service
- assumed weaknesses of the service
- data identification
- potential trade-offs between options such as effectiveness, volume and impact on travelling times for patients
- waiting list issues.

In addition, it may also be helpful to consider the disease process being addressed by the service under consideration as part of the problem-oriented conceptual model.

The problem-oriented conceptual model links to a 'design-oriented conceptual model' that is used as the quantitative basis to inform the structure, assumptions and data needs of the computer models ('implementation models') to be used to assess effectiveness and cost-effectiveness. The design-oriented conceptual model is an explicit simplification and abstraction of the problem-oriented conceptual model, mediated by what is feasible and by the availability of evidence and data. This design-oriented conceptual model can also be used to help structure the review questions and to specify precisely the data and evidence that will be needed by the implementation model to simulate the service decision problem, and so generate appropriate effectiveness and cost-effectiveness outcomes to inform recommendations.

[1] Tappenden P. Conceptual modelling for health economic model development. ScHARR Discussion Paper (number 12/05) 2012, University of Sheffield.

6 Developing review questions and planning the evidence review

The scope should identify key areas that the guidance will cover. There are various types of review question that may be considered for service guidance; for example, these may cover:

- The content, configuration or integration of services, including the allocation of:
  - medical equipment or tools
  - staff, such as:
    - skills, mix and experience of staff
    - training requirements of staff
    - staffing levels (numbers and staff mix)
- access to services for patients, including:
  - the availability of services
  - the uptake of services
- timing and delivery of services, including:
  - diagnosis
  - treatment
  - transfer and referral
  - waiting times
- location of services, in terms of:
  - setting for delivery
  - economies of scales
  - geographic variation

- feasibility, with regard to:
  - resource constraints (including capacity, queues and waiting lists)
  - policy constraints.

The questions will compare possible service configurations, which may be existing variations to current services (national and international variations) or a proposed service configuration, with a current service configuration with respect to effectiveness and cost-effectiveness.

Key outcomes of service delivery questions are likely to include measures of:

- service effectiveness:
  - health outcomes, including health-related quality of life
  - process outcomes (both directly and indirectly linked to outcomes)
  - compliance rates of staff
  - system failures
- service experience:
  - patient experience
  - family or carer experience
  - staff experience
- service resource use:
  - staff
  - equipment
  - time
  - costs
- service efficiency/optimisation:
  - cost effectiveness (cost–utility analysis)

  - cost consequence
  - cost saving
  - cost minimisations
- service equity (including health and geographical inequalities).

A key difference for service guidance compared with clinical guidelines is that, to adequately address the question, it is necessary to explore the underlying health and/or service concern first, and then assess the effectiveness of the various health service interventions in addressing this underlying issue. This requires an iterative approach to developing the review questions. The first step is to develop questions to explore the underlying problem, followed by developing questions around potential solutions and service models.

These types of review questions will often require the consideration of supplementary methodological approaches to identifying, assessing, synthesising and interpreting the evidence to those normally used.

Evidence reviews will be iterative, with new searches and/or analysis being planned depending on the outcome of the initial reviews. For example, a search for studies exploring the effectiveness of a particular intervention may not produce any results. The next step would be to consider whether to search for evidence for a similar condition or another healthcare system. Alternatively, primary data may need to be identified or requested to inform recommendations. The Committee should be consulted on the suitability of different types of evidence for developing recommendations.

7 Identifying the evidence

7.1 Searching for evidence

Evidence in this context refers not only to studies that directly inform the review question, but also to primary data that might inform parameters identified in the design-oriented conceptual model. A number of evidence and data sources might have been identified during scoping, but additional evidence and data should still be systematically searched for, including grey literature as needed.

When a systematic review of published literature is assessed as being appropriate, the methods outlined in chapter 5 of 'The guidelines manual' should be used to search for and identify evidence and data.

If undertaking a systematic search, the search methods should balance precision and sensitivity. The aim is to identify the best available evidence without producing an unmanageable volume of results. A search protocol should be developed prior to undertaking the search. The protocol will indicate the sources to be searched and the rationale for searching, plans to use any supplementary search techniques and any limits to be applied to the search.

It is important to ensure adequate coverage of the relevant literature and to search a range of sources, to minimise bias. However, there should be a clear rationale for the inclusion of additional sources, with only those likely to yield results being prioritised. Appendix 1 can provide a useful starting point for identifying potential sources.

Consideration may be given to identifying evidence about the efficiency and productivity of healthcare delivery. These include studies that examine the numbers of patients treated or identified for a given constraint (such as financial).

These additional requirements are likely to result in significantly more potential studies or data sources being identified than for clinical guidelines. Therefore consideration should be given to methods of identifying and including relevant evidence and data, such as iterative searching techniques.

As noted above, supplementary search techniques, such as citation searching on known

relevant studies, could be considered to make searching more efficient. The types of studies that will be needed to address the identified areas will be varied, and a search of the grey literature is likely to be needed. Some of the evidence may be in the form of modelling studies that may not be published within health-related journals, but rather in those related to operation research, statistical and mathematical methods. Because these are less likely to be indexed in healthcare databases such as Medline, consideration should be given to sources likely to retrieve this type of evidence. Science Citation Index is an example of a useful source.

Some of the evidence may be in the form of modelling studies that may not be published within health-related journals, but rather in those related to operation research, statistical and mathematical methods, which would not be indexed within Medline or Embase, but would be found in databases such as the Science Citation Index.

Searching for relevant observational data can be very time consuming, so the design-oriented conceptual model should be used to evaluate the usefulness of doing this. For example, registry data can be a potential source of estimates of treatment effects. But if the Committee cannot estimate the extent or direction of any inherent biases, and has no way of placing limits on these estimates, the value of these studies is greatly diminished and the usefulness of searching for this evidence is low.

Before requesting evidence or data from stakeholders, there needs to be careful consideration of how such evidence or data will inform either the working of the design-oriented conceptual model or how the model should be structured. For example, financial reports from health organisations may be used to quantify the financial trade-offs, but consideration needs to be given to whether single reports can provide sufficient detail.

7.2 Calls for evidence or data from stakeholders

For some questions, there may be good reason to believe that relevant and useful information exists outside of literature databases or validated national data sources. Examples include ongoing research in a field, if a service is relatively new, and studies that have been published only as abstracts (see section 6.1 of 'The guidelines manual').

Typically, the method for requesting information from stakeholders is through a 'call for evidence'. This is to allow all registered stakeholders to have an equal opportunity to provide relevant information they may have access to. It is anticipated that developing

service guidance will rely significantly on information from NHS, patient and other relevant clinical organisations, which is why a list should be compiled during scoping of stakeholders that might be able to provide such information. The type of evidence that might be requested includes:

- health needs assessments
- protocols
- local pilot studies
- business cases
- financial reports.
- analyses of primary data.

7.3 Registries and audits

Data from registries and audits may be used to inform both estimates of effectiveness and any modelling. To obtain such data, it may be necessary to negotiate access with the organisations and individuals that hold the data, or to ask them to provide a summary for inclusion in the guidance. Any processes used for accessing data will need to be reported in the protocol and in the guidance. Given the difficulties that organisations may have in extracting audit data, such requests should be focused and targeted: for example, identifying a specific audit and requesting results from the previous 3 years.

7.4 Economic studies

Searching for existing economic evidence relating to services may differ from economic searches for questions about interventions, diagnosis or prognosis. When undertaking a systematic search for economic evidence, the principles and guidance outlined in section 5 of 'The guidelines manual' should be followed. Examining the economic evaluations may also differ for service guidance. Existing economic evaluations are likely to focus on local or regional populations rather than national averages. Studies may not measure commonly used health outcomes such as mortality or measures to calculate the quality-adjusted life-year (QALY). Evidence is likely to include measures of resource use, and process measures such as length of lists, number of falls or throughput of patients. Searching for, selecting and reviewing such information will differ for service guidance.

The following inclusion criteria should be taken into account when deciding whether to include economic studies identified in the search; final decisions will depend on the service being assessed:

- An appropriate date range, because older studies may reflect outdated practices.
- The country or setting, because studies conducted in other countries might not be relevant to the UK. In some cases it may be appropriate to limit consideration to the UK or countries with similar healthcare systems. Studies on local settings should also be included.
- The type of economic evaluation. This may include cost-utility, cost-benefit, cost-effectiveness, cost-minimisation, or cost-consequence analyses, depending on what the Committee deems to be the most relevant and likely outcomes for the question. In the absence of comparative studies, non-comparative costing studies (such as econometric, efficiency, simulation, microcosting and resource use, and time-series) can be included. On occasion, the published economic evidence is extremely sparse. In such cases, search strategies may be broadened. The decision to do this is taken by the guideline Developer in consultation with NICE staff with responsibility for guideline quality assurance, when appropriate, with the Committee or its Chair.

8 Selecting relevant evidence and data

8.1 Identifying appropriate studies and data

Searches should have been developed using the conceptual models to help identify the most appropriate types of evidence and data to include in the guidance and to inform analyses. Selection of evidence and data should also be informed by planned subsequent synthesis, including any network meta-analysis. Otherwise the methods in chapter 6 of 'The guidelines manual' should be followed.

8.2 Extrapolation

In line with the methods outlined in section 9.1 of 'The guidelines manual', if very little, or no, good-quality evidence is identified, the Committee may wish to extrapolate from high-quality evidence in a related area. A number of service models have been developed from evidence for different clinical conditions. For example, recommendations in the Cancer service guidance[2] series were informed from evidence for a number of different cancer types and generalised across a number of others. The use of extrapolation must be considered carefully by the Committee, with explicit consideration of the features of the condition or interventions that allow extrapolation. This also applies when extrapolating findings from evidence in different healthcare settings. The Committee should comment on similarities in case mix, staffing, facilities and processes.

[2] To see a list of NICE cancer service guidance, type 'CSG' into the filter box on the linked 'Guidance list' page.

9 Evidence synthesis

9.1 Estimates of the relative effectiveness of service delivery interventions

It is helpful to distinguish between 2 general types of service delivery questions. One type concerns different pathways of care, different service configurations, interventions to be managed by different types of staff, whether a 'care team' approach is needed, and so on. These are questions for which trial evidence could in principle be found. For these kinds of questions, standard approaches to evidence identification and synthesis (for example, those described in 'The guidelines manual' and by the NICE Decision Support Unit) could, in principle, be used. However, for service guidance it is unlikely that 1 type of study or piece of evidence will be sufficient to inform recommendations. Therefore non-standard approaches to evidence synthesis will also need to be considered to enable the Committee to develop recommendations. Two specific problems that will often need to be addressed are:

- uncertainty about the quality and relevance of existing evidence on clinical outcomes
- the need to consider evidence on intermediate or surrogate outcomes, such as uptake of services or compliance, rather than (or in addition to) evidence on clinical outcomes.

A second type of service delivery issue relates to questions about the feasibility of providing access to services and procedures, or making them available within a certain timeframe, rather than whether the services or procedures are effective. In these questions, estimates of the effect of providing the service, compared with not providing it, are needed for decision making, whether based on cost-effectiveness analysis or on other criteria.

It should be emphasised that some service delivery guidance may present a combination of both access and availability issues as well as standard effectiveness issues.

Guidance on how to approach both kinds of problem, as well as on using consensus techniques when estimates based on published data cannot be obtained, is given in the following sections.

Finding studies that provide unbiased estimates of the effectiveness of service interventions is often difficult, for the following reasons:

- Service delivery interventions are inherently 'variable'. Even with a standard protocol, the precise way in which they are implemented at different sites or by different people is necessarily situation- and/or individual-dependent. This could be manifested by centre effects in multi-centre trials.
- The relative benefit of a new intervention over 'standard' or pre-existing care is likely to depend on the 'intensity' of the current care. For example, the beneficial effect of a new patient reminder system on the uptake of screening for breast cancer depends on what the current arrangements are, and on current uptake. For example, the effect of introducing a reminder system in the USA, where there is no systematic screening programme, will be quite different from the effect of adding the reminder system to existing infrastructure in the UK. In other words, results from studies carried out within other healthcare systems might not be easily generalised to the UK.

In these circumstances a standard systematic review is likely to identify a range of studies on interventions that are similar to the interventions being considered, but not necessarily the same. In this case, the Committee will need to consider carefully fidelity and applicability issues, and ensure these are accounted for in the 'Linking evidence to recommendations' section of the guidance.

In most cases, the expert opinion of the Committee will be used to explore and estimate any impacts on the confidence in the results of such evidence, but quantitative methods can be used. If quantitative methods are to be used, the NICE Clinical Guidelines Technical Support Unit should be contacted for advice on using such quantitative methods and on which types of evidence could be searched for.

9.2 Evidence on uptake and compliance outcomes

In some service delivery evaluations, measures of service uptake, patient satisfaction or compliance of health service staff are recorded, rather than data on clinical outcomes for patients. This is typically the case, for example, when the intervention is directed at changing staff behaviour or patient referral routes.

Such evidence can be used when analysing the effectiveness or cost effectiveness of a service delivery intervention, but only if there is also an estimate available – from whatever

source – of the underlying clinical effect of the procedure or treatment. It is then possible to combine estimates of the efficacy or effectiveness of the clinical intervention with estimates of the effectiveness of the service delivery intervention in ensuring that the clinical intervention is implemented. It is possible to combine evidence from trials reporting process outcomes alone, trials reporting clinical outcomes alone, and trials reporting both.

The NICE Clinical Guidelines Technical Support Unit can be consulted for advice on how the 2 kinds of evidence can be combined within a single modelling framework.

9.3 Estimates of relative effectiveness for questions about access and availability

For questions about access and availability, there is a particular difficulty in deriving an estimate of relative effectiveness, over and above those described in the previous section. This would be the case, for example, where a procedure such as endoscopy for upper gastrointestinal bleeding is indicated. The question is not about whether endoscopy should be done, but whether or not the procedure can be safely delayed (for example, at night or at weekends) in patients whose symptoms suggest they are at lower risk.

Studies based on individual patient 'audit' data that relate outcomes to treatment parameters while controlling for patient characteristics are difficult to interpret. This is because patients in whom the treatment was withheld or delayed are always likely to be those who were considered to be at lower risk.

It is likely that better estimates of the effectiveness of such interventions can be derived from nationally collected data in which between-unit variation in outcomes, or variation between different time periods, can be related to the local policies and practices (for example staffing levels) in operation at the time. For example, mortality rates within 1 or 2 days of hospital admission could be compared between weekends and weekdays, and hospitals where weekend cover was the same as weekday cover could also be compared with those where it is not. There are a number of examples where comparisons of this type have been published, for example by Dr Foster. Although these surveys avoid the problems of individual audit data, they are still observational and the use of aggregated data introduces further potential biases. The design of the data collection, and the analysis and interpretation of the data obtained, requires major input from clinical epidemiologists, expert clinicians, methodologists, operational research experts and people with relevant operational experience in the NHS.

A service delivery issue that is quite often examined in this way is the relationship between performance indicators and 'volume' (that is, number of cases seen per year). Such data are also used to establish 'institutional rankings'. Data of this type tend to show considerable overdispersion: in other words, there is far more variation between units than would be expected by chance. To determine whether individual units are performing at a level that requires some intervention, control charts can be used. There are also methods and processes for interpreting the relationships between performance and volume and the need to take into account general between-unit variation when trying to infer causal effects.

9.4 Formal consensus techniques

Formal consensus techniques are increasingly being used in developing clinical guidelines because of their explicit structure, process and output. A number of well-established formal consensus methods have been used in the health field; the 3 main approaches are the Delphi method, the nominal group technique and the consensus development conference. The Health Technology Assessment report 'Consensus development methods, and their use in clinical guideline development' (Murphy et al. 1998[3]) provides a useful summary of the strengths and limitations of each technique.

Since the concepts of appropriate and necessary care are fundamental to an efficient and equitable healthcare delivery system, the RAND Appropriateness Method (RAM) is often described as the preferred approach for developing service guidance. One of the advantages of RAM is that the process of developing consensus statements can be presented as a service pathway. In addition, the interactions and discussions during development can be structured to fit the current 'Evidence to recommendations' framework that is used in NICE clinical guidelines.

Developers should consult NICE if formal consensus methods are to be used. If formal consensus is used, the methods used should be clearly described in the guidance document.

[3] Murphy MK, Black NA, Lamping DL et al. (1998) Consensus development methods, and their use in clinical guideline development. Health Technology Assessment 2 (3).

10 Modelling and health economics considerations

The key challenge of service guidance is linking process developments to a health benefit. This obviously poses a challenge when conducting health economic analyses for service guidance, but it will also be difficult with respect to the quality and lack of evidence of effectiveness for service configurations, and so modelling will usually be needed to generate the health benefits used within the health economic analyses using scenario analyses. In addition, given the considerable resource and health impact of any service recommendations, there must be an explicit consideration of the opportunity cost of implementing a recommendation, preferably analytically or qualitatively.

Developing design-orientated conceptual models linked to each review question should help the health economist to decide what key information is needed for developing effectiveness and cost-effectiveness analyses. It is anticipated that developed effectiveness and economic models will relate to several review questions, so that almost all recommendations are underpinned by some form of modelled analysis.

The choice of appropriate model structure is a key aspect of the design-orientated conceptual model. When designing the implementation model, Brennan's taxonomy of model structures[4] should be considered for guidance on which of types of models may be appropriate to the service delivery decision problem.

Even if a fully modelled analysis is not possible, there is value in the process of development, as it will help to structure Committee discussions. For example, a model might be able to demonstrate how a service change will impact on demand for a downstream service or intervention.

For any cost-effectiveness analysis, the reference case remains that outlined in table 7.1 of 'The guidelines manual'; a cost–utility analysis should be aspired to, producing an incremental cost-effectiveness ratio (ICER). This allows the Committee to use the same decision rules as those outlined in chapter 7 of 'The guidelines manual'. Other methods of economic analyses such as cost–consequence, cost-effectiveness, cost–benefit, cost minimisation and microcosting analyses may be used if these can provide the Committee with sufficient information on which to base recommendations. For example, if a service is associated with better health outcomes and fewer adverse effects, then a cost-

minimisation analysis may be justifiable. However, given the complexity of services, a series of simple analyses may be misleading, by not accounting for interactions.

One main area where assessing cost effectiveness will differ from standard NICE methods is that any analysis will need to consider resource constraints. These might be monetary, but might also be resources such as staff, beds, equipment and so on. However, affordability should not be the sole driver for service recommendations, and there needs to be explicit consideration of the impact on quality of care of any proposed changes.

10.1 Health service related operational research

For the areas that are considered in service guidance, operational research methods are likely to be the most appropriate way to assess cost effectiveness. It is not appropriate for this guide to discuss in detail all available methods. Operational research in cost-effectiveness analysis of service delivery interventions, a report for NICE by the Clinical Guidelines Technical Support Unit, outlines the approaches that are available, in what circumstances they can be used and the data and resources required.

Experts in these operational research methods should be consulted for advice on the suitability of methods for certain types of service delivery question. This should be done when developing review protocols to identify whether operational research methods are likely to be useful. The use of these methods should be discussed and agreed with NICE, since additional resources and time may be required.

All methods used and results obtained should be described clearly in the full guidance, and should follow the principles outlined for statistical and health economics analyses in chapter 7 of 'The guidelines manual'.

10.2 Local considerations

Cost-effectiveness analyses will need to account for local factors, such as the expected number of procedures and the availability of staff and equipment at different times of the day, week and year. Models will need to incorporate the fact that each local provider may be starting from a different baseline of identified factors (for example, the number of consultants available at weekends). It is therefore important that these factors are identified and considered explicitly by the Committee. Results obtained from the analysis should include both the national average and identified local scenarios to ensure that

recommendations are robust to local variation.

10.3 Service failures

Service designs under consideration might result in occasional service failure – that is, where the service does not operate as planned. For example, a service for treating people with myocardial infarction may differ at the weekend compared with on weekdays – that is, the number of places where people can be treated might be reduced at weekends as a result of staffing considerations. Therefore more people will need to travel by ambulance and the journey time will also be longer. Given the limited number of ambulances, a small proportion may be delayed, resulting in consequences in terms of costs and QALYs. Such possible service failure events should be taken into account in effectiveness and economic modelling. This effectively means that analyses should incorporate the 'side effects' of service designs.

10.4 Perspective on costs

The perspective on costs should remain that of the NHS and personal social services (PSS); however, service recommendations are likely to have additional costs. These include implementation costs and costs to other government budgets, such as social care. Implementation costs should be included in economic analyses in a sensitivity analysis. Costs to other government budgets can be presented in a separate analysis to the base case.

10.5 Service demand

Introducing a new service or increasing capacity will often result in an increase in demand. This could mean that a service does not achieve the predicted effectiveness because there is more demand than was planned for. This should be explicitly addressed either in the analysis or in considerations.

10.6 Equity considerations

Basing economic evaluations on local circumstances may result in recommendations for a different provision of services in different areas. This could be perceived as being inequitable. The Committee should give careful consideration to equity concerns and this

should be explicitly addressed in the guideline, in particular consideration should be made of how recommendations that may lead to inequities of service provision would be mitigated.

10.7 Health benefits

The QALY remains the most suitable measure for assessing the impact of services, since it can incorporate benefits from extension to life and experience of care. In addition, it can explicitly include the trade-offs of benefits and adverse events.

If linking to a QALY gain is not possible, links to a clinically relevant or a related outcome should be considered. Consideration should be given to optimising outputs for the lowest resource use. Any surrogate outcome such as a process outcome (for example, bed days) needs to be justified explicitly in terms of linking it to a clinical outcome (either directly or indirectly), similar to when a clinical surrogate outcome is used instead of an outcome that is relevant to a patient. However, when QALYs are not used, issues such as trade-offs need to be considered explicitly.

[4] Brennan A, Chick SE, Davies R (2006) A taxonomy of model structures for economic evaluation of health technologies. Health Economics 15: 1295–310.

11 Reviewing the evidence

The quality of individual pieces of evidence or studies can be assessed using checklists. Appendix 2 gives examples of checklists for assessing the quality of different types of non-randomised and economic studies.

The criteria that are likely to be the most important indicators of quality should be agreed in advance and be tailored to the question being addressed. These criteria will be useful in guiding decisions about the overall quality of individual studies, and when summarising and presenting the body of evidence. Expert input may be needed to identify the most appropriate quality criteria.

When assessing the quality of evidence being used within effectiveness and economic analyses, the evidence should be assessed for potential bias (size and direction), and by conducting a sensitivity analysis to explore the impact of this where possible. Any analysis should be conducted in accordance with chapter 7 of 'The guidelines manual'.

The assessment of quality of evidence should be presented clearly in the full guidance. Assessment of quality will require input from experts to ensure that the study is assessed properly; these could include statisticians, epidemiologists and clinicians.

Appropriate methods should be used to present evidence. A GRADE-like approach (see appendix K of 'The guidelines manual') can be used if it is considered to add value to the presentation of the evidence.

Along with the presentation of the evidence, there should be an accompanying evidence statement to summarise the findings from the key features of the evidence on clinical and cost effectiveness.

12 Developing recommendations

The principles outlined in chapter 9 of 'The guidelines manual' should be used when developing service guidance recommendations. The link between the evidence on clinical and cost effectiveness and the recommendations should be clearly presented in the full guidance, and the Committee should assess the strength of recommendations (see section 9.2 of 'The guidelines manual'). The Committee will, however, have to consider additional factors when developing recommendations for service guidance.

Recommendations on service guidance could have a potentially significant impact on all aspects of care and resources. It is therefore important that the recommendations are underpinned by a clear and comprehensive review of the evidence, results from validated models, and documented and explicit considerations of the Committee to justify what may be a highly disruptive and expensive reorganisation of the service with potentially irrecoverable costs. This includes ensuring that recommendations will remain relevant into the future (usually for at least 5 years).

Although the aim of service guidance is to reduce variability in outcomes of and access to services, there may be circumstances when the Committee wants to ensure that local customisation is possible. Therefore the Committee may recommend a list of preferred options so that decision makers can choose appropriate models for their local circumstances. However, there needs to be an overarching statement about the objective of the recommendations.

Potential areas that the Committee could address or refer to include the following:

- The resources required for delivering clinically and cost effective services. These could include minimum specifications that a service must be able to deliver.
- Where patients should be treated or referred for treatment.
- How NHS staff should be organised.
- Designating staff who should be responsible for the provision and delivery of services.
- How services should interact and the sharing of information between services.

The Committee should identify potential areas for disinvestment. Development and

drafting of these recommendations should follow the same process, including consideration of the evidence and model development, as that followed for recommendations that lead to an increase in resources. However, the Committee should consider the potential impact of the withdrawal of a service on the health of the population, taking into account equalities considerations (for example, whether withdrawal might affect some groups more than others), as well as any other potential negative effects, and so the Committee should also consider how these effects may be mitigated.

NHS England will be the primary audience for service guidance. The Committee should not attempt to ensure that recommendations are directly relevant to the rest of the UK, as this will be the responsibility of the devolved administrations. Recommendations should meet the needs of both patients and decision makers. Therefore consideration should be given to aiming recommendations at specific audiences, such as commissioners and/or providers. This might include highlighting who is responsible for implementing recommendations.

The advice on the wording of recommendations in clinical guidelines (see section 9.2 of 'The guidelines manual') will not always be appropriate for service guidance. This is because recommendations about service guidance are not referring to a clinician's individual decision to offer or consider using a treatment. Therefore alternative wording should be explored by the Committee and discussed with NICE – in particular, the NICE editor.

In general, the wording of recommendations should be agreed by the Committee, and should:

- focus on the services that needs to be delivered, and where appropriate, who needs to provide these
- include what readers need to know
- reflect the strength of the recommendation
- emphasise the involvement of people using services, carers where appropriate, and the public in making decisions
- use plain English where possible and avoid vague language and jargon
- use language and terms that NICE has agreed to ensure consistency across guidelines and other products

- follow NICE's standard advice on recommendations about waiting times and ineffective interventions.

The recommendations should (wherever possible and if not obvious from the context of the guideline) clearly detail the intended audience for the recommendation (who is responsible for implementing it), the intended population, the setting (if relevant), what specifically should be done, and, where relevant, what the timeframe is for doing it.

12.1 Economic evidence and recommendations

When a cost per QALY can be obtained, the principles outlined in chapter 7 of 'The guidelines manual' should be followed. Additional considerations include the potential effects of identified factors that impact on the ICER, including the possible implications of legal and equity issues. If these factors have not been incorporated into the ICER, they need to be considered by the Committee.

The approach adopted by the Committee for interpreting cost-effectiveness or cost–benefit results, including the relative weight given to certain outcomes, should be clearly described in the evidence to recommendations sections of the guidance. The Committee should aim to maximise efficiency while maintaining the quality of services, and the considerations used should be outlined in the evidence to recommendations sections.

When there is no evidence of differences between different options, a cost-minimisation approach can be used. However, the Committee must be convinced that the 2 options do not differ for all relevant outcomes.

12.2 Research recommendations

Research recommendations will be vitally important to improve the available evidence base for health services. These should be formulated using the principles outlined in section 9.5 of 'The guidelines manual'.

13 Writing the guidance and next steps

We anticipate that the methods and process outlined in chapters 9 to 11 of 'The guidelines manual' for writing, consulting on and finalising clinical guidelines will also be appropriate for service guidance. Any variations should be agreed with NICE.

Given the potential impact of service guidance, the NICE implementation team should be contacted as soon as possible to provide support for implementing the guidance (see chapter 12 of 'The guidelines manual').

We anticipate that the processes for reviewing and updating service guidance, and for correcting errors in published guidance, will be the same as those for clinical guidelines, as outlined in chapters 13 and 14 of 'The guidelines manual'

14 Further reading

The following is a list of useful references on the methods outlined in this section.

Ades AE, Lu G, Higgins JPT (2005) The interpretation of random effects meta-analysis in decision models. Medical Decision Making 25: 646–54

Dias S, Welton NJ, Marinho V et al. (2010) Estimation and adjustment of bias in randomised evidence using mixed treatment comparison meta-analysis. Journal of the Royal Statistical Society Series A 173: 613–29

Higgins JPT, Thompson SG, Spiegelhalter DA (2009) A re-evaluation of random effects meta-analysis. Journal of the Royal Statistical Society Series A 172: 137–59

Kotiadis K, Tako AA, Vasilakis C. (2014) A participative and facilitative conceptual modelling framework for discrete event simulation studies in healthcare. JORS 65, 197–213

Robinson S. (2008) Conceptual modelling for simulation Part I: definition and requirements. JORS 59:278-290

Salanti G, Dias S, Welton NJ et al. (2010) Evaluating novel agent effects in multiple-treatments meta-regression. Statistics in Medicine 29: 2369–83

Spiegelhalter D (2002) Funnel plots for institutional comparison. Quality and Safety in Health Care 11: 390–2

Spiegelhalter DJ (2005) Funnel plots for comparing institutional performance. Statistics in Medicine. 24: 1185–202

Spiegelhalter DJ (2005) Handling over-dispersion of performance indicators. Quality and Safety in Health Care 14: 347–51

Spiegelhalter, DJ, Abrams, KR, Myles, JP (2004) Bayesian approaches to clinical trials and health-care evaluation. Chichester: Wiley

Turner RM, Spiegelhalter D, Smith GCS et al. (2009) Bias modelling in evidence synthesis. Journal of the Royal Statistical Society Series A 172: 21–47

Welton NJ, Ades AE, Caldwell DM et al. (2008) Research prioritisation based on expected value of partial perfect information: a case-study on interventions to increase uptake of breast cancer screening. Journal of the Royal Statistical Society Series A 171: 807–41 (with discussion)

Welton NJ, Ades AE, Carlin JB, et al (2009) A bias model for the combination of low and high quality evidence: empirically based priors. Journal of the Royal Statistical Society Series A 172: 119—36

Appendix 1 Examples of evidence sources

Databases

- Allied and Complementary Medicine (AMED)
- Applied Social Science Index and Abstracts (ASSIA)
- British Education Index (BEI)
- British Official Publications Current Awareness Service (BOPCAS)
- Campbell Database of Systematic Reviews
- Cochrane Central Register of Controlled Trials (CENTRAL)
- Cochrane Database of Systematic Reviews (CDSR)
- Cumulated Index to Nursing and Allied Health Literature (CINAHL)
- Database of Abstracts of Reviews of Effectiveness (DARE)
- Educational Information Resources Center (ERIC)
- Embase
- EPPI-Centre list of systematic reviews
- Health Business Elite
- Health Management Information Consortium (HMIC)
- Health Technology Assessment (HTA) database
- Joanna Briggs Institute Library of Systematic reviews
- MEDLINE/MEDLINE in Process
- National Institute for Health Research (NIHR) HTA Programme
- Physiotherapy Evidence Database (PEDro)

- PsycINFO
- Science Citation Index
- Social Care Online
- Social Policy and Practice
- Social Science Citation Index
- Social Services Abstracts
- Sociological Abstracts
- SportDiscus
- Transport
- UK Database of Uncertainties about the Effects of Treatments (DUETS)

Economics

- NHS Economic Evaluation Database (NHS EED)
- Health Economic Evaluations Database (HEED)
- Econpapers
- Econlit
- Econometics papers
- CEA Registry

Grey literature

- Evidence Search – NICE Evidence Services
- OpenGrey

Experiences of patients, service users and carers,

or the target population

- HealthTalkOnline
- YouthHealthTalk
- Service user organisation websites

Ongoing trials

- ClinicalTrials.gov
- Current Controlled Trials
- United Kingdom Clinical Research Network's (UKCRN) Portfolio Database

Conference abstracts

- Embase
- British Library Inside Conferences (BLIC)
- Google Scholar
- Conference websites

Understanding current service and variations

- Hospital episode statistics (HES)
- Health and social care information centre
- Care Quality Commission (CQC)
- National Audit Office
- Audit Commission
- National Clinical Audit and Patient Outcomes Programme
- Guidelines and Audit Implementation Network (GAIN)

- King's Fund
- Nuffield Trust
- NHS Institute for Innovation and Improvement
- Royal colleges and other professional bodies
- NICE's QIPP (Quality, Innovation, Productivity and Prevention) collection and Shared Learning Database
- National Patient Safety Agency

Appendix 2 Checklists

Please note that these checklists have not been previously used in guideline development.

1.1 Checklist: cost–benefit analysis (CBA) studies

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1.1 Is there a well-defined question?		
1.2 Is there a comprehensive description of alternatives?		
1.3 Was one of the alternatives designated as the comparator against which the intervention was evaluated?		
1.4 Is the perspective stated?		
1.5 Are all important and relevant costs and outcomes for each alternative identified? Check to see if the study is of money-costs and 'benefits' which are savings of future money-costs.		
1.6 Has the effectiveness of the intervention been established?		
1.7 Are costs and outcomes measured accurately?		
1.8 Are costs and outcomes valued credibly?		

1.9 Have all important and relevant costs and outcomes for each alternative been quantified in money terms? If not, state which items were not quantified, and the likely extent of their importance in terms of influencing the benefit/cost ratio.		
1.10 Are costs and outcomes adjusted for differential timing?		
1.11 Has at least 1 of net present value, benefit/cost ratio and payback period been estimated?		
1.12 Were any assumptions of materiality made?		
1.13 Were all assumptions reasonable in the circumstances in which they were made, and were they justified?		
1.14 Were sensitivity analyses conducted to investigate uncertainty in estimates of cost or benefits?		
1.15 To what extent do study results include all issues of concern to users?		
1.16 Are the results generalisable to the setting of interest in the review? • Country differences. • Question of interest differs from the CBA question being reviewed.		
1.17 Overall assessment: Minor limitations/Potentially serious limitations/Very serious limitations		
Other comments:		

Notes on Checklist: cost–benefit analysis (CBA) studies

Definition:

Cost–benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.

Source:

Adapted from Methods for the development of NICE public health guidance (third edition, 2012).

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.

Overall assessment:

The overall methodological study quality of the economic evaluation should be classified as 1 of the following:

- **Minor limitations** The study meets all quality criteria, or fails to meet 1 or more quality criteria but this is unlikely to change the conclusions about cost benefit.
- **Potentially serious limitations** The study fails to meet 1 or more quality criteria, and this could change the conclusions about cost benefit.
- **Very serious limitations** The study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost benefit. Such studies should usually be excluded from further consideration.

1.2 Checklist: cost–consequence analysis (CCA) studies

Study identification *Include author, title, reference, year of publication*		
Guidance topic:		**Question no:**
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1.1 Is there a well-defined question?		
1.2 Is there a comprehensive description of alternatives?		
1.3 Was one of the alternatives designated as the comparator against which the intervention was evaluated?		
1.4 Is the perspective stated?		
1.5 Who determined the set of outcomes that were collected to act as consequences?		
1.6 Are all important and relevant costs and outcomes for each alternative identified?		
1.7 Has effectiveness been established?		
1.8 Are costs and outcomes measured accurately?		
1.9 Are costs and outcomes valued credibly?		
1.10 Have all important and relevant costs and outcomes for each alternative been quantified? • If not, state which items were not quantified. • Were they still used in the CCA and how were they used?		
1.11 Are all costs and outcomes adjusted for differential timing?		

1.12 Were any assumptions of materiality made to restrict the number of consequences considered?		
1.13 Was any analysis of correlation between consequences carried out to help control for double counting?		
1.14 Was there any indication of the relative importance of the different consequences by a suggested weighting of them?		
1.15 Were there any theoretical relationships between consequences that could have been taken into account in determining weights?		
1.16 Were the consequences considered one by one to see if a decision could be made based on a single consequence?		
1.17 Were the consequences considered in subgroups of all the consequences in the analysis to see if a decision could be made based on a particular subgroup?		
1.18 Was an MCDA (multiple criteria decision analysis) or other published method of aggregation of consequences attempted?		
1.19 Were all assumptions reasonable in the circumstances in which they were made, and were they justified?		
1.20 Were sensitivity analyses conducted to investigate uncertainty in estimates of cost or benefits?		
1.21 How far do study results include all issues of concern to users?		
1.22 Are the results generalisable to the setting of interest in the review? • Country differences. • Question of interest differs from the CCA question being reviewed.		
1.23 Overall assessment: Minor limitations/Potentially serious limitations/Very serious limitations		
Other comments:		

Notes on Checklist: cost–consequence analysis (CCA) studies

Definition:

Cost–consequence analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with those of a suitable alternative. Unlike cost–benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.

Sources:

Methods for the development of NICE public health guidance (third edition, 2012).

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.

Overall assessment:

The overall methodological study quality of the economic evaluation should be classified as 1 of the following:

- **Minor limitations** The study meets all quality criteria, or fails to meet 1 or more quality criteria but this is unlikely to change the conclusions about cost consequences.
- **Potentially serious limitations** The study fails to meet 1 or more quality criteria, and this could change the conclusions about cost consequences.
- **Very serious limitations** The study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost consequences. Such studies should usually be excluded from further consideration.

1.3 Checklist: audit

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1 Objectives		
1.1 Are the objectives of the audit clearly stated?		
1.2 The clinical audit topic reflects a local service, speciality or national priority which merits evaluation and where care could be improved or refined through clinical audit		
2 Design		
2.1 The clinical audit measures against standards		
2.2 The clinical audit standards are based upon the best available evidence		
2.3 The clinical audit standards are referenced to their source		
2.4 The clinical audit standards are expressed in a form that enables measurement		

2.5 The patient group to whom the clinical audit standards apply is clearly defined		
2.6 The clinical audit standards take full account of patient priorities and patient-defined outcomes		
2.7 The timetable for the clinical audit is described, including timescales for completion and re-audit where necessary		
3 Methodology		
3.1 The methodology and data collection process is described in detail		
3.2 Systematic consideration is given to ethics, data confidentiality and consent issues, and Caldicott principles are applied		
3.3 The methods used in the audit are recorded so that re-audit can be undertaken later in the clinical audit cycle		
3.4 If a sample of the population was audited, the method for sampling is that which is best suited to measuring performance against the standards and is as scientifically reliable as possible		
3.5 Is the sample size sufficient to generate meaningful results?		
3.6 When necessary, the sample allows for adjustment for case mix		
3.7 The clinical audit uses pre-existing data sets where possible		
3.8 The data collection tool(s) and process have been validated		
3.9 The data collection process aims to ensure complete capture of data		
4 Analysis		
4.1 Data are analysed, and feedback of the results is given so that momentum of the clinical audit is maintained in line with the agreed timetable		
4.2 Results of the clinical audit are presented in the most appropriate manner for each potential audience to ensure that the audit results stimulate and support action planning		

4.3 The results are communicated effectively to all key stakeholders, including patients		
5 Sustaining improvement		
5.1 The topic is re-audited to complete the clinical audit cycle if necessary		
5.2 Where recommended action has not been achieved in full, the topic is re-audited at agreed intervals		
5.3 The results of re-audit are recorded and disseminated appropriately, including to patients		

Notes on Checklist: audit

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.

Definition:

Clinical audit is a quality improvement process that seeks to improve patient care and outcomes through systematic review of care against explicit criteria and the implementation of change. Aspects of the structure, process and outcome of care are selected and systematically evaluated against explicit criteria. Where indicated, changes are implemented at an individual, team or service level and further monitoring is used to confirm improvement in healthcare delivery. (This definition appears in Principles for best

practice in clinical audit (2002) and was endorsed by NICE.).

An audit is an examination or review that establishes the extent to which a condition, process or performance conforms to predetermined standards or criteria. Assessment or review of any aspect of healthcare to determine its quality; audits may be carried out on the provision of care, compliance with regulations, community response or completeness of records. (From: Porta M (2008) A dictionary of epidemiology [fifth edition]. Oxford: Oxford University Press.)

Source:

Healthcare Quality Improvement Partnership (HQIP)

1.4 Checklist: surveys

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1 Objectives		
1.1 Are the objectives of the study clearly stated?		
2 Design		
2.1 Is the research design clearly specified and appropriate for the research aims?		
2.2 Is there a clear description of context?		
2.3 If an existing tool was used, are references to the original work provided?		

2.4 If a new tool was used, have its reliability and validity been reported?		
2.5 Is there a clear description of the survey population and the sample frame used to identify this population?		
2.6 Do the authors provide a description of how representative the sample is of the underlying population?		
2.7 Did the subject represent the full spectrum of the population of interest?		
2.8 Is the study large enough to achieve its objectives? Have sample size estimates been performed?		
2.9 Were all subjects accounted for?		
2.10 Were all appropriate outcomes considered?		
2.11 Has ethical approval been obtained if appropriate?		
2.12 What measures were made to contact non-responders?		
2.13 What was the response rate?		
3 Measurement and observation		
3.1 Is it clear what was measured, how it was measured and what the outcomes were?		
3.2 Are the measurements valid?		
3.3 Are the measurements reliable?		
3.4 Are the measurements reproducible?		
4 Presentation of results		
4.1 Are the basic data adequately described?		
4.2 Are the results presented clearly, objectively and in sufficient detail to enable readers to make their own judgement?		
4.3 Are the results internally consistent, i.e. do the numbers add up properly?		
5 Analysis		

5.1 Are the data suitable for analysis?		
5.2 Is there a clear description of the methods of data collection and analysis?		
5.3 Are the methods appropriate for the data?		
5.4 Are any statistics correctly performed and interpreted?		
5.5 Is the method for calculating response rate provided?		
5.6 Are the methods for handling missing data provided?		
5.7 Is information given on how non-respondents differ from respondents?		
6 Discussion		
6.1 Are the results discussed in relation to existing knowledge on the subject and study objectives?		
6.2 Are the limitations of the study (taking into account potential sources of bias) stated?		
6.3 Can the results be generalised?		
6.4 Have attempts been made to establish 'reliability' and 'validity' of analysis (appropriate to methodology)?		
7 Interpretation		
7.1 Are the authors' conclusions justified by the data? Do the researchers display enough data to support their interpretations and conclusions?		

Notes for Checklist: surveys

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.

Definition:

A survey is a data collection tool used to gather information about individuals. Surveys are commonly used in clinical research to collect self-report data from study participants. A survey may focus on factual information about individuals, or it may aim to collect the opinions of the survey takers. A population survey may be conducted by face-to-face inquiry, self-completed questionnaires, telephone, postal service or in some other way. (From: Porta M (2008) A dictionary of epidemiology [fifth edition]. Oxford: Oxford University Press.)

Sources:

- BestBETs critical appraisal worksheet 'Survey (including pre-test probabilities)'
- Personal communication from Dr Susan Kirk (School of Nursing, Midwifery and Social Work, The University of Manchester) and Michelle Maden (Edge Hill University Library and Information Resources Centre).
- Bennett C, Khangura S, Brehaut JC et al. (2011) Reporting guidelines for survey research: an analysis of published guidance and reporting practices. PLoS Medicine 8: e1001069
- Crombie IK (1996) The pocket guide to critical appraisal: a handbook for healthcare professionals. London: BMJ Publishing

1.5 Checklist: studies of national, regional or local reports, assessments or evaluations

Study identification *Include author, title, reference, year of publication*	

Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1. Authority		
1.1 Does the report identify who is responsible for the intellectual content?		
1.2 Are they reputable?		
2 Accuracy		
2.1 Does the item have a clearly stated aim or brief?		
2.2 Does it have a stated methodology?		
2.3 Has it been peer-reviewed?		
2.4 Has it been edited by a reputable authority?		
3 Coverage		
3.1 Are any limits clearly stated?		
4 Objectivity		
4.1 Is the author's standpoint clear?		
4.2 Does the work seem to be balanced in presentation?		
5 Date		
5.1 Does the item have a clearly stated date related to content?		
6 Significance		
6.1 Is the item meaningful?		
6.2 Does it add context?		
6.3 Does it strengthen or refute a current position?		
6.4 Would the research area be lesser without it?		

Other comments:

Notes for Checklist: studies of national, regional or local reports, assessments or evaluations

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.

Definition:

The Fourth International Conference on Grey Literature held in Washington, DC, in October 1999 defined grey literature as: 'that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers'.

Source:

The AACODS checklist (adapted by NICE).

1.6 Checklist: longitudinal studies

Study identification *Include author, title, reference, year of publication*

Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1.1 Are the objectives of the study clearly stated?		
1.2 Was the study ethical?		
2. Sampling		
2.1 Were all members of the cohort entered at the beginning?		
2.2 Did the sampling scheme allow a representative sample?		
3. Participation		
3.1 Was loss to follow-up low – i.e. less than 20%?		
3.2 Was completion rate on individual items of the assessment instrument high?		
4. Measurement		
4.1 Were valid measures of disease (case definition) and risks used?		
4.2 Were the data gathered using the best-accepted techniques? (e.g. trained telephone interviewers or examiners, mail questionnaire)		
4.3 Were the data tested for accuracy and reliability?		
4.4 Are the age/sex distributions similar?		
4.5 Is there evidence of any systematic differences in prevalence or trends in disease between this group and the patients being considered?		
4.6 Is there evidence of any systematic differences in important environmental, behavioural or healthcare access factors between this group and the patients being considered?		

Other comments:

Notes for Checklist: longitudinal studies

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.

Definition:

In a longitudinal study, subjects are followed over time with continuous or repeated monitoring of risk factors or health outcomes, or both. Such investigations vary enormously in their size and complexity. At one extreme a large population may be studied over decades. For example, the longitudinal study of the Office of Population Censuses and Surveys prospectively follows a 1% sample of the British population that was initially identified at the 1971 census. Outcomes such as mortality and incidence of cancer have been related to employment status, housing and other variables measured at successive censuses. At the other extreme, some longitudinal studies follow up relatively small groups for a few days or weeks. Thus, firemen acutely exposed to noxious fumes might be monitored to identify any immediate effects. (From: Epidemiology for the uninitiated [fourth edition]. London: BMJ.)

Source:

Checklist to Assess Evidence of Prevalence and Incidence (Descriptive or Longitudinal Studies): University of Toronto (adapted by NICE).

1.7 Checklist: cross-sectional studies

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
1 Objectives		
1.1 Are the objectives of the study clearly stated?		
2 Design		
2.1 Is the research design clearly specified and appropriate for the research aims?		
2.2 Were the subjects recruited in an acceptable way?		
2.3 Was the sample representative of a defined population?		
3 Measurement and observation		
3.1 Is it clear what was measured, how it was measured and what the outcomes were?		
3.2 Are the measurements valid?		
3.3 Was the setting for data collection justified?		
3.4 Were all important outcomes/results considered?		
4 Analysis		
4.1 Are tables/graphs adequately labelled and understandable?		
4.2 Are the authors' choice and use of statistical methods appropriate, if employed?		
4.3 Is there an in-depth description of the analysis process?		

4.4 Are sufficient data presented to support the findings?		
5 Discussion		
5.1 Are the results discussed in relation to existing knowledge on the subject and study objectives?		
5.2 Can the results be generalised?		

Notes for Checklist: cross-sectional studies

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.

Definition:

A cross-sectional study is a study that examines the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at one particular time. The presence or absence of a disease and the presence or absence of the other variables are determined in each member of the study population or in a representative sample at one particular time. The relationship between a variable and the disease can be examined (1) in terms of the prevalence of the disease in different population subgroups defined according to the presence or absence of the variables and (2) in terms of the presence or absence of the variables in people with the disease compared with those without the disease. Note that disease prevalence rather than incidence is normally recorded in a cross-sectional study. The temporal sequence of

cause and effect cannot necessarily be determined in a cross-sectional study. (Adapted from: Porta M (2008) A dictionary of epidemiology [fifth edition]. Oxford: Oxford University Press.)

Sources:

- Cardiff University
- Wordpress

1.8 Checklist: secondary data studies

Study identification *Include author, title, reference, year of publication*		
Guidance topic:		**Question no:**
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
Screening questions		
1.1 Does the study address a clearly focused issue?		
1.2 Is a good case made for the approach that the authors have taken?		
1.3 Is there a direct comparison (for example, service configurations or models) that provides an additional frame of reference?		
Methods		
1.4 Were those involved in collection of data also involved in delivering a service to the user group?		
1.5 Were the methods used for selecting the users appropriate and clearly described?		
Results		

1.6 Was the data collection instrument/method reliable?		
1.7 What was the response rate and how representative was the sample under study?		
1.8 Are the results complete and have they been analysed in an easily interpretable way?		
1.9 Are any limitations in the methodology (that might have influenced results) identified and discussed?		
1.10 Are the conclusions based on an honest and objective interpretation of the results?		
Interpretation		
1.11 Can the results be applied to other service users?		
Other comments:		

Notes for Checklist: secondary data studies

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.

Definition:

Secondary data are data that have been already collected by and readily available from other sources.

1.1 Does the study address a clearly focused issue?

An issue can be 'focused' in terms of:

- the population (user group) studied
- the intervention (service or facility) provided
- the outcomes (quantifiable or qualitative) measured.

1.2 Is a good case made for the approach that the authors have taken?

Do the authors state how they identified the problem and provide a justification for why they have chosen to examine it? Do they state in what way their chosen methodology is appropriate to the question?

Consider, too, whether the study:

- refers to previous work that has looked at the **same user group**
- refers to previous work that has looked at the **same service or facility**
- utilises **a methodology or data collection instruments** that have been used in previous user studies.

1.3 Is there a direct comparison (for example, service configurations or models) that provides an additional frame of reference?

This may be either **external** or **internal**; for example, contrast with, or similarity to:

- other studies
- other user groups within the study
- the same group at different geographical locations or at a different time period.

1.4 Were those involved in collection of data also involved in delivering a service to the user group?

It may not always be possible to separate researchers from service deliverers, but consider whether the service deliverers' perspective has been acknowledged explicitly

and to what extent the questions in the user study have been generated elsewhere (for example, a previously trialled or validated instrument or from a focus group).

1.5 Were the methods used for selecting the users appropriate and clearly described?

Type of sample: Is it a convenience sample? Were participants self-selecting? Were key informants identified? Is it a randomly selected sample? Is it a comprehensive census or survey?

Size of sample: Has a sample size calculation been undertaken?

Representativeness of sample: Was the planned sample of users representative of all users (actual **and** eligible) who might be included in the study? Do the demographics of the sample (such as age, sex, staff grade, location) accurately reflect the demographics of the total population? Are any interests or motivations behind participation clearly identified? Are non-users included in the sampling frame?

1.6 Was the data collection instrument/method reliable?

If there is a questionnaire, survey form or interview schedule, do the authors include it in their report? Do they refer to where a full copy might be found? Has the data collection instrument been used before? Have the authors adapted an existing questionnaire and, if so, have they used it appropriately?

1.7 What was the response rate and how representative was the sample under study?

Consider not only the actual percentage of responses but also whether any specific subgroups were either over-represented or under-represented. Are reasons for non-response discussed? Have non-users been included in the analysis of responses?

1.8 Are the results complete and have they been analysed in an easily interpretable way?

Consider choices involved in analysis and in presentation. Have all variables identified earlier in the study been analysed? If not, why not?

1.9 Are any limitations in the methodology (that might have influenced results) identified and discussed?

Consider whether the authors give a clear picture of how the study might best be done. Would it be possible for you to replicate the study from the information given? Is there enough detail of any data collection instrument for you to reproduce it?

1.10 Are the conclusions based on an honest and objective interpretation of the results?

Do the authors base their conclusions on findings from their experimental data? Can you be sure that they are not presenting their data merely to substantiate some preconceived ideas?

1.11 Can the results be applied to other service users?

The burden of proof is on you to identify any ways in which your local population might differ from that in the study.

1.9 Checklist: grey literature

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No /Unclear /NA	**Comments**
Authority		

Identifying who is responsible for the intellectual content. **Individual author:** • Associated with a reputable organisation? • Professional qualifications or considerable experience? • Produced/published other work (grey/black) in the field? • Recognised expert, identified in other sources? • Cited by others? (use Google Scholar as a quick check) • Higher degree student under 'expert' supervision? **Organisation or group:** • Is the organisation reputable? (e.g. WHO) • Is the organisation an authority in the field?		
Does the item have a detailed reference list or bibliography?		
Accuracy		
Does the item have a clearly stated aim or brief?		
Does the item meet its aims?		
Does the item have a stated methodology?		
Has the item been peer reviewed?		
Has the item been edited by a reputable authority?		
Is the item supported by authoritative, documented references or credible sources?		
Is the item representative of work in the field?		
If no, is it a valid counterbalance?		
Is any data collection explicit and appropriate for the research?		

If the item is secondary material (e.g. a policy brief of a technical report), does it provide an accurate, unbiased interpretation or analysis of the original document?		
Coverage		
Are any limits to the item clearly stated?		
Objectivity		
Is the author's standpoint clear?		
Does the work seem to be balanced in presentation?		
Date		
Does the item have a clearly stated date related to content?		
If no date is given, but can be accurately ascertained, is there a valid reason for its absence?		
Has key contemporary material been included in the bibliography?		
Significance		
Is the item meaningful (i.e. does it incorporate feasibility, utility and relevance)?		
Does it add context?		
Does it enrich or add something unique to the research?		
Does it strengthen or refute a current position?		
Would the research area be lesser without it?		
Is it integral, representative, typical?		
Does it have impact (in the sense of influencing the work or behaviour of others)?		

Notes for Checklist: grey literature

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.

Definition:

The Fourth International Conference on Grey Literature held in Washington, DC, in October 1999 **defined grey literature** as: 'that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers.' [sic]

Grey literature includes theses or dissertations (reviewed by examiners who are subject specialists); conference papers (often peer-reviewed or presented by those with specialist knowledge) and various types of reports from those working in the field. All of these fall into the '**expert opinion**'.

Sources:

AACODS: archived at the Flinders Academic Commons.

Coverage:

All items have parameters that define their content coverage. These limits might mean that a work refers to a particular population group, or that it excluded certain types of

publication. A report could be designed to answer a particular question, or be based on statistics from a particular survey.

Objectivity:

It is important to identify bias, particularly if it is unstated or unacknowledged.

Date:

For the item to inform your research, it needs to have a date that confirms relevance. No easily discernible date is a strong concern.

Significance:

This is a value judgment of the item, in the context of the relevant research area.

1.10 Checklist: systematic reviews (non-randomised controlled trials)

Study identification *Include author, title, reference, year of publication*		
Guidance topic:	**Question no:**	
Checklist completed by:		
	Yes/ Partly/ No/ Unclear /NA	**Comments**
Reporting of background		

1.1 Reporting of background should include: • definition of problem • hypothesis statement • description of study outcome(s) • type of exposure or intervention used • type of study designs used • study population		
Reporting of search strategy		
1.2 Reporting of search strategy should include: • qualifications of searchers (e.g. librarians and investigators) • search strategy, including time period included in the synthesis and keywords • effort to include all available studies, including contact with authors • databases and registries searched • use of hand searching (e.g. reference lists of obtained articles) • list of citations located and those excluded, including justification • method of addressing articles published in languages other than English • method of handling abstracts and unpublished studies • description of any contact with authors		
Reporting of methods		

1.3 Reporting of methods should include: • description of relevance or appropriateness of studies assembled for assessing the hypotheses to be tested • rationale for the selection and coding of data (e.g. sound clinical principles or convenience) • documentation of how data were classified and coded (e.g. multiple raters, blinding, and inter-rater reliability) • assessment of confounding (e.g. comparability of cases and controls in studies if appropriate) • assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results • assessment of heterogeneity • description of statistical methods (e.g. complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose–response models, or cumulative meta-analysis) in sufficient detail to be replicated • provision of appropriate tables and graphics		
Reporting of results		
1.4 Reporting of results should include: • graphic summarising individual study estimates and overall estimate • table giving descriptive information for each study included • results of sensitivity testing (e.g. subgroup analysis) • indication of statistical uncertainty of findings		
Reporting of discussion		

1.5 Reporting of discussion should include: • quantitative assessment of bias (e.g. publication bias) • justification for exclusion (e.g. exclusion of non-English-language citations) • assessment of quality of included studies		
Reporting of conclusions		
1.6 Reporting of conclusions should include: • consideration of alternative explanations for observed results • generalisation of the conclusions (i.e. appropriate for the data presented and within the domain of the literature review) • recommendations for future research • disclosure of funding source		
Other comments:		

Notes for Checklist: systematic reviews (non-randomised controlled trials)

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.

Definition:

A non-randomised controlled trial is an experimental study in which people are allocated to different interventions using methods that are not random.

A systematic review uses explicit and systematic methods to identify, appraise and summarise the literature according to predetermined criteria. If the methods and criteria used to do this are not described or are not sufficiently detailed, it is not possible to make a thorough evaluation of the quality of the review.

Source:

From: Stroup DF, Berlin JA, Morton SC et al. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283: 2008–12

If this checklist is not considered appropriate, the NICE checklist for systematic reviews and meta-analyses (appendix B of 'The guidelines manual') can be used.

1.11 Checklist: mixed-methods reviews

For a mixed-methods study, use section 1 for appraising the qualitative component, the appropriate section (2, 3 or 4) for the quantitative component, and section 5 for the mixed-methods component.

Study identification *Include author, title, reference, year of publication*	
Guideline topic:	**Question no:**
Checklist completed by:	

	Yes/ Partly/ No /Unclear /NA	Comments
Section 1 – qualitative studies		
1.1. Are the sources of qualitative data (archives, documents, informants, observations) relevant to address the research question? *Consider whether (a) the selection of the participants is clear, and appropriate to collect relevant and rich data; and (b) reasons why certain potential participants chose not to participate are explained*		
1.2. Is the process for analysing qualitative data relevant to address the research question? *Consider whether (a) the method of data collection is clear (in-depth interviews and/or group interviews, and/or observations and/or documentary sources); (b) the form of the data is clear (tape recording, video material, and/or field notes, for instance); (c) changes are explained when methods are altered during the study; and (d) the qualitative data analysis addresses the question*		
1.3. Is appropriate consideration given to how findings relate to the context, such as the setting, in which the data were collected? *Consider whether the study context, and how findings relate to the context or characteristics of the context, are explained (how findings are influenced by or influence the context). 'For example, a researcher wishing to observe care in an acute hospital around the clock may not be able to study more than one hospital. Here, it is essential to take care to describe the context and particulars of the case [the hospital] and to flag up for the reader the similarities and differences between the case and other settings of the same type' (Mays and Pope, 1995[a])*		

1.4. Is appropriate consideration given to how findings relate to researchers' influence; for example, through their interactions with participants? *Consider whether (a) researchers critically explain how findings relate to their perspective, role and interactions with participants (how the research process is influenced by or influences the researcher); (b) the researcher's role is influential at all stages (formulation of a research question, data collection, data analysis and interpretation of findings); and (c) researchers explain their reaction to critical events that occurred during the study*		
Section 2 – quantitative studies (randomised controlled trials)		
2.1. Is there a clear description of the randomisation (or an appropriate sequence generation)? *In a randomised controlled trial, the allocation of a participant (or a data collection unit, e.g. a school) into the intervention or control group is based solely on chance, and researchers describe how the randomisation schedule is generated. A simple statement, such as 'we randomly allocated' or 'using a randomised design' is insufficient.* *Simple randomisation is defined as allocation of participants to groups by chance by following a predetermined plan/sequence. Usually it is achieved by referring to a published list of random numbers, or to a list of random assignments generated by a computer.* *Sequence generation: The rule for allocating interventions to participants must be specified, based on some chance (random) process. Researchers should provide sufficient detail to allow readers' appraisal of whether it produces comparable groups. Examples include blocked randomisation (to ensure particular allocation ratios to the intervention groups), stratified randomisation (randomisation performed separately within strata) and minimisation (to make small groups closely similar with respect to several characteristics).*		

2.2. Is there a clear description of the allocation concealment (or blinding when applicable)? *The allocation concealment protects assignment sequence until allocation. For example, researchers and participants are unaware of the assignment sequence up to the point of allocation; group assignment is concealed in opaque envelopes until allocation.* *The blinding protects assignment sequence after allocation. For example, researchers and/or participants are unaware of the group a participant is allocated to during the course of the study.*		
2.3. Are there complete outcome data (80% or above)? *For example, almost all the participants contributed to almost all measures.*		
2.4. Is there low withdrawal/drop-out (below 20%)? *For example, almost all the participants completed the study*		
Section 3 – quantitative studies (including non-randomised controlled trial, cohort study, case–control study, cross-sectional study)		
3.1. Are participants (organisations) recruited in a way that minimises selection bias? *At the recruitment stage:* • *for cohort studies, consider whether the exposed (or with intervention) and non-exposed (or without intervention) groups are recruited from the same population* • *for case–control studies, consider whether the same inclusion and exclusion criteria were applied to cases and controls, and whether recruitment was done independently of the intervention or exposure status* • *for cross-sectional analytical studies, consider whether the sample is representative of the population.*		

3.2. Are measurements appropriate (clear origin, or validity known, or standard instrument; and absence of contamination between groups when appropriate) regarding the exposure/ intervention and outcomes? *At the data collection stage:* • *consider whether (a) the variables are clearly defined and accurately measured; (b) the measurements are justified and appropriate for answering the research question; and (c) the measurements reflect what they are supposed to measure.* • *for non-randomised controlled trials, the intervention is assigned by researchers, and so consider whether there was absence/presence of a contamination.* • *consider whether the control group may be indirectly exposed to the intervention through family or community relationships.*		
3.3. In the groups being compared (exposed versus non-exposed; with intervention versus without; cases versus controls), are the participants comparable, or do researchers take into account (control for) the difference between these groups? *At the data analysis stage:* • *for cohort, case–control and cross-sectional studies, consider whether (a) the most important factors are taken into account in the analysis; (b) a table lists key demographic information comparing both groups, and there are no obvious dissimilarities between groups that may account for any differences in outcomes, or dissimilarities are taken into account in the analysis.*		
3.4. Are there complete outcome data (80% or above), and, when applicable, an acceptable response rate (60% or above), or an acceptable follow-up rate for cohort studies (depending on the duration of follow-up)?		

Section 4 – quantitative descriptive studies (including incidence or prevalence study without comparison group, case series or case report)		
4.1. Is the sampling strategy relevant to address the quantitative research question (quantitative aspect of the mixed-methods question)? *Consider whether (a) the source of the sample is relevant to the population under study; (b) when appropriate, there is a standard procedure for sampling, and the sample size is justified (using power calculation, for instance)*		
4.2. Is the sample representative of the population understudy? *Consider whether (a) inclusion and exclusion criteria are explained; and (b) reasons why certain eligible individuals chose not to participate are explained*		
4.3. Are measurements appropriate (clear origin, or validity known, or standard instrument)? *Consider whether (a) the variables are clearly defined and accurately measured; (b) measurements are justified and appropriate for answering the research question; and (c) the measurements reflect what they aresupposed to measure*		
4.4. Is there an acceptable response rate (60% or above)? *The response rate is not pertinent for case series and case reports (for example, there is no expectation that a case series would include all patients in a similar situation)*		
Section 5 – mixed methods[1] (including sequential explanatory design, sequential exploratory design, triangulation design and embedded design)		
5.1. Is the mixed-methods research design relevant to address the qualitative and quantitative research questions (or objectives), or the qualitative and quantitative aspects of the mixed-methods question? *For example, the rationale for integrating qualitative and quantitative methods to answer the research question is explained*		

5.2. Is the integration of qualitative and quantitative data (or results) relevant to address the research question? *For example, there is evidence that data gathered by both research methods was brought together to form a complete picture and answer the research question; the authors explain when integration occurred (during the data collection analysis or/ and during the interpretation of qualitative and quantitative results); they explain how integration occurred and who participated in this integration*		
5.3. Is appropriate consideration given to the limitations associated with this integration, such as the divergence of qualitative and quantitative data (or results)?		

[1]**Mixed-methods study designs**

A. Sequential explanatory design: The quantitative component is followed by the qualitative. The purpose is to explain quantitative results using qualitative findings. For example, the quantitative results guide the selection of qualitative data sources and data collection, and the qualitative findings contribute to the interpretation of quantitative results.

B. Sequential exploratory design: The qualitative component is followed by the quantitative. The purpose is to explore, develop and test an instrument (or taxonomy), or a conceptual framework (or theoretical model). For example, the qualitative findings inform the quantitative data collection, and the quantitative results allow a generalisation of the qualitative findings.

C. Triangulation design: The qualitative and quantitative components are concomitant. The purpose is to examine the same phenomenon by interpreting qualitative and quantitative results (bringing data analysis together at the interpretation stage), or by integrating qualitative and quantitative datasets (for example, data on same cases), or by transforming data (for example, quantisation of qualitative data).

D. Embedded design: The qualitative and quantitative components are concomitant. The purpose is to support a qualitative study with a quantitative substudy (measures), or to better understand a specific issue of a quantitative study using a qualitative substudy; quantisation, the efficacy or the implementation of an intervention based on the views of participants.

Key references: Creswell and Plano Clark (2007)[b]; O'Cathain (2010)[c].

[a] Mays N, Pope C (1995) Qualitative research: rigour and qualitative research.BMJ 311: 109–12.

[b] Creswell JW. Plano Clark VL (2007) Designing and conducting mixed methods research. Thousand Oaks, CA: Sage Publications.

[c] O'Cathain A (2010) Assessing the quality of mixed methods research: toward a comprehensive framework. In: Tashakkori A, Teddlie C (editors), Handbook of mixed methods research, 2nd edition, pp. 531–55. Thousand Oaks, CA: Sage Publications.

Notes for Checklist: mixed-methods reviews

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.

Definition:

Mixed-methods reviews evaluate studies that employ qualitative, quantitative and mixed methodology.

Sources:

- Pluye P, Robert E, Cargo M et al. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. Retrieved on 25 July 2012 from http://mixedmethodsappraisaltoolpublic.pbworks.com. Archived at www.webcitation.org/5tTRTc9yJ

- Health Care Practice R&D Unit (HCPRDU) at the University of Salford.

About this manual

This manual describes the methods used in the development of NICE service guidance.

Citing this document: National Institute for Health and Care Excellence (October 2014) Interim methods guide for developing service guidance 2014. Manchester: National Institute for Health and Care Excellence. Available from: www.nice.org.uk

Produced by the National Institute for Health and Care Excellence

First issued October 2014

Copyright

ISBN: 978-1-4731-0808-0