HANDBOOK FOR THE PREPARATION OF EXPLICIT EVIDENCE-BASED CLINICAL PRACTICE GUIDELINES

NEW ZEALAND GUIDELINES GROUP

www.nzgg.org.nz
## CONTENTS

Acknowledgements 3

1 Introduction to Evidence-Based Clinical Practice Guidelines 4
   1.1 Introduction 4
   1.2 What are clinical practice guidelines? 5
   1.3 Do clinical practice guidelines make a difference to the delivery of health care? 6
   1.4 Guideline development in New Zealand 6
   1.5 Principles of guideline development 7
   1.6 Overview of steps in clinical practice guideline development 9

2 Topic Identification and Suitability Screen 11
   2.1 Topic identification 11
   2.2 Suitability screen 11

3 Question Formulation for Clinical Practice Guidelines 15
   3.1 The development of a series of questions 15
   3.2 The 5 parts to well formulated clinical questions 16
   3.3 Examples of well formulated clinical questions 17

4 Internal Data Acquisition 21
   4.1 Steps in data acquisition 21

5 Finding the Evidence - Literature Searching for Guideline Development 23
   5.1 Literature Searches 23
   5.2 Principles of searching OVID databases 26
   5.3 Saving search strategies and results 28
   5.4 Search filters 28
   5.5 PubMed 28
   5.6 Observations of the process 31
   5.7 Example 31
   5.8 Search filters for Medline OVID 34
   5.9 Sources of assistance 35

6 The Role of Qualitative Research in the Development of Clinical Practice Guidelines 38
   6.2 Qualitative research methods 38
   6.3 Differences between qualitative and quantitative research 39
   6.4 Why include qualitative data in evidence based guidelines 39
   6.5 Specific role for using qualitative data in guidelines 40
   6.6 How to evaluate qualitative data 40
   6.7 An example of qualitative research 41

7 Assessing and Applying the Evidence 42
   7.1 Step 1 Critical appraisal 44
   7.2 Step 2 Quality rating and development of evidence tables 50
7.3 Step 3 Developing graded evidence recommendations from the body of evidence.

8 Balance Sheets
8.1 What is a balance sheet?
8.2 What value is a balance sheet?
8.3 Objectives of balance sheets
8.4 Perspective
8.5 Who develops the balance sheet
8.6 Prior questions
8.7 Data sources
8.8 Feasibility of balance sheet development
8.9 Pitfalls
8.10 Process for development of a balance sheet
8.11 Examples of balance sheets

9 Pulling it all together, decision aids
9.1 The recommendations
9.2 Steps in algorithm formulation
9.3 Writing instructions
9.4 Other decision aids

10 Dissemination and implementation of clinical practice guidelines
10.1 What works in implementation of clinical practice guidelines?
10.2 Dissemination and implementation

11 Appraisal and evaluation guidelines
11.1 AGREE evaluation tool
11.2 Evaluation and adaption of other guidelines
11.3 Evaluation of the impact of the guideline

12 Revision and adaptation of clinical practice guidelines
12.1 Revision and updating of guidelines
12.2 Local and national adaptation of guidelines

13 Getting Started
13.1 Convene multidisciplinary guideline development team to oversee the development
13.2 Guideline development team process
13.3 Define the aim and target audience for the guideline
13.4 Involving the stakeholders
13.5 Pre-testing the guideline
13.6 Time line for guideline development

References

APPENDICES
1 Guidelines For the Involvement of Consumers in Guideline Development
2 Generic Appraisal Tool for Epidemiology (GATE)
3 Considered judgement form
4 The AGREE Tool for appraising guidelines
5 Suggested Template for Guiding
6 Commitment to Maori
7 Glossary
Acknowledgements:

This manuscript was produced by the following members of the Effective Practice Institute of the University of Auckland:

Ms Anne Lethaby (Chapter 7), Dr Sue Wells (Chapter 3), Ms Sue Furness (Chapter 5), Ms Judi Strid (Chapter 1, Chapter 10, 11, Appendix 1) Assoc Prof Bruce Arroll (Chapter 7) Assoc Prof Richard Milne (Chapter 8) and Prof Rodney Jackson (Chapter 7, Appendix 2). Dr David Thomas commented on an early draft of Chapter 6.

Assoc Prof Cynthia Farquhar was responsible for developing an early draft of the manuscript and the overall editing of the manuscript. An early draft was circulated to guideline developers in New Zealand, Australia, the United States and the United Kingdom and we are grateful for their feedback.

Mrs Vivienne Topping assisted with the preparation of the manuscript. We are also thankful to the NZGG and in particular Catherine Marshall for providing guidance during the process and Miss Stephanie Dixon who was responsible for the formatting of the handbook.

Sources of information:

This handbook is a compilation of a number of different guideline development processes. The handbooks and manuals of Group Health Cooperative, Puget Sound, Washington, the National Health and Medical Research Council (NHMRC) of Australia and Scottish Intercollegiate Guideline Network (SIGN) have all been invaluable resources. Chapters 2, 3 and 8 have drawn on a previous handbook from GHC. Chapters 7 and 10 have adapted work from NHMRC and research papers from ICSI (Institute for Clinical Systems Improvement), Minnesota, and USPSFC (US Preventive Services Task Force). Chapter 9 is adapted from SIGN and GHC. Chapter 11 was developed from the AGREE tool.

A literature search commissioned by the NZGG and produced by the Department of General Practice, Wellington Clinical School, University of Otago also proved invaluable during the writing of this handbook and particularly influenced Chapters 1, 10 and 12. The glossary was developed from the NHMRC and SIGN glossaries.

Copyright 2001 New Zealand Guidelines Group (NZGG)

P O Box 10 665, The Terrace
Wellington, New Zealand
Phone 64-4-471 4180
Facsimilie 64-4-471 4185
Email info@nzgg.org.nz

Published November 2001

ISBN 0-473-08181-4
INTRODUCTION TO EVIDENCE-BASED CLINICAL PRACTICE GUIDELINES

AIMS OF CHAPTER:

- to introduce the topic of Clinical Practice Guidelines (CPG)
- to provide background to CPG development in New Zealand
- to describe principles of CPG development
- to provide guidance on convening a CPG a guideline development team
- to provide guidance on the ground rules for the guideline development team
- to define the aim and target audience for the CPG
- to outline the steps of clinical practice guideline development

1.1 INTRODUCTION

Over the past decade, there has been a marked increase in the development of clinical practice guidelines. Clinical practice guidelines (CPG) are ‘systematically developed statements to assist practitioners and consumer decisions about appropriate health care for specific clinical circumstances’¹. They are tools used by healthcare professionals to assist in clinical decision making and to improve healthcare for consumers. The concept of recommendations providing guidance is not novel and for many years clinicians have used treatment recommendations, immunisation schedules, algorithms, textbooks and practice bulletins, to guide practice². The difference over the last decade has been the increasing focus on systematically summarising research evidence in order to develop more evidence based recommendations. This more rigorous approach involves multidisciplinary teams representing various stakeholders and perspectives, who systematically locate and appraise research evidence, to produce explicit, evidence based guidelines.

The purpose of clinical practice guidelines is to identify effective diagnostic, screening and treatment strategies and encourage the use of these to improve the quality of healthcare delivered and thus consumer outcomes. Defining quality in healthcare can be difficult. One simple definition of quality in healthcare is “providing the right care, at the right time for the right person in the right way”³. Quality healthcare should be appropriate, accessible, effective, safe and provided by someone who is competent and accountable for practice. Improvements in healthcare quality should result from better decisions recommended by practice guidelines at a clinical level and by influencing policies that promote allocation of resources and better delivery systems based on evidence.

This handbook is designed to be used by members of guideline development teams, to assist them to produce evidence-based clinical practice guidelines. Our expectation is that this approach will have a positive impact on the quality of healthcare delivery.
1.2 WHAT ARE CLINICAL PRACTICE GUIDELINES?

Clinical practice guidelines have been defined as decision tools to close gaps between current and optimal practice\(^4\). They are also described as

- mechanisms to improve the quality of health care and decrease costs and utilisation\(^5\)
- recommendations devised to influence decisions about health interventions\(^6\)
- tools to outline procedures to be followed thus helping doctors make decisions\(^7\)
- processes to operationalise the implementation of evidence-based practice\(^8\).

The NZ Guidelines Group (NZGG) has a broad definition of clinical practice guidelines:

> “Guidelines provide guidance in decision making at each level of interaction; between health professional and consumer, between purchaser and provider, and between ‘funder’ and ‘purchaser’” (http://www.nzgg.org.nz).

There is some disagreement over which documents should be called guidelines; the term often being used interchangeably with protocols when implying a greater degree of compliance (NHMRC 1999).\(^9\) It has been suggested that the term ‘guideline’ be applied only to a systematically developed advisory statement devised according to validated scientific methodologies\(^8\).

The New Zealand Guidelines Group (NZGG) has defined five different types of guidelines (http://www.nzgg.org.nz).\(^10\)

**Best Practice Guideline:** (also called practice guidelines, clinical guidelines, statements of best practice, and boundary guidelines). “…systematically developed statements to assist practitioner and consumer decisions about appropriate health or disability care for specific circumstances, taking into account evidence for effectiveness and competing claims,...and form a fundamental basis for planning”.

**Protocol:** “… specific guidelines which are expected to be followed in detail with little scope for variation…” Protocol guidelines are used in specialty high-risk areas e.g. emergency resuscitation, or where legislation regulates the practice e.g. forensic psychiatry.

**Consensus Based Guideline:** The most common form of guideline developed is agreement among a group of experts.

**Evidence Based Guideline:** Developed after the systematic retrieval and appraisal of information from the literature. “They usually include strategies for describing the strength of the evidence, and try to clearly separate opinions from evidence ...they make statements not just about which of two treatment options is ‘better’, but quantify the absolute differences in outcome, including both benefits and harms”.

**Explicit Evidence Based Guideline:** Developed as an evidence based guideline, “…but also projects the healthcare outcomes (benefits, harms, utilization and costs) of the change in practice on a defined population”.
Guidelines that have recommendations that are based on evidence are considered to be of greater value to practitioners and consumers because the decisions are likely to result in improved consumer outcomes.

1.3 DO CLINICAL PRACTICE GUIDELINES MAKE A DIFFERENCE TO THE DELIVERY OF HEALTH CARE?

Prior to and alongside the development of the evidence based practice (EBP) ‘movement’ and its subsidiary, the clinical guideline movement, has been the quality assurance/improvement ‘movement’ with the development of clinical audit as a means of monitoring the quality of health care. The motivation for these movements has been to assure the overall quality and outcomes of health care. Quality assurance/improvement places emphasis on the system, process and outcomes of all health care functions, whereas evidence based practice focuses on the provision (and use) of evidence based clinical information based on clinical research. Prior to the development of national guidelines, primary, secondary and tertiary level care organisations, with a commitment to quality assurance or quality improvement, developed local practice protocols in an attempt to govern the way practice was delivered. These protocol statements were mainly developed in a consensus manner and were often multidisciplinary in nature.

Systematic reviews of guideline evaluations have shown that clinical practice guidelines can be an effective means of both changing the process of healthcare delivery and improving outcomes. A review of 59 guideline evaluation studies found that, in all but 4, statistically significant improvements occurred in clinical practice after implementation. A systematic review of 87 studies on the use of guidelines concluded that 81 studies revealed evidence of improved patient outcomes.

Traditionally, guidelines have been based on consensus amongst experts. However, this process has its limitations, it usually only includes some but not all perspectives and can lead to flawed conclusions. Expert opinion does not always reflect the state of current knowledge. Furthermore, it is necessary for research literature to be analysed systematically in order to avoid biased conclusions. It is now widely accepted that guideline recommendations should be based on systematic identification and synthesis of the best available scientific evidence. This may be a daunting task given the size of research activity in some clinical areas.

The development of clinical practice guidelines involves a number of steps; topic identification, suitability screen, developing the clinical questions, searching for all available evidence, critically appraising and synthesizing the evidence, and drafting recommendations. Dissemination and implementation of clinical practice guidelines are particularly important steps, as is the need to review, update and evaluate each guideline.

1.4 GUIDELINE DEVELOPMENT IN NEW ZEALAND

The New Zealand Guidelines Group (NZGG) was established by the National Health Committee (NHC) of the NZ Ministry of Health in 1996 as an informal network of expertise in guidelines development and implementation. The NZGG became an Incorporated Society in
July 1999. Funding for the NZGG came from the National Health Committee for 3 years and then the Health Funding Authority and later the Ministry of Health. One of the primary purposes of the NZGG is to train professionals in the health and disability sectors, and consumers of these services. The long-term aim is to facilitate a culture change amongst all stakeholders in healthcare and disability to improve the quality, effectiveness and equity of service provision\textsuperscript{10}. The NZGG provides leadership and oversight of guideline training for a range of healthcare professionals and consumers in guideline development and implementation. The process initially adopted was based on that used by Group Health Cooperative, Puget Sound Seattle, USA.\textsuperscript{4} Several guidelines have now been developed within New Zealand and other guidelines have been adapted locally.

### 1.5 PRINCIPLES OF GUIDELINE DEVELOPMENT

The principles of partnership and consultation between the Crown (or its agencies) and Mana Whenua (local tribal Iwi) embodied in the Treaty of Waitangi and the Code of Health and Disability Services Consumer Rights should underpin the process undertaken by guideline development teams.

<table>
<thead>
<tr>
<th>GUIDING PRINCIPLES FOR GUIDELINE DEVELOPERS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Processes for developing and evaluating guidelines should focus on outcomes valued by consumers (eg. survival rates, quality of life measures)</td>
</tr>
<tr>
<td>2 Guidelines should be based on the best available evidence and should include a statement about the strength of the recommendations.</td>
</tr>
<tr>
<td>3 The method used to synthesise the evidence should be the strongest applicable (considered judgement)</td>
</tr>
<tr>
<td>4 The process of guideline development should be multidisciplinary and should include consumers.</td>
</tr>
<tr>
<td>5 Guidelines should be flexible and adaptable so individual circumstances can be taken into consideration.</td>
</tr>
<tr>
<td>6 Guidelines should be developed with some consideration of limitations such as resource constraints that may influence implementation strategies.</td>
</tr>
<tr>
<td>7 Guideline development should include a dissemination and implementation plan.</td>
</tr>
<tr>
<td>8 The usefulness and impact of the guidelines should be evaluated.</td>
</tr>
<tr>
<td>9 Guidelines should be reviewed and updated regularly.</td>
</tr>
</tbody>
</table>

(adapted from the NHMRC).\textsuperscript{9}
The following principles have become established from both research studies reporting on the attributes of good guidelines and the experience of guideline development teams.  

1.5.1 GUIDELINES SHOULD BE FOCUSED ON IMPROVED PATIENT/CONSUMER OUTCOMES

A health outcome has been defined as ‘a change in the health of an individual, a group of people or population which is attributable to an intervention or a series of interventions’. For example, outcome measures might include survival rates and quality of life measures. Outcomes can be positive or negative and may differ for different subgroups in the population. Many evaluations of guidelines report health processes rather than health outcomes, which does not provide information on the trends in the health of individuals or populations. Some studies report outcomes that are measurable and of interest to the researcher (such as biochemical parameters) which may be of little value to the consumer. These surrogate outcomes may or may not be associated/correlated with an important consumer outcome. An example of a surrogate outcome measure is bone mineral density measurements in older people, whereas the outcome that consumers and their providers are interested in is the fracture rate.

1.5.2 USING THE BEST AVAILABLE EVIDENCE

Guideline recommendations should be based on the best possible evidence linking the intervention and the clinical outcomes of interest. Decision making may be based on anecdote, pathophysiological information, expert opinion, or research evidence. The best possible evidence comes from well-designed research. The actual study design to provide this evidence will depend on the nature of the question asked.

1.5.3 MAKING RECOMMENDATIONS BASED ON EVIDENCE

Evidence of effectiveness is necessary but not sufficient on its own when making recommendations for treatment. Other information, such as information about individual circumstances, evidence of harm, and cost should also be considered.

1.5.4 A MULTIDISCIPLINARY APPROACH WHICH INCLUDES CONSUMERS

If guidelines are to be relevant, those who are expected to use them and to benefit from their use should play a part in their conception and development. Involving a range of generalists and specialist clinicians, allied health providers, experts in research methodology, and consumers will improve the relevance of the guideline and make it more likely to be accepted and adopted. Wider consultation with consumers, such as focus groups, may also be necessary.
1.5.5 FLEXIBILITY AND ADAPTABILITY

Guideline development should take into account
- different populations of people (e.g., age, ethnicity)
- different geographical settings (e.g., rural, urban)
- different resource allocations (e.g., access to services, PHARMAC subsidies, self-funding)
- different consumer expectations, values, and preferences.

1.5.6 RESOURCE CONSTRAINTS

Guideline developers should be mindful of available resources when developing recommendations. Where possible, an economic appraisal should be included in a guideline especially where cost data may be helpful for choosing between treatment options, as well as influencing purchasing decisions in favour of effectiveness.

1.5.7 GUIDELINE DISSEMINATION AND IMPLEMENTATION

Guidelines are disseminated and implemented in ways that take into account the particular audiences they are for. They need to be disseminated in such a way that practitioners and consumers become aware of them and are able to easily access and make use of them.

1.5.8 EVALUATION

The impact of the guideline should be evaluated at an appropriate time after the guideline has been disseminated and implemented. It is helpful if the guidelines group identifies a recommended timeframe and appropriate process.

1.5.9 REVISION

Guidelines should be reviewed, revised, and updated after an appropriate time period (usually 3-5 years) and when new evidence becomes available.

1.6 OVERVIEW OF STEPS IN CLINICAL PRACTICE GUIDELINE DEVELOPMENT

There are a number of steps in guideline development that are outlined in Fig 1.1. The steps should be consistent with the principles discussed previously in section 1.4. The order of the steps may vary and do not necessarily follow a specific timeline. The following chapters will deal with each of these topics in order.

- Topic Identification
- Suitability Screen
- Question Formulation
- Data Acquisition
- Literature searching
- Critical appraisal and application of the evidence
- Balance Sheet
- Development of recommendations and algorithm
- Dissemination and implementation
Fig 1.1 STEPS IN GUIDELINE DEVELOPMENT

Problem Identification

Formation of Guideline development team

Suitability Screen

Is development of guideline an appropriate solution? NO

YES

Proceed with guideline development process

Question Formulation

Current data acquisition and literature searching

Identify evidence

Assess evidence

Determine benefits & harms

Balance sheet*

Develop recommendations and algorithm

Disseminate

Implement

Evaluate

Improve

NO

Project outcomes unlikely to benefit from a guideline being developed

Share Evidence

Consider the options

* Balance sheets with cost may be omitted if there are too many assumptions to be made
2.1 TOPIC IDENTIFICATION

Topic or problem identification is identifying the areas where a CPG could make a difference to the delivery of health care. A topic is suitable for evidence-based guideline development if:

- The topic is clinically important affecting large numbers of people with substantial morbidity or mortality (the burden of illness);
- The topic is complex enough to initiate debate about the recommendations;
- There is evidence of variation between actual and appropriate care;
- There are no existing valid or relevant guidelines available to use;
- There is evidence available to support evidence-based guideline development;
- The recommendations will be acceptable to the potential users (consumers/communities);
- Implementation of the guideline is feasible, will not exhaustively use the communities resources and barriers to clinical change are not so high that they cannot be overcome.

If there is already a guideline in existence that addresses the same issue or problem, another guideline may not be needed. An evaluation and/or adaptation of that guideline may be more appropriate than developing a new guideline. The AGREE tool can be used for the evaluation of a current guideline. There may also be some concern that the lack of treatment or over treatment may lead to poorer consumer outcomes, even harm.

2.2 SUITABILITY SCREEN

A suitability screen is a systematic process used to establish how successful the development of a guideline in a particular clinical area is likely to be. Given limited resources for guideline development, efforts are best directed to projects that can demonstrate significant positive changes in outcomes are likely based on valid scientific studies. The convenor of the group should consider the questions in the suitability screen and discuss at the first meeting. The following questions make up the suitability screen:
(1) Does the project have an owner (an individual who takes responsibility for organising and leading the effort)? Projects that have an identified and motivated owner/leader are more likely to proceed in a timely fashion and be completed.

(2) Can the proposed change be measured? For example, can changes in health outcomes, practice, prescribing patterns, the number of procedures or interventions, or changes in resource utilisation/cost, be quantified? Before significant resources are committed to the project, it is prudent to ensure the resources allocated are likely to result in a benefit for consumers. This is described in Chapter 4.

(3) Is there a suitable guideline already available that could be adapted? Use AGREE instrument to evaluate. (Appendix 4)

(4) What does a brief literature search reveal? Is there adequate literature to make an evidence based decision about appropriate practice? The evidence based approach may not be appropriate for evaluating all practices but it is best suited to clinical areas where there is valid scientific literature which supplies information about the potential benefits, harms and costs of therapy. If this is not available, assumptions will need to be substituted. The more significant the assumptions, the less valid the results may be. In some clinical circumstances where interventions have known harms and unproven benefit, a lack of evidence may be utilised to make recommendations about appropriate practice. Chapter 5 provides information on how to go about searching.

(5) Will the proposed change in practice result in sufficient changes in outcomes to justify the efforts? How big is the gap between what the evidence supports and current practice? How much effort will it take to close the gap? Answer this by considering the internal data (eg from audit) and the external data (from the literature) and make some assessment of the size of the gap between current and optimal practice.

(6) Is there a reasonable likelihood that the change could be implemented? Both short and long term steps may be needed.
**Suitability Screen Tool**

Given limited resources, efforts are best directed to projects that can demonstrate significant changes in outcomes based on valid scientific studies. This tool will address the six issues that are predictors of success.

1. **Is there an owner for the project (preferably an individual)?**

2. **Can the proposed change be measured (health status, cost)?**
   - Is there a gap between current and optimal practice?
   - Are there outcomes that can be measured?
   - Can the data be captured?

3. **Is there a suitable guideline already available that could be adapted?**

4. **Brief literature search**
   - Is there adequate literature to make an evidence-based decision about appropriate practice?
   - In some clinical circumstances (where interventions have known harms and unproven benefit) a lack of evidence may be utilised to make recommendations about appropriate practice.

5. **Would the proposed practice change result in sufficient change in outcomes (health status, provider and consumer satisfaction and cost) to justify the effort?**
   - How big is the gap?
   - How much effort will it take to close the gap?

6. **Is there a reasonable likelihood that we could implement the change?**
Example:

**SUITABILITY SCREEN FOR GUIDELINES IN HEAVY MENSTRUAL BLEEDING (1997)**

A suitability screen was performed at the beginning of the Guideline process to ensure that it was likely that investing in a guideline on this topic would be worthwhile. The screening process followed these steps and involved the project leader with assistance from other group members.

1. *Was there an owner for the project?* Yes

2. *Could the proposed change be measured?* Yes. There are outcomes that can be measured which were identified as follows:
   - hysterectomy rate in women under the age of 50 years
   - prescribing rates of medical alternatives
   - number of consultations and referrals (GP databases)

3. *Is there a suitable guideline already available that could be adapted?* No suitable guideline was identified.

4. *What does a brief literature search reveal?* A brief literature search was carried out. There were approximately 50 RCTs of medical and surgical therapy of heavy menstrual bleeding and at least 20 studies of diagnostic tests identified.

5. *Is the best treatment supported by the evidence being used?* The prescribing rate of Tranexamic acid [effective therapy] amongst gynaecologists in this country was low with only 10% using it as a third line measure while the prescribing rate for luteal phase progestogen [ineffective therapy] was 50%. There was no current use of the levonorgestrel intrauterine system.

6. *Would the proposed change result in sufficient change in outcomes and justify the effort? How big is the gap between current practice and optimum care?* New Zealand has the fifth highest hysterectomy rate in OECD countries. 21% of New Zealand women have a hysterectomy before menopause. In a population of premenopausal women in the United Kingdom, the rate is 17%, whereas in Denmark the lifetime prevalence is 10%.

7. *How much effort would it take to close the gap?* It would require an educational and implementation programme for primary and secondary care doctors as well as changes in access to certain diagnostics and medication.

8. *Is there a reasonable likelihood that changes could be implemented?* Yes, with appropriate funding.
QUESTION FORMULATION FOR CLINICAL PRACTICE GUIDELINES

AIM OF CHAPTER:

- To teach principles of good question formulation in order to support the development of a literature search strategy, recommendations and algorithms for CPGs.

Effective and efficient guideline development involves asking and answering clinical questions using research evidence where possible. Although it is not always recognised, each interaction between a clinician and a patient/consumer generates several clinical questions. Sometimes the answers are known or the search for information is brief as the information is at hand. On other occasions, the information is further afield and effort is required to access it. In these cases, it is important to have well designed clinical questions in order to make the search for the information more efficient. Research has shown that each practitioner has about 30 needs for new information each day and that only 30% of information needs are met\(^\text{18}\). In many cases, the information comes from colleagues.

Well-developed clinical questions can form the basis of the guideline structure. They assist guideline developers in a number of different ways. They help the guideline development team to focus on evidence that is relevant to the consumers and clinicians, making the searching more efficient. Although there may be a lot of other interesting clinical research evidence available, unless it is answering a question that clinicians, policy makers or consumers are asking it is probably not worthwhile pursuing.

3.1 THE DEVELOPMENT OF A SERIES OF QUESTIONS

Getting the question right is not always easy. Asking “why is this important to answer?” may help focus the guideline development team and ensure a shared understanding of the scope of the guideline. Once questions are formulated, the systematic search for evidence from the literature can be planned in an efficient manner. Question formulation also involves deciding what type of question you are asking. Is it a question about which diagnostic test is best? Which therapy is best? What prognosis does the patient have? etc. Each of these questions may be addressed by a different type of study. The multiple perspectives of the guideline development team may be helpful in getting the question or questions right.

As it can be difficult to translate information needs into questions that can be answered by the literature, a series of examples of clinical episodes or scenarios have been developed. At
each step of a typical clinical scenario, key questions arise and need to be addressed in a logical sequence.

Examples:

What are the symptoms suggestive of the clinical condition?

What is the prognosis for this condition?

What diagnostic test/s (if any) should be arranged to confirm the diagnosis?

If the diagnosis is confirmed what are the best treatment options?

What other diagnostic/screening should be considered?

What other treatment/s are indicated for this patient?

What are the potential benefits and harms of the treatments?

What are the costs of the various steps?

What co-morbidities change the recommended approach?

It is highly likely that there is even greater complexity to consider depending on the clinical situation or health service being addressed. If these clinical steps are translated into 5-part “PECOT” questions, this provides a framework for the literature searching strategy, evidence summaries, recommendations and algorithm development.

3.2 THE 5 PARTS TO A WELL FORMULATED CLINICAL QUESTION

Each question has 5 parts that should be considered:

1. **Patients/consumers/participants**: Which patients or participants are we interested in? How can they be best described? Are there subgroups that need to be considered?

2. **Exposure**: Which intervention, treatment, factor, disease or approach are we interested in?

3. **Comparison**: What is/are the main alternative/s to compare with the exposure?

4. **Outcome**: What is really important to the patient/consumer? What does this exposure affect? Which outcomes should be considered? - intermediate or short term measures (eg lowering in cholesterol levels), mortality, morbidity and treatment complications, rates of relapse, late morbidity and readmission, return to work and physical and social functioning and other measures such as quality of life, general health status. Cost to the consumer is another outcome that should be considered.
5 Time: Over what time frame is it reasonable to expect an effect? Time frame for outcomes should be considered in all cases.

### 3.3 EXAMPLES OF WELL FORMULATED CLINICAL QUESTIONS

**Clinical Scenario 1**

Mathew, a 20 year old student asks his GP to check a mole on his back that has been bleeding intermittently. The mole is large, black and raised but has regular edges. The doctor thinks that this looks like a melanoma but possibly it is just a mole that has been repeatedly traumatised by rubbing/catching on clothing. There is no lymph node enlargement in his neck or armpits and his chest is clear.

The mole is excised with wide margins and sent for analysis. The report comes back stating that it is a malignant melanoma of x mm depth, x length and x mm width and the excision is complete (the margins are clear of abnormal cells). The GP was relieved not to have missed this important diagnosis but wonders what is the diagnostic accuracy of visual inspection of moles and skin lesions?

Mathew is referred to the oncology specialist for review and advice on further treatment. Although the excision is complete and other tests have shown no spread of the cancer to other parts of the body, the specialist still advises a course of chemotherapy. Matthew isn’t sure about this and seeks his GP’s advice—what are his chances of getting a recurrence of the cancer and if it does come back, is he going to die? Does he really need chemotherapy or are there other alternatives?

From this scenario 3 major questions arise:

**Diagnosis Question**: What is the diagnostic accuracy of visual inspection of moles and skin lesions?

**Prognosis Question**: What are his chances of getting a recurrence of the cancer and if it does come back, what is the likely prognosis?

**Therapy Question**: Does he really need chemotherapy or are there other alternatives?

These questions need to be translated into 5-part PECOT questions.

**Diagnosis Question**: In young European adults aged 18-30yrs with a skin lesion or mole (Participant), what is the accuracy of visual inspection and clinical examination (Outcome) to detect those who have malignant melanoma confirmed by skin biopsy (Exposure), compared to those who do not have a malignant melanoma at skin biopsy (Comparison) at one point in time (Time)?

This can be put into a PECOT chart as follows:

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>EXPOSURE</th>
<th>COMPARISON</th>
<th>OUTCOME</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>In young European adults (aged 18-30 years) with a mole or skin lesion</td>
<td>Malignant melanoma on biopsy</td>
<td>No malignant melanoma on biopsy</td>
<td>Accuracy of diagnosis by visual inspection and clinical examination</td>
<td>One point in time</td>
</tr>
</tbody>
</table>
**Therapy Question:** In young European adults aged 18-30yrs who have a malignant melanoma fully excised from the back and no evidence of metastatic spread (Participants), does chemotherapy with drug X at this dose and duration (Exposure) compared with no chemotherapy (Comparison) improve survival (Outcome) over a five year period (Time)?

<table>
<thead>
<tr>
<th>Participants</th>
<th>Exposure</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>In young adults (aged 18-30 years) with a malignant melanoma</td>
<td>Does size, depth, central body location, clear tissue margins, no evidence of secondary spread</td>
<td>Different size, depth, location or spread</td>
<td>Influence the likelihood of recurrence or survival</td>
<td>Over 5 years</td>
</tr>
</tbody>
</table>

There are many therapy questions that could be asked as seen in PECOT chart. Also, there may be several other outcomes of interest eg. quality of life, harms and side-effects.

**Prognosis Question:** In young European adults aged 18-30 years who have a malignant melanoma (Participants), does size, depth, central body location, clear tissue margins and no evidence of secondary spread (Exposure=Possible prognostic factors), compared to different size, depth, location or spread (Comparison) influence recurrence of malignant melanoma and survival (Outcome) over a five year period (Time)?

<table>
<thead>
<tr>
<th>Participants</th>
<th>Exposure</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>In young adults (aged 18-30 years) with a malignant melanoma that is fully excised with no evidence of metastatic disease</td>
<td>Chemotherapy Drug X (dose and duration) OR Drugs X, Y and Z OR combination of drugs OR alternative therapies eg. homoeopathy, naturopathy using A, B or C</td>
<td>No chemotherapy OR alternative therapies using A, B or C OR single drug therapy</td>
<td>Survival</td>
<td>Over 5 years</td>
</tr>
</tbody>
</table>
Clinical Scenario 2

The coronary care team in a moderate sized hospital is having a meeting with its service manager. The team wishes to set up a comprehensive, cardiac rehabilitation programme, including exercise, lifestyle education and psychosocial counseling for patients after discharge from hospital following an acute myocardial infarction (heart attack). The service manager says he needs some evidence that this would be a good use of their already overstretched budget. During the course of the ensuing conversation, the coronary care nurse observed that those who seem depressed after their heart attack are more likely, in her experience, to be readmitted or die. Maybe, more needed to be offered to these people. In reply, the cardiologist said, “surely everyone after a heart attack is feeling anxious and low. How can you tell who needs extra psychological help?” The nurse replied that she’d heard of a questionnaire called the Beck Depression Inventory that could be used to determine those who were clinically depressed or not.

There are several questions coming from this scenario:

- **Diagnosis:** How accurate is the Beck Depression Inventory (BDI) for diagnosing clinical depression?
- **Prognosis:** Does depression influence the risk of death or re-hospitalisation following a heart attack?
- **Therapy:** Does comprehensive, multifactorial cardiac rehabilitation improve quality of life, reduce hospital readmissions and cardiac mortality?

These questions need to be translated into 5-part PECOT questions.

- **Diagnosis:** In patients following a heart attack who are admitted to a coronary care unit (Participants), how accurate is the Beck Depression Inventory (Outcome) for determining whether a person has depression according to DSM IV criteria (Exposure) or no depression (Comparison) at one point in time (Time)?

This can be put into a PECOT chart as follows:

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>EXPOSURE</th>
<th>COMPARISON</th>
<th>OUTCOME</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>In consumers following a heart attack who are admitted to a coronary care unit</td>
<td>Depression according to DSM IV</td>
<td>No depression according to DSM IV</td>
<td>Diagnosis of depression using BDI</td>
<td>One point in time</td>
</tr>
</tbody>
</table>

**Prognosis:** In patients following a heart attack who are admitted to a coronary care unit (Participants), does the presence of clinical depression according to DSM IV criteria, (Exposure) compared with those who are not depressed (Comparison) affect re-hospitalisation rate or mortality (Outcome) over the next year and over the next 5 years (Time)?

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>EXPOSURE</th>
<th>COMPARISON</th>
<th>OUTCOME</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>In consumers following a heart attack admitted to a coronary care unit</td>
<td>Clinical depression according to DSM IV criteria</td>
<td>No depression</td>
<td>Affect rehospitalisation rate and mortality</td>
<td>Over 1 year and/or Over 5 years</td>
</tr>
</tbody>
</table>
Therapy: In consumers following a heart attack who are admitted to a coronary care unit (Participants), does comprehensive cardiac rehabilitation including exercise, lifestyle education, psychosocial education and counseling (Exposure), compared with usual care (Comparison), improve quality of life and reduce the risk of rehospitalisation, reinfarction or cardiac and total mortality (Outcome) over the next year and over the next 5 years (Time)?

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>EXPOSURE</th>
<th>COMPARISON</th>
<th>OUTCOME</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>In consumers following MI who are admitted to a coronary care unit</td>
<td>Comprehensive cardiac rehabilitation programme including exercise, lifestyle, education and psychosocial counseling</td>
<td>Usual care</td>
<td>Improve quality of life. Reduce risk of reinfarction, rehospitalisation, cardiac mortality, total mortality</td>
<td>Over 1 year and/or Over 5 years</td>
</tr>
</tbody>
</table>
INTERNAL DATA ACQUISITION

AIM OF CHAPTER:

- To locate data on current practice in order to perform the suitability screen and develop the balance sheet volumes.

Internal data usually means data from the health service within which you are working or researching (either local or national) that may be a source of information for evaluating proposed changes in clinical practice. Many health care organisations have data systems which can be used to monitor care including the patients treatment details, track outcomes, adverse effects and resource utilisation. In many cases, audits may be available. These data can be used to determine the demographic characteristics of a patient population, current clinical outcomes and the processes of care used including interventions.

Internal data define and quantify current practice and will provide an indication or baseline for determining the gap between current and optimal behaviour as outlined in the evidence. The amount of data collected should be limited to essential items that reveal the size of the problem from the perspective of policy makers, management, clinicians and consumers. The items that should be collected include recorded clinical outcomes (mortality, morbidity), intermediate outcomes for example (blood pressure measurement), rates of use, consumer satisfaction and cost (consumer and health service).

There are pitfalls in seeking internal data. Considerable resources and a certain level of analysis are required. Administration datasets may not necessarily have been assembled with accuracy in mind. In some cases there may be no information available and an audit of some kind may be necessary. Often these are retrospective chart reviews.

A series of questions should be used to locate these data which are then summarized. It may be helpful to create internal data tables. The following questions should be asked of these data:

4.1 STEPS IN DATA ACQUISITION

Step 1: What information would you like to help define current practice?

- Define the population (demographics).
  Are there data available to help determine whether this is a major/significant health and disability consideration (condition, diagnostic test, or therapeutic intervention) found in the patient population?
• What are the current clinical and disability outcomes for these patients?
  • Health and disability outcomes:
    Diagnoses
    Morbidity
    Mortality
    Functional status

• Intermediate outcomes (eg blood pressure measurement and laboratory values)
• Rates of use and process information
  Visit rates
  Levels of support
  Laboratory utilization
  Drug utilization
  Diagnostic interventions (and adverse outcomes)
  Therapeutic interventions (and adverse outcomes)
• Is there a significant local variation in care or outcomes?
• Are there data available on the differences in resource utilization/cost with different internal practices?
• What is the total cost of care of the population with current practices?

Step 2: What data can you readily obtain?

Efforts should be directed at data that the organisation already captures. This means limiting your data acquisition to the vital few. Things to be considered:

• Time frame
• Regional versus system wide data
• Are the data internal or external (service provided by the system or purchased outside)?
• What are the limitations of the data?
• Format for data presentation

Step 3: Should you commit resources to targeted data collection? What data could you reasonably collect and how would it help the guideline process?

New Zealand sources of data:
• GP Waimeca (McAvoy 1993)
• GP database (RNZCGP Dunedin, Computernet)
• Drug Utilisation databases, e.g. IMS, PHARMAC
• Specific studies e.g. Auckland Heart Study for cardiovascular disease
• hospital data
• Ministry of Health
• labs
• funded research
• New Zealand Health Information Service
• National Mortality / Morbidity figures
• Health Benefits Ltd
• Statistics New Zealand
• Accident Compensation Commission
FINDING THE EVIDENCE - LITERATURE SEARCHING FOR GUIDELINE DEVELOPMENT

AIM OF CHAPTER:

- To describe the sources of information and aspects of the process of conducting literature searches

5.1 LITERATURE SOURCES

The biomedical literature is huge and growing daily with a wide range of paper journals, electronic publications, abstracts, posters and books available. There is no single source or electronic search that will yield all the required evidence. A search strategy that incorporates a range of sources including several relevant electronic databases, abstracts of relevant scientific meetings, printed bibliographies/reference lists together with direct contact with researchers and expert practitioners in the field is more likely to locate all of the relevant evidence.\textsuperscript{19,20} Guideline development teams may need to use health information specialists to assist them to ensure a comprehensive search of the literature is achieved. The following is a list of the more common sources.

5.1.1 ELECTRONIC DATABASES

1. MEDLINE encompasses information from Index Medicus, Index to Dental Literature, and International Nursing, as well as other sources of coverage in the areas of allied health, biological and physical sciences, humanities and information science as they relate to medicine and health care, communication disorders, population biology, and reproductive biology. MEDLINE contains bibliographic citations and author abstracts from over 4,000 journals published in the United States and in 70 foreign countries. It has 11 million records dating back to 1966. Abstracts are included for about 67\% of the records.

2. Cochrane Library – available on CD or Internet by subscription (http://www.update-software.com/cochrane/cochrane-frame.html); some parts available free on Internet. Includes 5 major databases that are searched simultaneously.
   - Cochrane Database of Systematic Reviews (CDSR) – contains completed reviews and protocols of reviews in progress on interventions (therapies) and preventive strategies across all areas of healthcare.
• Cochrane Controlled Trials Register (CCTR) Contains more than 250,000 references to randomised controlled trials of interventions. Includes references located by hand searching, conference abstracts, letters as well as references listed in other databases.

• Database of Reviews of Effectiveness (DARE) is prepared by the NHS Centre for Reviews and Dissemination at the University of York, England. This database provides information on non-Cochrane published reviews of the effects of health care.

• The NHS Economic Evaluation Database is a companion to the CRD Database of Abstracts of Reviews of Effectiveness (DARE). It is a register of published economic evaluations of health care interventions.

• Health Technology Assessment Database contains reviews of various topics not restricted to therapy/interventions

3. EMBASE, the Excerpta Medica database, produced by Elsevier Science, is a major biomedical and pharmaceutical database indexing over 3,500 international journals in the following fields: drug research, pharmacology, pharmaceutics, toxicology, clinical and experimental human medicine, health policy and management, public health, occupational health, environmental health, drug dependence and abuse, psychiatry, forensic medicine, and biomedical engineering/instrumentation. There is selective coverage for nursing, dentistry, veterinary medicine, psychology, and alternative medicine.

4. CINAHL, Cumulative Index to Nursing and Allied Health. Indexes virtually all English language publications in the field - more than 1200 journals. The database also provides access to healthcare books, nursing dissertations, selected conference proceedings, standards of professional practice, educational software and audiovisual materials in nursing.

5. PsycINFO covers the professional and academic literature in psychology and related disciplines including medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, linguistics, and other areas. Coverage is worldwide, and includes references and abstracts to over 1300 journals in more than 30 languages, and to book chapters and books in the English language. Over 50,000 references are added annually.

6. ERIC (Education Resources Information Center) is presently the largest education database in the world. ERIC contains over one million citations covering research documents, journal articles, technical reports, program descriptions and evaluations, and curricular materials in the field of education.
5.1.2 INTERNET

On the Internet, there are numerous databases and sites containing information and reviews of research. There is wide variation in the quality of material and scientific rigour of the evaluation. Guidelines produced in New Zealand are available through the website of the New Zealand Guidelines Group (http://www.nzgg.org.nz). This website also contains links to a number of other useful websites for those interested in evidence based health care.

Free electronic copies of guidelines produced in other countries can be accessed through Health Technology assessment websites, for example

- Canadian Medical Association – CMA Infobase (http://www.cma.ca/cpgs/index.asp)
- National Guideline Clearing House (structured abstracts of guidelines only at http://www.guidelines.gov)
- US guidelines can be accessed through Health Services Technology Assessment Text (HSTAT) at http://www.hstat.nlm.nih.gov/

Other websites to note

- National Research Register - database of ongoing and recently completed trials funded by NHS in UK (http://www.doh.gov.uk/research/nrr.htm)
- NLM Gateway (http://gateway.nlm.nih.gov/gw/Cmd) includes Clinical Trials (a database of ongoing US trials)
- TRIP Database (US) Translating Research into Practice (http://www.tripdatabase.com/)
- PubMed free - internet version of Medline available through NLM Gateway or directly at http://www.pubmed.gov/entrez/query.fcgi
- SUMSearch smart search engine - http://SUMSearch.uthscsa.edu/searchform4.htm

5.1.3 HANDSEARCHING

Time consuming page-by-page searching through publications such as the abstract books from relevant conferences and scientific meetings can yield references to ongoing or unpublished research that may be very relevant, and not available electronically. It is recommended that contact be made with a relevant Cochrane Review Group to avoid duplication of effort.

5.1.4 REFERENCE LISTS

Searching through the reference list and bibliographies of relevant reviews and research already obtained can yield new research that may not be identified through electronic searches.
Grey literature, such as unpublished technical reports, dissertations and theses may be identified in this way. Scisearch Database gives information about where a given reference has been subsequently cited. (for example, it may identify an economic evaluation published subsequent to an RCT).

5.1.5 DIRECT CONTACT WITH PHARMACEUTICAL COMPANIES/RESEARCHERS

Requests for unpublished information or preliminary results may be successful.

5.2 PRINCIPLES OF SEARCHING OVID DATABASES (WITH PARTICULAR REFERENCE TO MEDLINE OVID)

Sensitive search strategies increase the probability that the searcher will locate all the relevant information on a given topic. However, a sensitive strategy will also retrieve a large amount of irrelevant material. For guidelines, it is recommended that searchers use a sensitive search strategy and subsequently use filters or limits to improve the precision. Evidence based clinical practice guidelines require that all the evidence is located and appraised and therefore literature searches should err on the side of sensitivity.

5.2.1 INCREASING SENSITIVITY (WHERE THERE ARE TOO FEW HITS)

- Subject Headings/Trees

Medline and Embase both assign index terms/subject headings to the references (indicated by / after the word eg asthma/). This controlled vocabulary assists the searcher to obtain information by reducing the chances that differences in terminology may cause the searcher to miss valuable information. Subject headings are arranged in a tree structure with broad headings over more specific headings.

Subheadings (eg /dt) for drug therapy are not recommended for use in a sensitive search because it is “estimated that 50% of articles in Medline are inadequately or incorrectly classified by subheading.”

- Scope notes (i)

These state the definition of the subject heading as used in the database, the year the heading was introduced, other related subject headings and possible synonyms for textword searching.

- Explode

A command which causes the database to search on the given subject heading and the heading(s) beneath it on the tree. Indexers are instructed to use the most specific index term possible so if you search on ‘asthma/’ you will miss the references indexed ‘asthma in children/’ unless you use ‘exp asthma/’
• Textword Searching

This is also known as free text searching where the database is asked to search for a word in the text fields – usually title & abstract, (and sometimes subject heading fields).

• Truncation and Wildcards

This is another tool to increase sensitivity. When searching for information on pregnancy the use of the text term pregnan$ will retrieve references including pregnant, pregnancy, pregnancies (but also pregnandiol!)

• Adjacency

Many databases include a phrase list of words that commonly occur together which the database searches as a phrase, eg blood pressure. When searching for phrases which may not be on the phrase list, the use of the adjacency command increases the sensitivity by retrieving reference in which both words appear, but not necessarily consecutively or in the order specified by the searcher. Eg acute adj3 haemorrhage.tw will retrieve acute subarachnoid haemorrhage, but acute haemorrhage.tw will not

• Boolean ‘OR’

Combining truncated text word search terms OR exploded MeSH terms on the same topic will give the best sensitivity – a belt and braces approach!

• Pearl growing – a technique for improving search sensitivity.

Review results of preliminary search, look at the references retrieved so far, and the associated MeSH terms. Identify new terms previously overlooked, new spellings, word endings, broader/narrower MeSH terms. Modify the search strategy to incorporate the new terms.

5.2.2 INCREASING PRECISION (WHERE THERE ARE TOO MANY HITS)

• Boolean AND

Combining groups of terms related to different aspects of the question with AND will give a result set in which both aspects of the question are addressed. Eg (exp asthma/ or asthma$.tw) and (respiratory tract infections/ or respirat$ adj3 infect$.tw)

• Boolean NOT

This can be used to remove a narrow group of references that are not required
Eg searching on non drug therapy for asthma try exp asthma/ NOT dt.fs)

- Quality filters

There are many validated search filters (aka search hedges or quality filters) that are designed to select specific types of study design. These can be added to a search on a given topic, eg in order to identify relevant randomised controlled trials about asthma, one could do a sensitive search on asthma and add a quality filter for therapy. Filters of different sensitivity and specificity are available (see appendix) but filters are generally more sensitive than limits.

- Limits

Databases enable searchers to limit the search to eg year(s) of publication, specified age groups, specified languages, publication types, human etc.

Notes:
- In Medline, limit to English is not always required because many non-English references in Medline have an English abstract.
- Limit to publication type may miss relevant references because this term is relatively new and inconsistently applied.

5.3 SAVING SEARCH STRATEGIES AND RESULTS

Search strategies can be easily saved on Ovid and rerun to retrieve references recently added to the database. The SDI feature allows one to save the search strategy and have the system e-mail the new references found each time the database is updated or at regular intervals. PubMed also has a facility to save searches but long searches are more difficult to save on this platform.

It is strongly recommended that the electronic records of the references found be stored in a bibliographic database such as Endnote, Reference Manager or ProCite. This allows the use of Cite-as-you-write and facilitates the management of the many references collected during the preparation of a guideline.

5.4 SEARCH FILTERS (AKA QUALITY FILTERS/SEARCH HEDGES)

These are strings of search terms devised by experts and validated by research which are designed to retrieve defined types of study designs. There are many variations and the ones included here are from the CASPFEW website and were developed at the Institute of Health Sciences, Oxford. They are designed for Medline OVID and require modification for use on other platforms or databases.

5.5 PUBMED

PubMed is available free on the Internet, has records added daily and comprises MEDLINE and PreMEDLINE and a selection of citations electronically supplied by publishers. It is
mentioned separately here because the PubMed search engine is significantly different from Ovid. Searchers who wish to use PubMed for comprehensive systematic searches are advised to study the Help/FAQ section. However, the functionality of this database is continuously improving and it is expected that future refinements will address current difficulties.

The following is a summary of the features of PubMed (taken from the online FAQ section at time of writing).

- **Automatic Term Mapping** is the default search that matches the query against MeSH (exploded), Journals, Phrase list and Author index. Automatic Term Mapping can be ‘turned off’ by the use of truncation symbol eg heart attack* or by entering a field descriptor eg [AU]

- **Boolean operators** are in CAPITALS - AND, OR, NOT

- **Only the Medline part of PubMed has MeSH Terms** – affects the use of limits

### 5.5.1 INCREASING SENSITIVITY ON PUBMED – WHEN THERE TOO FEW HITS

- Avoid truncation and wildcards eg infection* will retrieve infection/s/ious but not infection control (because * turns off the automatic mapping function)

- Increase the use of “OR”

  Use synonyms, spelling variations, abbreviations, - combine with OR

  Eg “Vitamin H” OR biotin, diarrhoea OR diarrhea, UTI OR Urinary Tract Infection, Acetaminophen OR paracetamol

- Decrease the use of “AND”

  You may only need to search on the intervention and comparison

- Check for LIMITS – once set the limit stays until you remove it.

- Try NOT animal [MeSH] – instead of limit to human

### 5.5.2 INCREASING PRECISION – WHEN THERE ARE TOO MANY HITS

- Increase the “AND”

- Use additional terms eg from the patient/exposure/comparison/outcomes columns of the PECOT chart.

- Use NOT to remove unwanted references (noise). Choose narrow terms eg NOT animal will remove any reference where the word is used in the text - NOT animal [MeSH] is better

- Limits –use the limits available in PubMed - year of publication/age group/language

- Fields - limit your query to a specific field(s) eg “Yang YL” [AU]
• Clinical Queries - use these built in filters to retrieve the types of reference that you require

5.5.3 SAVING SEARCHES ON PUBMED - 2 WAYS

• Cubby – set up a cubby with username and password and you can retrieve searches for future updating

• URL button from the Details screen. This embeds the search strategy as part of the URL, which can be added to your bookmarks and renamed appropriately using the Edit function of your browser.

NB Both methods are unable to save searches that include a search statement number eg #3 AND #7

Working in the Preview/Index screen allows one to develop a search in a format that can be saved. Alternatively, the final search can be re-entered as one statement in the query box.

5.5.4 OTHER USEFUL FEATURES OF PUBMED

• Details - shows how PubMed has translated your query – and allows you to modify

• Preview/Index - useful to browse fields for search terms, while viewing the last 3 queries from history. (Similar to the main search page in OVID)

• Clipboard Citations (up to a max of 500) can be selected (place a tick in the box) and added to the clipboard. Click on ‘Add to Clipboard’ without selecting and the first 500 citations from a search will be added. Citations on the clipboard can be sorted according to criteria from the pull down menu and may be displayed in different formats. The contents of the Clipboard can be printed or saved as a text file.

• Single Citation Matcher - a fill-in-the-blank form that allows you to find a citation if you know some of author, journal name and date. If the information that you have is title word and date you can use field searches in PubMed.

• Link Out - shows where the citation is available as full text on line. (Sometimes free, sometimes by subscription)

• Clinical Queries – contains built in validated search filters for therapy, diagnosis, etiology and prognosis (either sensitive or specific) which can be combined with a short search string.
5.6 OBSERVATIONS OF THE PROCESS

For a guideline development team, there are likely to be at least 2 stages of literature searching.

- A quick look at the published literature as part of the suitability screen, perhaps looking specifically for systematic reviews and guidelines published in other countries.

- A second much more thorough, comprehensive search to identify all relevant information, published and unpublished, to answer all the questions in the guideline. This search strategy will cover multiple databases and other sources, will need to be documented and perhaps set up in such a way that will enable team members to rerun the search to identify new information.

The guideline development team may want to assign 2 people or groups to carry out the electronic searches independently and pool their results. As a comprehensive search is often one of low precision, guideline development teams may need to develop a set of criteria by which abstracts of articles identified by a search are evaluated, and a consensus decision made on whether to retrieve a copy of the full article for appraisal.

For every guideline, there will be a number of questions concerning each of the areas of diagnosis, prognosis, therapy, and harm, which will form the various chapters in the document. These questions need to be clarified before the search strategy is devised. The resultant search strategy will comprise a group of terms (combined with “OR”) relating to the participants/condition of interest, and a group of terms relating to each specific area of questions such as diagnosis, prognosis etc. Various combinations (using “AND”) of these groups of search terms will be used to identify the research that contains the information related to each chapter of the guideline. The basic search strategy will of course require modification to suit the subject headings from different databases.

Searchers are advised to create a bibliographic database containing all the literature retrieved. The use of an electronic database to contain all the references facilitates the insertion of accurate, complete references into the final document.

5.7 EXAMPLE

The search example that follows is based on one of the scenarios introduced in chapter 3. The search is for papers to answer a single question but should serve as a template for creating a guideline search strategy that would have a similar structure, but would be longer and more comprehensive.

Break the question into the following components

<table>
<thead>
<tr>
<th>Participants</th>
<th>Exposure</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>In consumers following a heart attack who are admitted to a coronary care unit</td>
<td>Depression according to DSM IV</td>
<td>No depression</td>
<td>Re-admission Rehospitalization Mortality</td>
<td>1 year 5 years</td>
</tr>
</tbody>
</table>
**Question:** Does depression affect the risk of death or rehospitalization after a heart attack?

**PECOT SEARCH PLAN**

<table>
<thead>
<tr>
<th>Participants</th>
<th>Exposure</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression disorders/ OR Depression/ OR Depres$.tw*</td>
<td></td>
<td></td>
<td>Hospitalisation/ OR Patient readmission/ OR Exp Mortality/ OR Mortalit$.tw OR Survival analysis/</td>
<td></td>
</tr>
</tbody>
</table>

To increase the precision of this search further, one could
- add another limit – eg age group
- remove the text search on depression (line 6) – includes depressed heart rate
- use a quality filter for prognosis (see Table 5.1)

*This is a very broad term; including a text search here may be too imprecise.*
Problem – sensitivity/precision balance

This topic illustrates the importance of using multiple strategies to identify the relevant literature.

There is a systematic review:

Hemingway H, Marmot M. (1999) Evidence based cardiology: psychosocial factors in the aetiology and prognosis of coronary heart disease. BMJ 318 (7196) 1460-7. which is BOTH relevant to address this question, AND is included in the Medline database.

The search demonstrated above did not find this review because

- The MeSH term coronary disease/ was used instead of myocardial infarction/ (Indexers are instructed to use the most specific relevant MeSH but errors can occur)
- There is no abstract from Hemingway et al included on Medline. This means that text word search terms search the title field only, where the word ‘psychosocial’ is used to include ‘depression’.

(Hemingway et al was identified by a search through the reference list of another study identified by the Medline search.)
### 5.8 SEARCH FILTERS FOR MEDLINE OVID FROM

- CASPFEW Institute Of Health Sciences, University Of Oxford, UK, &
- Health Information Research Unit McMaster University, Canada.

<table>
<thead>
<tr>
<th>Filter Type</th>
<th>Filter Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis sensitivity filter</strong></td>
<td>1. exp “sensitivity and specificity”/ 2. sensitivity.tw. 3. di.xs. 4. du.fs. 5. specificity.tw. 6. or/1-5</td>
</tr>
<tr>
<td><strong>Diagnosis specificity filter</strong></td>
<td>1. exp “sensitivity and specificity”/ 2. (predictive and value$).tw. 3. #1 or #2</td>
</tr>
<tr>
<td><strong>Therapy sensitivity filter</strong></td>
<td>1. randomized controlled 2. dt.fs. 3. tu.fs. 4. random$.tw. 5. or/1-4</td>
</tr>
<tr>
<td><strong>Therapy specificity filter</strong></td>
<td>1. (double and blind$).tw. 2. placebo.tw. 3. 1 or 2</td>
</tr>
<tr>
<td><strong>Prognosis sensitivity filter</strong></td>
<td>1. incidence/ 2. exp mortality/ 3. follow-up studies/ 4. mo.fs. 5. prognos$.tw. 6. predict$.tw. 7. course.tw. 8. or/1-7</td>
</tr>
<tr>
<td><strong>Prognosis specificity filter</strong></td>
<td>1. prognosis/ 2. survival analysis/ 3. 1 or 2</td>
</tr>
<tr>
<td><strong>Aetiology or Harm sensitivity filter</strong></td>
<td>1. exp cohort studies/ 2. exp risk/ 3. (odds and ratio$).tw. 4. (relative and risk).tw. 5. (case and control$).tw. 6. or/1-5</td>
</tr>
<tr>
<td><strong>Aetiology or Harm specificity filter</strong></td>
<td>1. case-control studies/ 2. cohort studies/ 3. 1 or 2</td>
</tr>
</tbody>
</table>
5.9 SOURCES OF ASSISTANCE

Identifying the primary studies, systematic reviews and existing Guidelines from NZ and elsewhere is fundamental to the production of an evidence-based guideline. Specialist skills in identifying and appraising the literature are available from the following sources.

1 MEDICAL LIBRARIANS

Medical libraries employ specialist staff who have the expertise to provide assistance with the design and execution of electronic searches and the use of bibliographic database software. Medical librarians will also be able to assist guideline development teams to obtain hard copies of material through interlibrary loans.

2 NZ HEALTH TECHNOLOGY ASSESSMENT

(Clearing House for Health Outcomes and Health Technology Assessment in New Zealand)
e-mail nzhta@chmeds.ac.nz
Website http://nzhta.chmeds.ac.nz

Health Technology Assessment (NZHTA) has been established at the Department of Public Health and General Practice within the Christchurch School of Medicine. The main function of NZHTA is to identify effective health care interventions and technologies and thereby facilitate evidence-based policy making and purchasing by the New Zealand funders of health and disability services.

NZHTA Services

- **Quick reference enquiries/ advisory service** Free but limited to quick assistance and advice on sources and strategies or a quick search of in-house resources. Electronic information from level one searches is also available free of charge. Note that this includes only the search documentation (sources and strategies) and the references from bibliographic databases. It does not include full text reviews or other information located in the course of the search.

- **Level One**: comprehensive literature Search (using COSI search protocol) of multiple bibliographic databases, library catalogues, selected grey literature sources (including internet sites), work by other HTA agencies, and in house resources. Does not include handsearching or contacting experts. Information is collated into an information package but is not summarized or evaluated. Does not include retrieval of documents.

- **Level Two**: A comprehensive search as above plus obtaining of relevant documents and a critical appraisal of the evidence by NZHTA staff and a consultant expert. Output is a report with an appraisal of the evidence, with conclusions, recommendations.
THE COCHRANE LIBRARY

All of the databases that comprise the Cochrane library require a subscription - either CDROM or Internet version. Website: http://www.cochrane.org

Abstracts of completed reviews are available free on internet and indexed in Medline

There are 2 Cochrane Groups based in NZ

Cochrane Menstrual Disorders and Subfertility Group
For more information please contact
Michelle Proctor
Review Group Co-ordinator, Cochrane Menstrual Disorders & Subfertility Group
Division of Obstetrics & Gynaecology
University of Auckland
National Women’s Hospital
PO Box 92189
Auckland 1003
New Zealand

Telephone: +64 9 630 9943 ext. 3250
Fax: +64 9 630 9858
Email: m.proctor@auckland.ac.nz

Cochrane Musculo-Skeletal Injuries Group
For more information please contact
Leeann Morton
Review Group Co-ordinator, Cochrane Musculo-Skeletal Injuries Group
Dunedin School of Medicine
University of Otago
PO Box 913
Dunedin
NEW ZEALAND

Telephone: +64 3 474 0999 ext 5085
Fax: +64 3 474 7881
Email: cmsig@stonebow.otago.ac.nz

• **Level Three**: As for Level Two above, plus a critical evaluation of evidence by an expert panel of three or more consultants.

• All levels are fee for service unless for NZ Ministry of Health. Other work is scheduled around MoH priorities.

• Copies of all NZHTA reports are available to download free from NZHTA website (http://nzhta.chmeds.ac.nz)
  Abstracts of completed reviews available free on internet and indexed in Medline
EFFECTIVE PRACTICE INSTITUTE (EPI)

Based at the Division of Community Health, University of Auckland, EPI is directed by Prof Rod Jackson, and was set up initially to develop teaching programmes in evidence-based practice. EPI also initiates and develops new guidelines and offers an advisory service for other guideline development teams. EPI staff are either health professionals or individuals who have skills in biostatistics, epidemiology, health economics, health outcomes, health information, health management and policy, and consumer representation.

For more information please contact
Manager
Effective Practice Institute
Division of Community Health
Private Bag 92 019
Auckland, NZ

Telephone +64 9 373 7599 ext 2363
Fax +64 9 373 7494
Email epi-manager@auckland.ac.nz
THE ROLE OF QUALITATIVE RESEARCH IN THE DEVELOPMENT OF CLINICAL PRACTICE GUIDELINES

Aims of Chapter:

- To describe qualitative research methodologies
- To define the role of qualitative research in clinical practice guideline development

Qualitative research answers questions about meaning and how people feel about and experience situations. The focus of quantitative research is the frequency or the occurrence of events whereas qualitative research aims to provide insights into the meaning of the events. Qualitative research aims to understand how people interpret health and disease and make sense of their health experiences.

6.2 QUALITATIVE RESEARCH METHODS

There are 3 types of qualitative studies:\(^{20}\)

*Phenomenology* which is concerned with describing the human or “lived” experience using the subjective or first-person experience as a source of knowledge. Examples are a study to describe parents’ day-to-day experiences of managing their child’s chronic, life threatening and progressive illness. The parents were able to report the challenges of their daily lives and in doing so could define adversity and how to manage it.

*Grounded theory studies* are designed to develop theory grounded in the real world of participants. A grounded theory study would attempt to explain process or disease. An example is a study of women recovering from depression. Each woman was interviewed for 90 minutes and transcripts of the interviews were analysed. A theory emerged that explained a “six stage process of redefining myself” that the women went through as they recovered from depression.

*Ethnographic studies* seek to learn about how people interpret their experience and adapt their behaviour within the context of their own culturally defined environment. For example, risk taking behaviour amongst adolescents.
6.3 DIFFERENCES BETWEEN QUALITATIVE AND QUANTITATIVE RESEARCH

There are major methodological differences between qualitative and quantitative research in terms of the type of question asked, sample selection, data collection and analysis and reporting the results. Qualitative study questions are designed to enquire about how people feel or experience certain situations. The number of people studied in qualitative studies is generally much lower as the focus is on themes rather than effective treatments and adverse events. Participants are often chosen on the basis of their knowledge or experience of the content area (known as an iterative approach) whereas in quantitative research samples are often chosen to be representative of the population, generally by using a random sampling method. The main methods to collect qualitative data are case studies, participant observation, in depth interviews and focus groups. Whereas the unit of analysis in quantitative studies is primarily numbers, the unit of analysis in qualitative studies is a thought, a concept or a theme. Study findings are usually presented in the form of detailed narratives that describe these themes and attempt to analyse how the themes are interrelated. Discourse analysis is another way to analyse the data.

6.4 WHY INCLUDE QUALITATIVE DATA IN EVIDENCE BASED GUIDELINES?

It is recognized that different research methodologies provide different types of information. There are limits to the usefulness of quantitative data and other kinds of research will be complementary to guidelines. There are some instances where numerical data may be misleading, reductionist and irrelevant to the real issues. For example, the averaging effect of the statistical analyses of clinical trials and systematic reviews means that there is a tendency to treat a patient as if they were “average” or will have an “average response” to treatment. It is often felt that there is little place to accommodate variability in either clinical outcomes or patients preferences within the evidence based health care model. A qualitative approach to complement the trial data by collecting information on patient preferences and the values that patients place on the outcomes, would perhaps help get over the averaging problem of statistics. Qualitative methods should also help bridge the gap between scientific evidence and clinical practice by informing on issues around acceptability and implementation of the recommendations.

There are three kinds of qualitative research data:

- In depth, open ended interviews with individuals or groups
- Direct observation and description of people’s activities, behaviours, actions and interactions
- Written data, usually excerpts, quotations or entire passages from records, diaries, official records, or open ended responses to questionnaires
6.5 USING QUALITATIVE DATA IN GUIDELINES

1. Using consumer oriented outcomes that matter.
   As established in Chapters 1, 2 and 3, guideline recommendations seek to focus on outcomes that matter to consumers and to avoid using surrogate or intermediate outcomes such as blood tests. The most obvious and appropriate way of identifying outcomes relevant to consumers is by direct contact either by interview or by focus groups. This may be particularly relevant to the topic identification and suitability screen questions in establishing what is meaningful.

2. Understanding health behaviour in an everyday context.
   The environment of many clinical trials is artificial and may not reflect everyday clinical practice or the every day experience of consumers seeking health care. Qualitative research that seeks to understand how consumers and clinicians interact in the health care sector and within their own environment should strengthen the development of recommendations that are applicable, generalisable and can be implemented.

3. Implementation
   In order to establish successful strategies/identify barriers between consumers and clinicians, it may be necessary to conduct focus groups or interviews with both consumers and clinicians.

6.6 HOW TO EVALUATE QUALITATIVE DATA

It is suggested that the following questions are considered in evaluating qualitative research:

- Is a qualitative approach appropriate to stated purpose?
- How were the setting and subjects selected?
- Is the researcher’s perspective clearly stated?
- Are methods of data collection clearly described?
- Are methods of data analysis systematic - quality control described?
- Are the results credible/based on the data?
- Are the conclusions “grounded in evidence”
- How generalisable is the information to other consumers?
6.7 AN EXAMPLE OF QUALITATIVE RESEARCH

Although there are many asthma guidelines available and a considerable body of evidence about the effective management of asthma, questions remain about the uptake of asthma guidelines and about the patient's adherence to therapy.

A qualitative study provided useful insights into managing asthma. Half of asthmatics interviewed did not see themselves as asthma sufferers and interpreted their “bad chests” as an acute and temporary problem, better treated with acute medications rather than daily prophylactic medications.

The acceptance of daily medications was seen as “stigmatising” and asthma patients would rather avoid triggers such as physical exertion than take daily medications. The need for people with asthma to integrate symptoms and management with both the practicalities of everyday life and the psychological “self” over time was acknowledged. Other qualitative studies provided further useful information. For example, many doctors make the assumption that patients are seeking medications whereas patients are concerned about becoming physically and psychologically dependent on bronchodilators and have deep seated concerns about the long term effects of inhaled corticosteroids.23,24
AIM OF CHAPTER:

- To describe the steps in assessing the evidence and developing graded recommendations to assist practice

Guideline recommendations need to be based on the best available evidence. There should be explicit links between the strength of the available evidence and the grade of the recommendation.

Grading is a 2 tier process:
- firstly based on an (objective) assessment of the design and quality of each individual study (quality scores) and
- secondly on a judgment (which may be more subjective) on the consistency, relevance and applicability of the whole body of evidence to the questions the guideline seeks to answer (graded recommendation).

The evaluation system that is suggested below incorporates several approaches. There are three steps to undertake in a rigorous evaluation process (see Figure 7.1)

1. Assessing evidence relevant to guideline questions
   - Determination of study design for each study (e.g. randomized controlled trial, cohort study, systematic review etc)
   - Critical appraisal of individual studies using the relevant checklist for the particular study design

2. Assigning quality scores (+, -) for each study and developing evidence tables of the total body of evidence

3. Developing graded recommendations from the body of evidence based on volume, consistency, clinical relevance and applicability.
FIGURE 7.1 STEPS FOR ASSESSING AND APPLYING SCIENTIFIC EVIDENCE

RESEARCH EVIDENCE FROM SYSTEMATIC REVIEW OF THE LITERATURE

STEP ONE: DETERMINE STUDY DESIGN AND ASSESS EVIDENCE RELEVANT TO GUIDELINE QUESTION
(critical appraisal with GATE* checklist)

STEP TWO: ASSIGN QUALITY SCORES (+, ∅ or -) AND DEVELOP EVIDENCE TABLES

STEP THREE: DEVELOP GRADED RECOMMENDATIONS
(Using considered judgment form)

CONSIDER:
- Volume of evidence
- Consistency
- Applicability
- Clinical impact
7.1 Step One: Assessing the evidence

**Critical appraisal**

Assessment of the evidence is essentially critical appraisal, a process of critically looking at clinical research studies and asking questions.

Each study that has been identified from the search strategy should be appraised in a systematic way to identify potential problems such as bias within the studies. This can be a huge undertaking and may require considerable time and resources. If good quality systematic reviews are available then it may not be necessary to critically appraise the individual studies making up the systematic review.

A full description of the process of critical appraisal used in this handbook is contained in Appendix 2 (GATE = Generic Appraisal Tool for Epidemiology, Appraisal Module).

For each primary research study or systematic review, regardless of design, the study can be divided into parts: the Participants, the Exposure group (e.g. an intervention), the Comparison group, the Outcomes and the Timeframe. The acronym “PECOT” is used to describe the 5 components of all studies (Participants, Exposure, Comparison, Outcome, Time). This is summarised in Figure 7.2 and Table 7.2.

---

**FIGURE 7.2 THE 5-PART PECOT DIAGRAM: THE GENERIC DESIGN OF ALL EPIDEMIOLOGICAL STUDIES**

![PECOT Diagram]

- **Study Population**
  - Source pop.
  - Randomised

- **Exposure (intervention)**
  - $D_E$
  - $D_C$

- **Comparison (control)**
  - $N_E$
  - $N_C$

- **Outcome**
  - +
  - -

**Time**

$D_E = \text{Denominator (D) for exposure (intervention) group, } D_C = D \text{ for comparison (control) group}$

$N_E = \text{Numerator (N) for exposure group, } N_C = N \text{ for comparison group}$
Participants: Consider the appropriateness of the inclusion/exclusion criteria, deciding if the population from which the patients were selected was relevant and recognisable and whether or not participants are representative of that population. These questions apply to all study designs.

Exposure (to the intervention or risk factor). The questions on exposure and comparison differ slightly for each study design. For studies of therapy, the focus is on how were patients assigned to each group, what was the process of randomisation, did blinding occur, was adherence to therapy measured? For diagnostic studies, the exposure group is the participants who have the disease or condition (i.e. reference standard positive) and the comparison group is those without the disease/condition (i.e. reference standard negative). The questions include: “was the reference standard applied regardless, was the independent test valid and was assessment blind?”

Comparison: There is always a comparison group in any study although it is sometimes implicit. Some therapeutic studies compare an intervention with usual therapy while others use a placebo or no intervention as a comparison.

Outcomes: The questions relate to the appropriateness of the outcomes, the reliability of the measurement and whether the outcomes were measured by investigators who were unaware of the exposure. Measurement of the outcomes is either dichotomous or continuous. Typical dichotomous data are the proportion of people with myocardial infarction (MI) requiring rehospitalisation or dying from MI in a specified time frame. Continuous outcomes are measured on a continuous scale eg blood pressure, number of days before returning to work or quantity of blood required for transfusion.

Timeframe for outcomes: Studies either measure the incidence of outcomes (events per number of participants per unit of time) or the prevalence of outcomes (events or states per number of participants). Many studies include measures of both prevalence and incidence. It is important to document the timeframe of each study as the results may be very different if the timeframe varies. Questions relate to the length of follow-up and whether it is adequate.

Determination of study design

In order to undertake critical appraisal, one needs to identify the type of study design. An overview of the types of studies can be found on Pg 2 of the GATE appraisal module in Appendix 2. The optimal study design varies according to the clinical question being asked; eg. a therapy question is best answered by an RCT or systematic review of RCT’s, but this type of study design is not appropriate for a diagnostic accuracy question (see Table 7.1). Most questions that are relevant to guideline development are therapy and diagnostic test questions, although questions of prognosis and risk factors also arise.

Once the study design has been determined, the assessor then is able to choose the appropriate checklist for the critical appraisal. Different study designs require slightly different types of assessment when quality is being evaluated (see Table 7.2) so separate checklists are provided for systematic reviews and meta-analyses, randomized controlled trials (RCT’s), cohort studies, case-control studies and diagnostic studies (see Appendix 2). It is unlikely that other types of study design will need to be assessed as evidence in the development of a guideline.
The checklists contain 3 main sections:

1. study validity (steps made to minimise bias)
2. study results (size of effect and precision)
3. study relevance (containing applicability/generalisability)

### TABLE 7.1: DIFFERENT QUESTIONS NEED DIFFERENT STUDY DESIGNS

<table>
<thead>
<tr>
<th>Clinical Questions</th>
<th>Most Appropriate Study Design</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>Cross Sectional Cohort</td>
<td>Sensitivity, specificity, Likelihood ratios, Number needed to test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient expected event rate</td>
</tr>
<tr>
<td>Harm Therapy</td>
<td>RCT or Cohort or Case Control</td>
<td>Number needed to harm</td>
</tr>
<tr>
<td></td>
<td>Systematic reviews or Randomised controlled trials</td>
<td>Absolute Risk Reduction, Number needed to treat</td>
</tr>
</tbody>
</table>
### TABLE 7.2: SUMMARISING THE QUESTIONS USED FOR CRITICAL APPRAISAL FOR DIFFERENT STUDY DESIGNS (SUMMARY OF GATE WORKSHEETS)

<table>
<thead>
<tr>
<th>Clinical question and optimal study design</th>
<th>Validity</th>
<th>Study results (Magnitude and Precision)</th>
<th>Study relevance (Application)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Who were the study participants (What were the key inclusion and exclusion criteria)?</td>
<td>Was there a clearly defined independent &amp; valid reference standard?</td>
<td>How strong was the association between reference and test results?</td>
<td>Study findings applicable in typical settings</td>
</tr>
<tr>
<td>Were selection criteria appropriate given study question?</td>
<td>Was the reference standard applied regardless of the test result?</td>
<td>Were likelihood ratios for the test results presented or data necessary for their calculations provided?</td>
<td>Is it feasible to apply the test in usual settings?</td>
</tr>
<tr>
<td>Appropriate spectrum of participants?</td>
<td>Was the reference standard assessed blind to the test result?</td>
<td>Did different levels of the test have strong associations with the reference standard?</td>
<td>Are the benefits of using the test worth the potential harm and costs?</td>
</tr>
<tr>
<td>Source population relevant &amp; recognisable?</td>
<td>Was the test validated in the second independent group of participants?</td>
<td>How precise were the estimates of association? (were CIs or p values given for LRs, sensitivity/specificity)?</td>
<td>Were all important outcomes considered?</td>
</tr>
<tr>
<td>Participants representative of source</td>
<td></td>
<td>If no significant effects were detected did the study have sufficient power to detect important associations?</td>
<td>Will the results change management?</td>
</tr>
<tr>
<td><strong>Exposure/comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes/Time</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validity</td>
<td>Study results (Magnitude and Precision)</td>
<td>Study relevance (Application)</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td><strong>Exposure/comparisons</strong></td>
<td><strong>Outcomes/Time</strong></td>
<td><strong>How strong was the association between exposure and outcomes?</strong></td>
</tr>
<tr>
<td>Inclusion/exclusion criteria appropriate?</td>
<td>Subgroups similar except exposure (ie: prognostic factors)?</td>
<td>Assessment blinded to exposure</td>
<td>How large were the intervention effects?</td>
</tr>
<tr>
<td>Source population relevant &amp; recognisable?</td>
<td>If not, were differences stratified/adjusted in analyses?</td>
<td>Completeness of follow up sufficient</td>
<td>What measures of occurrence were used?</td>
</tr>
<tr>
<td>Participants representative of source?</td>
<td>Clearly defined exp/comp subgroups?</td>
<td>Length of follow up sufficient</td>
<td>What measures of effect/association were used? (RR, RD, etc)</td>
</tr>
<tr>
<td><strong>Harm/causatory benefit cohort, case-control studies</strong></td>
<td><strong>Assessment blinded to exposure</strong></td>
<td></td>
<td>Could NNTs be calculated for harms as well as benefits?</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria appropriate?</td>
<td>Completeness of follow up sufficient?</td>
<td></td>
<td>Was there a dose-response relationship?</td>
</tr>
<tr>
<td>Source population relevant and recognisable?</td>
<td>Length of follow up sufficient?</td>
<td></td>
<td>How precise were the estimates of effect/association?</td>
</tr>
<tr>
<td>Participants representative of source?</td>
<td>Temporal association sufficient?</td>
<td></td>
<td>Were CIs or p-values given?</td>
</tr>
<tr>
<td><strong>Study results</strong> (Magnitude and Precision)</td>
<td><strong>Study relevance</strong> (Application)</td>
<td></td>
<td>If significant effects were not detected, did study have sufficient power?</td>
</tr>
<tr>
<td><strong>Prognosis:</strong></td>
<td><strong>How large were the intervention effects?</strong></td>
<td><strong>Results applicable in typical settings?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cohort Study</strong></td>
<td></td>
<td></td>
<td>Were all important outcomes considered?</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria appropriate?</td>
<td><strong>What measures of occurrence were used?</strong></td>
<td><strong>Harms vs benefits</strong></td>
<td></td>
</tr>
<tr>
<td>Source population relevant &amp; recognisable?</td>
<td><strong>What measures of effect/association were used?</strong> (RR, RD, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>Exposure/comparisons</td>
<td>Outcomes/Time</td>
<td>Study results (Magnitude and Precision)</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>- Who were the study participants?</td>
<td>- What interventions were evaluated?</td>
<td>- What outcome measures were used?</td>
<td>- What measures of effect were used? (RR, RD, etc)</td>
</tr>
<tr>
<td>- Inclusion/exclusion criteria appropriate?</td>
<td>- How many patients were in each group?</td>
<td>- Were benefits and harms considered?</td>
<td>- What was the scale and direction of the effect?</td>
</tr>
<tr>
<td>- Source population relevant and recognisable?</td>
<td>- Assignment randomised?</td>
<td>- Were the outcomes appropriate?</td>
<td>- How precise were the estimates of intervention effect? (CIs etc)</td>
</tr>
<tr>
<td></td>
<td>- Randomisation successful?</td>
<td>- Assessment blinded to exposure?</td>
<td>- If no significant effects were detected, did the study have sufficient power?</td>
</tr>
<tr>
<td></td>
<td>- Randomisation process concealed?</td>
<td>- Completeness of follow up sufficient?</td>
<td>- Could NNTs be calculated for harms as well as benefits?</td>
</tr>
<tr>
<td></td>
<td>- Patients &amp; researchers blinded?</td>
<td>- Length of follow up sufficient?</td>
<td>- If multicentred RCTs were the results homogenous between sites?</td>
</tr>
</tbody>
</table>
7.2 STEP TWO: QUALITY RATING AND DEVELOPMENT OF EVIDENCE TABLES

Once the study design has been determined and the study critically appraised using Sections 1-3 of the relevant checklist, a quality score can be determined for each section. There are 3 quality categories that can be assigned based on the extent to which the study design has met the criteria:

- plus (+) (strong study where all or most of the validity criteria are met – ie “in the shaded boxes)
- minus (-) (weak study where very few of the validity criteria are met and there is a high risk of bias) or
- neutral (∅) (study where not all of the criteria are met but the results of the study are not likely to be affected)

on the basis of answers to questions contained in the relevant sections of the Summary Checklist.

Different study designs may have different types of bias and characteristics that are important for quality, so the assessor will need to answer these questions in relation to the study being evaluated (see Table 7.2 for a summary of the types of quality features important for each study design).

This quality assessment is intended to identify exceptionally strong or exceptionally weak studies in terms of quality. A study that does not meet all the quality criteria in 1 and where the potential omissions in the study are not likely to invalidate the results (ie cannot be designated as + or -), should be designated as neutral (∅).

It is difficult to give explicit criteria for the decision to allocate +, - or ∅ to a study. Quality scores should be assigned by 2 or more independent assessors and disagreement resolved by consensus. When the quality of a study is the subject of significant disagreement, the entire Guideline development team may need to be involved and the final quality score assigned after debate and discussion. The rationale for this time consuming procedure is that the process of scoring is unavoidably subjective and bias is minimised by adopting a consensus process.

Steps One and Two of the assessment of evidence are applied to individual studies that have been independently selected as being relevant to the questions that are being answered in the guideline. Both the study design and the quality rating for the 3 sections are entered in the evidence table that is part of the body of evidence that is considered in the forming of recommendations for the guideline (see evidence table template at end of Appendix 2).
7.3  STEP THREE: DEVELOP GRADED RECOMMENDATIONS FROM THE BODY OF EVIDENCE

Recommendations are developed jointly by the guideline development team based on consideration of the whole body of evidence that is summarised in the evidence tables. The body of evidence is the sum of the evidence of all the individual studies and the quality ratings of each study. Their separate relevance to the guideline question(s) has also been considered (Section 3 of the checklist). Before recommendations can be prepared, the body of evidence relating to the guideline question(s) as a whole needs to be assessed. All of the evidence needs to be considered giving greater weight to studies of higher quality.

This is perhaps the most difficult part of the process and requires the exercise of judgment based on experience as well as knowledge of the evidence and the methods used to generate it. It is surprisingly rare for the evidence to show clearly and unambiguously what course of action should be recommended for any given question. The quality of all the evidence needs to be linked with other aspects that require “considered judgment”. Some of the most important issues that should be considered in the formation of recommendations are as follows:

- **Quantity and consistency**
  Development groups need to consider both the number of studies and the number of participants studied in the development of recommendations. If the findings across all studies are reasonably consistent, then the group will have greater confidence in its conclusions. If not, greater weight should be given to designs that are most appropriate for the study question and studies with minimal bias.

- **Applicability**
  Recommendations should be based on the best evidence directly applicable to a New Zealand setting. In extrapolating findings from overseas trials to a NZ setting, the guideline development team needs to be aware of unique cultural factors, provider or organizational factors and characteristics of the study population.

- **Clinical impact**
  An important consideration is whether the potential benefit from an intervention is sufficiently great to justify a recommendation that it be used in practice. This will depend on: the size of the effect compared with no intervention or other management, the duration of treatment required to achieve the effect and the balance of risk and benefit (see Figure 7.3). Other factors such as cost may have to be taken into consideration but should impact more on the implementation strategies needed than on how the evidence is expressed.

The guideline development team considers these issues and completes a Considered Judgement Form (see Appendix 3 adapted from SIGN\textsuperscript{25}). A summary statement based on the synthesis of the entire body of evidence is also recorded on this form together with the separate quality scores of evidence recorded in the right hand column.
Making recommendations involves considerations that extend beyond scientific evidence and require judgment. Increasing the role of subjective judgment in forming guideline recommendations necessarily increases also the risk of re-introducing bias into the process. However, this is minimized if decisions are made based on a consensus of the entire development group. The Considered Judgement form provides a focus for discussion of the issues relating to the forming of recommendations. When consensus is reached, comments can be recorded directly on the Considered Judgement form. The recommendation can be developed from these statements and graded according to its strength (see Table 7.5).

**FIGURE 7.3 RISK BENEFIT APPROACH TO APPLYING THE EVIDENCE TO INDIVIDUAL PATIENTS (BASED ON MODEL OF LUBSEN AND TIJSSEN 1989)**

![Risk Benefit Approach Diagram](image)

The strength of the evidence that is being considered in this grading process is linked to the question that is being answered. A question on therapy is best answered by a well-designed systematic review or randomised controlled trial, a diagnostic question by a cross sectional study and a prognostic question by a cohort study. Case series studies or expert opinion do not provide strong evidence for answering guideline questions but may be the only evidence available.
Thus, the grade of the recommendation is based on consideration of a number of factors:

- the design and quality of individual studies that have been identified to provide answers to the question posed – these are converted into a summary evidence statement reflecting the body of evidence that is recorded on the Considered Judgment Form together with the different quality scores.
- quantity, consistency, applicability and clinical impact of the body of evidence that is applicable the guideline question that is also recorded on the Considered Judgment Form.
- the consensus of the guideline development team.

A recommendation can be made that an action is beneficial, harmful, has no effect or the harms and benefits of the action are equally balanced. The grade of the recommendation is determined by the strength of the evidence that supports it. For example, although Danazol is very effective at reducing heavy menstrual bleeding, it is not recommended for the treatment of heavy menstrual bleeding because it causes unacceptable side effects in women (balance of benefit and harm).

These recommendations form a series of guiding statements that propose a course of action and should be reproduced together as a summary for ease of implementation.

The Key Elements in the Considered Judgement Form  (see Appendix 3)

1. Volume of evidence
2. Applicability and generalisability
3. Consistency
4. Clinical impact
5. Other factors
6. Evidence statement
7. Recommendation
<table>
<thead>
<tr>
<th>Grade</th>
<th>Details</th>
</tr>
</thead>
</table>
| A     | **The recommendation (course of action) is supported by good evidence**  
The evidence consists of results from studies of strong design for answering the question addressed. |
| B     | **The recommendation (course of action) is supported by fair evidence**  
The evidence consists of results from studies of strong design for answering the question addressed but there is some uncertainty attached to the conclusion either because of inconsistencies among the results from the studies or because of minor flaws; or the evidence consists of results from weaker study designs for the question addressed but the results have been confirmed in separate studies and are reasonably consistent. There is fair evidence that the benefits of the course of action being proposed outweigh the harms. |
| C     | **The recommendation (course of action) is supported by expert opinion only**  
For some outcomes, trials or studies cannot be or have not been performed and practice is informed only by expert opinion. |
| I     | **No recommendation can be made because the evidence is insufficient**  
Evidence for a course of action is lacking, of poor quality or conflicting and the balance of benefits and harms cannot be determined. |
8.1 WHAT IS A BALANCE SHEET

A balance sheet is a formal itemisation of the major costs and health benefits of a healthcare programme or intervention. It distinguishes an ‘explicit evidence-based guideline’ from an ‘evidence based guideline’. A balance sheet is not a full economic analysis because it does not fully account for all the economic costs or consumer benefits such as gains in quality of life. It also cannot take into account changes that may occur over time, other than those that result directly from the intervention under consideration. In some cases a background decision analytic model or a Markov model (e.g. for an illness with recurring health states) may be required to support the development of a balance sheet, although it is preferable to avoid this step if possible because of its potential complexity and expense. However, it may be the only way to obtain information about the economic outcomes of current or proposed clinical practice. Not all guidelines contain a balance sheet, especially if there is insufficient or inadequate information, as too many assumptions may be necessary.

A balance sheet simply itemises and sums the economic costs and benefits of the interventions under consideration in the Guideline, under the conditions of current practice as well as the practice under consideration (such as a screening programme or a therapeutic intervention). Some schools of thought would include health benefits such as mortality gains in the balance sheet, however adding these levels of complexity turns a conceptually (but not necessarily) simple exercise into a full economic evaluation, which is not usually necessary. It also can put the balance sheet out of reach of most practitioners and well beyond the limited budget that is available for guideline development. The balance sheet could be considered as a first step in a full economic evaluation.

A full economic analysis would attempt to take into account outcomes that are important but not quantifiable, such as intangible aspects of quality of life and indirect benefits such as avoidance of lost productivity. However, this is not the role of a balance sheet. Further
A good balance sheet demonstrates the value of the clinical guideline. It determines what things cost and who pays for what. It informs purchasers, prescribers and consumers about potential changes in the costs of screening, diagnosis and interventions. It can also provide a thumbnail sketch of potential economic outcomes such as hospitalisations avoided by delaying or preventing morbid events. Finally, it also tests the practicality of the guideline: an intervention that proves to be very costly may well be impractical and a poor use of valuable resources.

It is desirable to provide a balance sheet along with a clinical practice guideline because it summarises the financial aspects of both benefits and harms of an intervention to purchasers, providers and consumers and their caregivers. It also highlights the financial hurdles to implementation.
8.4 PERSPECTIVE

Because a New Zealand clinical practice guideline is set in the context of New Zealand healthcare delivery, most of which is publicly funded, a balance sheet will usually take a healthcare perspective rather than a full societal perspective. It is therefore likely to exclude indirect costs such as lost productivity of the family unit; loss of taxation revenue; welfare benefits etc. It may also include costs to consumers such as co-payments for GP consultations, but it is unlikely to account for productivity gains and losses in families or other economic units. It does not usually take the consumer’s perspective, although it can/should partition costs into those incurred by the public system or the consumer (e.g. co-payments on GP consultations and pharmaceuticals; payment for private surgery).

8.5 WHO DEVELOPS THE BALANCE SHEET?

The primary requirement is an ability to draw the key elements of the clinical discipline, its epidemiology and its economics into a quantitative, population-based framework. In order to ensure that the balance sheet fully reflects evidence-based clinical practice, the individual who takes primary responsibility for the balance sheet should be a full member of the guideline development team from the beginning, not a late addition. The balance sheet can be developed in parallel with other aspects of the guideline, but most of the work will come towards the end of the guidelines process.

A balance sheet does not necessarily require a health economist. The requisite skills are: an analytical mind; ability to program a spreadsheet; familiarity with population sciences including basic statistical methods; and the ability to communicate with clinical leaders who are developing the guideline.

8.6 PRIOR QUESTIONS

Before a decision is taken to develop a balance sheet, a number of questions should be asked:

- What is the objective of the balance sheet?
- From whose perspective is the balance sheet being prepared? (Society? The healthcare system? Public healthcare funders only?).
- Is there a time horizon to consider? (1 year? 5 years? A consumer’s lifetime?)
- What incidence and prevalence data are available?
- Are any ‘hard’ outcomes available or is a model required?
- What data are available about quality-of-life outcomes?
- What data are available about quality clinical management?
- What data are available about the major cost items?
- How practical is it to undertake a balance sheet? Adequate time? Budget?
- Will it provide good value or is it an optional extra?
8.7 DATA SOURCES

The data available to construct a guideline will come from a number of sources. The predicted health benefits and harms should come from the systematic review of the literature that has already taken place (see chapters 5 and 7). Where information is not available from well-designed studies then lower levels of published evidence and expert opinion may be included. Information on disease incidence and prevalence may be obtained from the international literature supplemented by local clinical trials, observational studies, databases, patient registries and surveys.

A clinical practice guideline should use local costs rather than international costs. Cost and volumes data should come from the internal data process (chapter 4). There may be multiple sources and the costs will depend on the perspective, i.e. public / private; societal / healthcare.

Some outcomes data may require economic modelling based on incidence, prevalence, and cost: e.g. long term outcomes of screening or vaccination programmes; ‘hard’ clinical outcomes of interventions for which only surrogate end-points are available. Do not underestimate the time taken to track down costs from diverse sources, some of which may be reluctant to release information!

Table 8.2. gives a few NZ sources of cost information. The list is not exclusive, and the relevant sources depend on the project.

8.8 FEASIBILITY OF BALANCE SHEET DEVELOPMENT

In some cases it may not be practical to develop a balance sheet if the data sources are inadequate or key data are missing. If the outcomes of an intervention are complex, or if events recur, economic modelling may be more appropriate and it may be essential to avoid giving a misleading impression of economic outcomes. If it is necessary to make very broad assumptions that cannot be validated or checked, then it may not be worthwhile pursuing a balance sheet approach. Furthermore the development of the balance sheet should be included in the planning of the guideline, as it does take considerable time and expertise.
<table>
<thead>
<tr>
<th>Source</th>
<th>Information provided</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient registries</td>
<td>Individual health outcomes</td>
<td>Available in limited therapeutic areas: e.g. chronic renal failure, diabetes. Can require work to clean and extract relevant information.</td>
</tr>
<tr>
<td>Local publications (e.g. NZMJ)</td>
<td>NZ epidemiology</td>
<td>Use international data if necessary and state caveats</td>
</tr>
<tr>
<td>Ministry of Health (e.g. &quot;Purchasing for your Health&quot;)</td>
<td>Mortality and morbidity; surgical costs and volumes by DRG; some GP and pharmaceuticals data; limited burden-of-disease information; health benefits expenditures; rest home and disability expenditure</td>
<td>Broad picture material; much of this information is too coarse for use in a Guideline</td>
</tr>
<tr>
<td>Hospital cost databases</td>
<td>Funding by clinical service unit</td>
<td>Available through clinical directors</td>
</tr>
<tr>
<td>RNZCGP database (c/o University of Otago)</td>
<td>GP consultations by indication</td>
<td>Useful background but not fine grained enough for many applications</td>
</tr>
<tr>
<td>NZ Health Information Services</td>
<td>Population statistics including life expectancy, mortality and hospital based morbidity</td>
<td>Reasonably accessible via E-mail and quite useful</td>
</tr>
<tr>
<td>New Ethicals Catalogue</td>
<td>Drug regimens*</td>
<td>Community and hospital pharmaceuticals</td>
</tr>
<tr>
<td>Pharmaceutical Schedule</td>
<td>Drug costs</td>
<td>Mostly community pharmaceuticals; available as hard copy or on the PHARMAC website</td>
</tr>
<tr>
<td>PHARMAC</td>
<td>Drug volumes and expenditure</td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical manufacturers</td>
<td>Drug costs; sales volumes</td>
<td>Sales figures are usually confidential; pharmaceuticals may be used in several indicators</td>
</tr>
<tr>
<td>Statistics NZ</td>
<td>1996 census data*</td>
<td></td>
</tr>
<tr>
<td>Diagnostic laboratories</td>
<td>Reimbursement schedules for laboratory tests and radiology</td>
<td></td>
</tr>
<tr>
<td>Private hospitals and insurance companies</td>
<td>Reimbursement schedules for acute surgery</td>
<td></td>
</tr>
</tbody>
</table>

DRG = diagnosis-related group (based on the system used in Victoria, Australia); NZMJ = NZ Medical Journal; PHARMAC = Pharmaceutical Management Agency Ltd; RNZCGP = Royal NZ College of GPs

* Soon to be updated
8.9 PITFALLS

Balance sheets have several potential pitfalls:

- Important assumptions and controversies can get overlooked.
- Some of the cost data can be difficult or impossible to locate with any certainty.
- It can be difficult to validate.
- Charges are often used as proxy for costs; this fails to account for cross subsidisation of services or profits in private medical systems.
- Because it does not formally link costs to outcomes, it cannot be used to compare the value of different types of interventions.
- Inclusion of balance sheets in clinical practice guidelines has led to the criticism that guidelines provide a tool to help managers trim budgets.

Another difficulty is that potential cost offsets such as avoidance of surgery by utilisation of a novel intervention may not be realisable in practice, because of a 'silo' approach to purchasing. For example, the Ministry of Health may purchase a fixed volume of surgical operations each year, and resources are rationed indirectly by allowing a waiting list to develop and the private sector to take up some of the load. A new intervention that has the potential to avoid surgery may relieve pressure on a surgical waiting list but will have little or no effect on the annual cost of surgery in the public sector. Furthermore, it provides an incentive for surgeons in the private sector to encourage more patients to accept surgery that may not be essential, with resulting deterioration of efficiency in healthcare delivery.

Nevertheless, it is important to spell out the potential for cost avoidance in a balance sheet, if only to demonstrate the value of a new intervention and potentially lead to better ways of purchasing healthcare. A good balance sheet is a step in the direction of evidence-based purchasing. It aligns with government objectives to move towards purchasing population outcomes.

8.10 PROCESS OF DEVELOPMENT OF A BALANCE SHEET

1. DECIDE WHETHER IT IS WORTH DOING, AND WHY

- Which professional and managerial groups might be interested in the outcome?
- Is the available information of sufficient quality so that the result is informative, not misleading?
- How long is it likely to take?
- What resources are available to perform the task?
2. DECIDE WHETHER IT IS FEASIBLE

- Types of information adequate? Simple or complex?
- Does it require economic modelling? E.g. do the costs and benefits last beyond the time frame of the clinical source data? (E.g. antihypertensive agents; ‘statins’; vaccines); is information about current therapy adequate? (novel therapies are usually accompanied by good information on safety, efficacy and optimal use but older therapies may lack the high quality data that are necessary for a balance sheet)
- Are the data accessible? Are they clean? (E.g. cost databases well organised? Patient registries cleaned and anonymised? Pivotal data available? Data on comparator therapies available? Need further clinical input on patient management, from focus groups? Any data on referrals to high cost services?).

3. DEFINE THE TASK

- What is the perspective of the balance sheet?
- What types of information do you require to perform this task?

NZ incidence and prevalence of the disease or condition that is to be screened for, prevented or treated.

- Health outcomes (including adverse effects) of all relevant therapies:
  - Reduces symptoms? (Specify, any QOL data? utility data?)
  - Avoids or delays the need for other therapies? (e.g. older drugs)
  - Reduces disease progression?
  - Reduces disease burden on the State and other carer(s)
  - Adverse effects? (e.g. GI, CNS)

- Costs: of the illness (current therapies, non-medical care) new therapy (drug + process), rehabilitation
  - Direct medical: drug, hospital, institutional, home (personal health /disability; State/private budget?)
  - Indirect (Appropriate to the perspective? Relevant to the disease or condition? Available? Credible?)
  - Consumer and provider satisfaction: carer, GP

4. CONSTRUCT A COMPREHENSIVE SPREADSHEET ANALYSIS

It is useful to split it into 2 halves, with current practice on one side and recommended practice on the other. Distinguish between costs to the public and private sector including consumers.

5. OBTAIN PEER REVIEW

Clinical Directors, the Ministry of Health and colleagues in NZ and abroad can provide useful input at this stage. Full peer review is obtained when it is sent out as part of the draft guideline.
8.11 EXAMPLES OF BALANCE SHEET DEVELOPMENT

Based on availability and quality of information it may be possible to develop a comprehensive spreadsheet based balance sheet, or a partial balance sheet, or nothing at all. Efforts to develop a balance sheet can be thwarted by scanty information or alternatively by excessive complexity in the published outcomes studies and/or models. Table 8.2 describes the balance sheets in some recent NZ guidelines.

<table>
<thead>
<tr>
<th>Indication</th>
<th>Interventions</th>
<th>Feasibility of balance sheet</th>
<th>Economic model available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menorrhagia</td>
<td>Medical and surgical</td>
<td>Yes (published on NZGG website)</td>
<td>Partial only</td>
</tr>
<tr>
<td>Leg Ulcers</td>
<td>Compression bandaging</td>
<td>Yes, in progress</td>
<td>In progress</td>
</tr>
<tr>
<td>Dyspepsia</td>
<td>Medical and surgical</td>
<td>Too complex for a balance sheet; limited aspects may be considered</td>
<td>Yes (UK)</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>Acetylcholinesterase inhibitors</td>
<td>Inadequate information about long term cost implications of delayed institutionalisation</td>
<td>Yes (UK and US)</td>
</tr>
</tbody>
</table>

The balance sheet for implementation of the Guideline for the Management of Heavy Menstrual Bleeding is shown on the next page.

This balance sheet implies that substantial costs could be avoided by shifting towards more effective and better-tolerated medical therapies, and away from expensive surgery. Although this balance sheet is comprehensive (and it was time consuming to develop) it has been challenged on the grounds that surgical costs cannot be avoided by replacement with medical therapies (see section 8.9 above). However, the balance sheet still provides a good model that could guide evidence-based purchasing; in fact its main value lies in showing the clinical and economic advantages of uptake of better therapies.

The balance sheet is divided into 2 parts: the left side represents current practice and the right side represents a potential shift towards more cost-effective medical therapies such as a novel intra-uterine levonorgestrel-releasing device. The costs of GP and specialist consultations, diagnostic tests, medical and surgical therapies are shown, along with estimates of shifts in volume expenditures that could occur when the guideline is fully implemented.
AIMS OF CHAPTER:

- To make CPG more accessible and easy to use
- To describe the process of summarising the recommendations
- To describe the steps in algorithm formulation

The majority of readers and users of CPG’s will not have the time to read the whole document and it is essential that a summary of the guideline with recommendations is produced. Algorithms are also useful tools that can improve guideline utilisation.

9.1 THE RECOMMENDATIONS

Recommendations are a series of guiding statements that propose a course of action. The recommendations should progress through the questions prepared earlier in a logical manner following clinical care episodes. Each recommendation should advise a course of action, followed by an indication of the strength of the recommendation. This step cannot occur until the evidence has been graded and evidence tables compiled (Chapter 7). Simple statements of the evidence with the levels are not sufficient in most cases.

The development of the recommendations is perhaps the most difficult and least evidence-based part of the process as it is necessary to get the group to agree and usually this is achieved by consensus. The considered judgement form links the clinical questions, the evidence and the recommendation25 (see Appendix 3). The group needs to make a decision at the beginning of the process about how to resolve differences.

9.2 STEPS IN ALGORITHM PREPARATION

Algorithms are flow charts that have a clear beginning and end. In the middle there will be a number of linked process steps and decision points.

One convention that is popular is to have the start and end in round oblong boxes, the process steps in sharp boxes and decision points in diamonds. Linking parts of the algorithm to the evidence tables or recommendations may be helpful. (See Fig 9.1)
The first draft of the algorithm is usually undertaken by the team leader who presents it at one of the early meetings. This provides a framework for the team to consider and allows the team to see how the process is possible. The algorithm will evolve slowly by a process of using the evidence and recommendations and considering the process of care for consumers. Many drafts will be considered. Some testing with clinical scenarios is useful. The final version should reflect all the options of care recommended in the guideline.
9.3 WRITING INSTRUCTIONS FOR GROUP MEMBERS

Although one person (usually the convenor) usually has overall responsibility for preparing and editing the manuscript, many members of the group will contribute portions of the guideline. The convenor should discuss the layout of the document with the New Zealand Guideline Group. A suggested template for both the written and web version is given in Appendix 5. Writing in committee can be challenging and requires prior agreement about the consistent use of terminology, evidence tables and use of statistics. The writing style should be as consistent and succinct as possible, keeping in mind the broad readership of the guideline.

9.4 OTHER DECISION AIDS

Tables that simplify the risks and benefits or can be used to quantify the risk are also helpful.

**EXAMPLE: MEDICAL TREATMENTS FOR FIBROIDS**

<table>
<thead>
<tr>
<th>Rx</th>
<th>Level of evidence</th>
<th>Improving symptoms</th>
<th>Shrinking Fibroid</th>
<th>Maximum duration</th>
<th>Harmful side effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC pill (HMB only)</td>
<td>B</td>
<td>✓</td>
<td>X</td>
<td>Unlimited</td>
<td>Nausea, headache, breast tenderness</td>
</tr>
<tr>
<td>Danazol</td>
<td>B</td>
<td>Not studied</td>
<td>✓ 57%</td>
<td>6 months</td>
<td>Androgenic side effects</td>
</tr>
<tr>
<td>Gestrinone</td>
<td>A</td>
<td>Not studied</td>
<td>✓ 15-36%</td>
<td>6 months</td>
<td>Androgenic side effects</td>
</tr>
<tr>
<td>GnRHa</td>
<td>A</td>
<td>✓</td>
<td>✓</td>
<td>6 months</td>
<td>Androgenic side effects</td>
</tr>
<tr>
<td>LNG IUS</td>
<td>C</td>
<td>Not studied</td>
<td>Not studied</td>
<td>5 years</td>
<td>Irregular menses, perforation, expulsion</td>
</tr>
<tr>
<td>RU486</td>
<td>B</td>
<td>Not studied</td>
<td>✓ 49%</td>
<td>Not yet investigated</td>
<td>Not studied</td>
</tr>
</tbody>
</table>
DISSEMINATION AND IMPLEMENTATION OF CLINICAL PRACTICE GUIDELINES

AIMS OF CHAPTER:

- To describe dissemination and implementation strategies for CPGs that have been successful in medical practices
- To outline the steps that need to be undertaken for successful dissemination and implementation strategies

Nothing could be more frustrating to the guideline development team than developing a guideline that is then ignored by not being disseminated or implemented. This phase of the process may not necessarily be the responsibility of the guideline development team but any guideline development should include a plan of the steps and options for dissemination and implementation and a recommended timeframe. Unplanned clinical guideline implementation that has not been tailored to the various audiences to which the guideline applies usually results in unsuccessful guideline adoption. 33,34,35

There are three levels of implementation change that should be considered:

- The practitioner – patient level eg. changing clinician/patient behaviour and attitudes
- At the systems level eg. enabling clinicians to make changes easily by providing access to computer decision support systems
- At the policy level eg. by providing coverage decisions that enable access to health interventions

10.1 WHAT WORKS IN IMPLEMENTATION OF CLINICAL PRACTICE GUIDELINES?

There is no single answer to what is a successful implementation strategy although the limited research carried out suggests a range of approaches is more likely to succeed than a single approach. Ideally, the research literature should guide this phase of guideline development but methodological limitations of the research base mean that this is not necessarily possible. For example, problems of sample sizes, length of time required before data analysis can begin, resources issues, and different health systems often mean that the research can not be transferred or generalised to other settings.
However, some interventions have been shown to be consistently effective while others have variable, little or no effectiveness. These are summarised in the following table (Adapted from Bero).  

### TABLE 10.1 WHAT WORKS AND DOESN’T WORK IN GUIDELINE IMPLEMENTATION

<table>
<thead>
<tr>
<th>Consistently effective</th>
<th>Variably effective</th>
<th>Little or no effect</th>
<th>Unknown effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Educational outreach visits</td>
<td>- Audit and feedback</td>
<td>- Dissemination of educational materials alone</td>
<td>- Financial incentives</td>
</tr>
<tr>
<td>- Decision-support systems and other reminders</td>
<td>- Local opinion leaders</td>
<td>- Didactic educational meetings</td>
<td>- Administrative interventions</td>
</tr>
<tr>
<td>- Interactive educational meetings</td>
<td>- Local consensus processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Multifaceted interventions</td>
<td>- Patient-mediated interventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Mass media interventions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 10.1.1 IMPLEMENTATION STRATEGIES THAT HAVE BEEN SHOWN TO BE CONSISTENTLY EFFECTIVE

#### 10.1.1.1 Educational outreach visits

Outreach visits are also known as academic detailing. These are face-to-face visits by trained personnel to clinicians within their practices. This is the approach that the pharmaceutical industry has undertaken for many years and believes it to be worthwhile for influencing decisions about prescribing. Other groups have also used it such as in Sudden Infant Death Syndrome research to influence midwives about babies sleeping positions.
10.1.1.2 *Decision support systems and other reminders*

Decision support systems include anything manual or automated that prompt health professionals to perform a clinical action. Examples are reminders about screening, laboratory reports where results to note are highlighted, follow up appointment systems and stickers on charts. In particular, computerised decision support systems have led to improvements in doctors’ decision making on drug dosage, provision of preventive care and general clinical management of patients.\(^{37}\) The advantage of these systems is that they are fairly easy to implement and are available to clinicians at the time required (just in time reminders).

10.1.1.3 *Interactive educational meetings*

These involve the active participation of health professionals in workshops, small group discussions, problem based learning or a range of other approaches. They should not be passive learning environments.

10.1.1.4 *Multifaceted interventions*

Multifaceted approaches are more effective than single interventions. Combinations of audit and feedback, face to face educational outreach/visitation using opinion leaders, reminders, local consensus processes and marketing are more effective than a single approach as they will tap into different parts of the change process.\(^{38}\)

10.1.1.5 *Mass media campaigns*

Both planned and unplanned approaches can be successful in the mass media although the effect of health services utilisation may be modest.\(^{39}\) The mass media is a potential means of reaching all audiences at the same time.

10.1.2 **IMPLEMENTATION STRATEGIES THAT ARE VARIABLY EFFECTIVE**

10.1.2.1 *Audit and feedback*

Audit and feedback is a process in which the clinical performance over time is measured (audited) and fed back to the clinician.\(^{40}\) It does not necessarily include recommendations for action and there may not be other comparative data. Audit and feedback are most successful if the clinician receiving the feedback recognises clinical practice must change, is able to change within a supportive environment and can respond to the feedback immediately.\(^{41}\)

10.1.2.2 *Use of local opinion leaders*
Local opinion leaders are people who are trusted by their colleagues to evaluate new medical information and technology and apply this to the local context. This context may be hospital or community based, regional or national. Local opinion leaders are approached frequently for clinical advice, should have good listening skills and are perceived as clinically competent and caring. They should be able to influence clinical practice amongst their peers. However, randomised controlled trials in this area have not shown consistent results.

10.1.2.3 Local consensus process

The involvement of local healthcare providers in solving local problems and in particular, analysing barriers that may not be otherwise apparent to those in central planning authorities is invaluable and can help tailor solutions, particularly if consumer input is included.

10.1.2.4 Consumer mediated intervention

These can include anything that aims to change clinicians’ behaviour by the intervention of the consumer. Providing consumers with written information about the findings of an evidence-based guideline can influence the questions they ask and the expectations they have about therapies and treatments.

10.1.3 IMPLEMENTATION STRATEGIES SHOWN TO HAVE LITTLE OR NO EFFECT

10.1.3.1 Dissemination of educational material

Single dissemination of educational materials including recommendations for clinical care, audio-visual materials, electronic publications and journal articles appears to have only a small effect in altering practice if carried out in isolation. Dissemination of guidelines by postal mailout alone is unlikely to be effective in influencing clinical behaviour, although it may raise awareness and knowledge of the guideline. Unfortunately, this knowledge does not necessarily translate into changes in clinical practice. It is however a necessary and useful adjunct within a multifaceted implementation process.

10.1.3.2 Didactic educational sessions

The standard educational approach of lectures, personal visits or workshops, in which there is no explicit effort made to change practice, have often failed to change performance. However, they are relatively cheap and easy to set up.
10.1.4 IMPLEMENTATION STRATEGIES THAT HAVE UNKNOWN EFFECTIVENESS

10.1.4.1 Incentives and penalties

Many incentives already exist within the health system that are designed to influence clinical behaviour. For example, these include additional fee charges for giving immunisation, removal of items for reimbursement and provision of funds for retraining, personal satisfaction, professional incentives such as accreditation or Continuing Medical Education points, and receipt of personalised relevant data through the evaluation process. Clinician behaviour is less likely to change if there are disincentives to do so such as increasing workload, extra time required, no familiarisation, the need for extra resources and the need for specialised skills and equipment. Disincentives for consumers also exist such as complications and the need for travel.

10.1.4.2 Administrative interventions

Administrative or management interventions that encourage or compel health professionals to change their practices are used widely but are infrequently evaluated. These include removing barriers such as requiring approval of a specialist for certain tests, simplifying order forms, as well as providing incentives.

10.2 DISSEMINATION AND IMPLEMENTATION

Steps in planning dissemination and implementation:

- Develop a statement of purpose of the implementation strategy
- Identify who should be involved in an implementation project
- Identify the target audiences
- Analyse the current gaps between ideal and current practice
- Distil the key messages of the guideline
- Consider the potential barriers
- Decide on appropriate strategies given the local setting
- Consider the cost implications of the strategies
- Evaluate the success of the implementation

10.2.1 Clarify the objective of the implementation strategy

The first step in implementing a clinical guideline is to establish the objectives of the implementation strategy. It is important to be specific and maintain a focus on improvements and the change in practice you are trying to achieve. Major objectives such as reducing the number of caesarean sections in a hospital are unlikely to succeed
but specific objectives such as reducing the proportion of women having caesarean sections in a hospital to 15% may be achievable.

10.2.2 **Identify who should be involved in the implementation project**

An effective implementation team should have multidisciplinary representation. The major stakeholders and those who are able to effect change at policy and administrative levels should all be included:

- patient or consumer representatives
- policy makers such as the Ministry of Health, District Health Board members
- administrative decision makers - managers, executives, Medsafe, PHARMAC and other purchasers
- clinical decision makers
- representatives of the various practitioners who provide care to patients groups developing protocols and clinical pathways
- industry – for example pharmaceutical companies

By the end of this step, you will have a list of people you consider it important to influence. These are people whose cooperation is required for successful changes to occur. These are also some of the people to whom the guideline should be disseminated.

10.2.3 **Identify the different target audiences**

It is also important to list the different target audiences as there is a range of strategies that are available but strategies will not be suitable for all groups. You will almost certainly have to tailor the approach for different groups. Implementation at a local level may have an entirely different set of target groups than the national implementation group. (see Table 10.2)
### TABLE 10.2: POSSIBLE AUDIENCES FOR GUIDELINE IMPLEMENTATION AND THEIR DIFFERING REQUIREMENTS

<table>
<thead>
<tr>
<th>Audience</th>
<th>Type of decision maker</th>
<th>Information needs</th>
<th>Preferred format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legislative</td>
<td>Politician Bureaucrat</td>
<td>Problem definition</td>
<td>Person to person</td>
</tr>
<tr>
<td></td>
<td>Interest group</td>
<td>Affirmation of assumed causes</td>
<td>Overview in brief memorandum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Policy ‘ideas’</td>
<td>Media</td>
</tr>
<tr>
<td>Administrative</td>
<td>Programme manager</td>
<td>Programme evaluation</td>
<td>Special contacts</td>
</tr>
<tr>
<td></td>
<td>Regional administrator</td>
<td>Practice variation</td>
<td>Summary report</td>
</tr>
<tr>
<td></td>
<td>Hospital executive</td>
<td>Cost effectiveness</td>
<td>Dedicated seminar</td>
</tr>
<tr>
<td>Clinical</td>
<td>Practitioner</td>
<td>Effectiveness</td>
<td>Colleagues</td>
</tr>
<tr>
<td></td>
<td>Professional society</td>
<td>Ethics</td>
<td>Action oriented synthesis</td>
</tr>
<tr>
<td></td>
<td>Expert panel member</td>
<td>Patient preference</td>
<td></td>
</tr>
<tr>
<td>Industrial</td>
<td>Company scientist</td>
<td>Marketable product</td>
<td>Depends on scientist versus non scientist</td>
</tr>
<tr>
<td></td>
<td>Corporate executive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Venture capitalist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>Patient</td>
<td>Effectiveness</td>
<td>Personal</td>
</tr>
<tr>
<td></td>
<td>Friend/relative</td>
<td>Safety</td>
<td>Provided in a range of formats (written, verbal, video etc)</td>
</tr>
<tr>
<td></td>
<td>Person with chronic illness</td>
<td>Cost</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Well person</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consumer organisations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10.2.4 Analyse the current gaps

This should have been done at an earlier stage (suitability screen) but it is important to revisit them, as circumstances may have changed and plan strategies to close the gaps between ideal and current practice.

10.2.5 Identify the key messages of the guideline

What are the key messages that need to be conveyed? These are likely to be the key recommendations that you identified from the guideline development process that highlight the gap between the evidence and current or accepted practice. Adoption of these recommendations is considered likely to reduce the gaps. These messages should be listed and it may be that the messages have different target audiences requiring different strategies. The messages should also be ranked in order of priority so the important ones are more likely to receive a higher profile and emphasis.

The available information needs to be assessed for dissemination and it is important to ensure that the summary, recommendations and algorithm are in a format that is suitable as different groups may need different approaches. In particular preparation of consumer publications should be a separate task.

10.2.6 What are the barriers?

An assessment of the barriers to successful implementation should be considered at this stage. There are likely to be barriers at a number of levels and stages of the process. The barriers are also likely to be different according to different providers and situations. For example, specialists may identify barriers that are different from those of general practitioners. Similarly, allied health professionals may identify barriers that are different from those of doctors. Barriers may be more difficult to overcome in a large institution such as a hospital than in a small community clinic or practice.

In an ideal world, assessment of the barriers could take place with surveys, interviews or focus groups involving the primary target group. Resources for this would obviously have to be made available. However, an alternative is to utilise the knowledge and expertise of the implementation group who are collectively likely to be familiar with sector difficulties and barriers. Forcefield analysis at this stage may be a helpful way to clarify potential problems and opportunities. The diagrammatic picture allows us to see an overview of what influences can either drive or impede change.

Specific characteristics of guidelines that have been identified as barriers include:

- Guidelines which, although well developed, are too large, complex and difficult to use
- Guidelines that are not user friendly due to publication formatting and layout
- Guidelines which are not applicable or workable locally5
10.2.7 What are the appropriate strategies given the range of audiences and their locations?

Implementation strategies range from simple and low cost strategies to multifaceted strategies. These are described in detail in Section 10.1. The degree of complexity in the guideline recommendations will need to be reflected in the implementation plan. Fewer strategies are usually required for simpler guidelines. The different settings should be identified if these could augment or deter guideline uptake. Implementation strategies by local groups such as Independent Practitioner Associations and by District Health Boards may include action such as introducing a new guideline with their annual practice improvement plans.

10.2.8 Consider the cost implications of the strategies

The implementation budget for each guideline should be determined by the particular strategies identified as appropriate and necessary for influencing changes in practice. Depending on the topic of the guideline, stakeholders may be prepared to offer implementation funding perhaps in tandem with local groups. Potential cost savings of putting the guideline into practice can be helpful in influencing the willingness of organisations to provide implementation funding.
10.2.9 Evaluate the impact of the guideline implementation strategy

In order to evaluate the impact of the guideline and the implementation strategies, it is necessary to collect data on who has seen and is familiar with the guideline, as well as the actual use of the guideline. It is also appropriate to collect information on the use of additional health resources and changes in health outcomes. Ideally, this evaluation should be planned in advance of the implementation plan so before and after data can be compared. It is also important to not only look at these things in the short term but also to evaluate whether the initial interest and changes achieved are sustained over time.

Questions to ask when evaluating guideline implementation include:

- The target groups’ awareness and understanding of the context of the guidelines (post implementation survey);
- The guideline’s relevance to these groups (consumer group vs general patient population);
- Degree to which consumer groups were involved and informed;
- The extent to which barriers to implementation were overcome;
- The extent to which guidelines affected or changed clinical practice;
- Whether the guidelines were sufficiently flexible for use in different local and regional circumstances and whether they were interpreted consistently by different providers.
APPRAISAL AND EVALUATION OF CLINICAL PRACTICE GUIDELINES

AIMS OF CHAPTER:

- To describe evaluation of the process and content of guideline development
- To describe evaluation of CPGs using the AGREE instrument
- To evaluate the effectiveness of the guideline in influencing changes in practice

11.1 AGREE EVALUATION INSTRUMENT

There is a large body of literature identifying attributes that define high quality clinical practice guidelines. The term “quality of clinical practice guidelines” refers to the extent that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid and feasible for practice. The assessment includes judgments about the methods used for developing the guidelines, the content of the final recommendations, and the factors linked to their uptake.

The purpose of the Appraisal of Guidelines Research and Evaluation in Europe (AGREE) instrument is to provide a framework for assessing the quality of clinical practice circumstances.17 The primary aim of the project was to develop an appraisal instrument to compare guideline development approaches in Europe. A secondary aim was to identify potential areas for integration of guideline development in order to minimise duplication of efforts and maximize efficiency of development in individual countries.

This project was a collaborative effort of 11 European countries (and Canada) to produce a validated tool for guideline appraisal. The project was funded by the European Union and took 4 years to complete.

The AGREE instrument has 23 items in 6 domains. The domains are:

1. Scope and purpose
2. Stakeholder involvement
3. Rigour of development
4. Clarity and presentation
5. Applicability
6. Editorial independence
The instrument does not provide an overall quality score for a guideline. Aggregate scores are not reported but the individual domain scores may help identify specific areas of strengths and weaknesses in a guideline.

The AGREE instrument is presented in Appendix 4 and includes user instructions. Guideline developers may use the tool in 3 ways:

1. To assess available guidelines at the suitability screen phase
2. To plan guideline development and use the tool as a checklist so that the quality of the guideline is ensured
3. To assess available guidelines for local adaptation

11.2 EVALUATION AND ADAPTATION OF OTHER GUIDELINES

Once the evaluation of a guideline using the AGREE instrument has occurred and meets an acceptable level for local adaptation, then issues for local utilisation of the guideline must be considered.

11.3 EVALUATION OF THE IMPACT OF THE GUIDELINE

The approach to this question should be guided by the clinical questions regarding the gap between current and ideal care that was identified in the suitability screen. This may not be the role of the guideline development team but certain approaches and analyses should be suggested.

These may include:

- How successful was the dissemination project? Did the guideline reach the ‘right’ people and are they familiar with it?
- Has the gap between current and best practice narrowed?
- How has resource utilisation changed (decreased/increased)?
AIMS OF CHAPTER:

- To consider approaches to the revision and adaptation of clinical practice guidelines
- To provide guidance for local and national adaptation of guidelines produced by other guideline groups

12.1 REVISION AND UPDATING OF GUIDELINES

Guidelines should be kept up to date. Some guideline development panels are ongoing and meet at intervals to consider if new evidence has come to light and whether there are other influences that might lead to the recommendations needing to be revised.

12.2 LOCAL AND NATIONAL ADAPTATION OF GUIDELINES

There is a growing recognition that it is not possible for national guidelines to be produced on every health problem of concern. Considerable resources are involved in developing the evidence base from which that guideline recommendations are developed and it is a measure of the collaboration of national guideline programs that this information is now more readily available. Local adaptation addresses ownership and recognition of the clinical issue to be addressed (the gap) by obtaining backing from respected opinion leaders, (involvement of local clinical staff), organisational structures, and contextual issues and taking into account specific local circumstances acknowledging involvement of local clinical staff. For example, SIGN (the Scottish Inter-collegiate Guidelines Group) sponsors the modification of its national guidelines to local protocols for local application, retaining responsibility for local implementation. This local ownership step has been shown to increase the uptake of local protocols.

12.2.1 Adapting guidelines for the NZGG.

If a guideline development team decides to adapt a guideline developed by others they should address the following questions:

- Are the patients/consumers referred to in the guideline similar to the comparable NZ population?
- Are the recommendations feasible within the NZ setting?
12.2.2 Local adaptation of national or other guidelines

Local adaptation of guidelines is generally accepted to improve ownership of the guidelines by local professionals or a professional body. This may take place as part of a quality improvement programme and is thought to be the key to an effective and credible implementation strategy. Local adaptation creates local innovation and indirectly fosters an environment where ‘change’ is more acceptable. A possible disadvantage is diminished validity of the guideline.
AIM OF CHAPTER:

• To describe the ‘milestones’ in the process of development of a clinical practice guideline

13.1 CONVENE MULTIDISCIPLINARY GUIDELINE DEVELOPMENT TEAM

The precise composition of the team will depend on the nature of the guideline but should be designed to encourage the expression of diverse interests. Some members may be corresponding members to limit the size of the working group yet ensure that all perspectives are considered. In general, consideration should be given to involving people from the following groups as appropriate:

• Clinicians with expertise from relevant disciplines
• Other relevant health professionals
• Representatives of relevant consumer groups*
• Experts in research methods relevant to guideline development, eg epidemiologists, biostatisticians
• Health economists
• Public health specialists
• Representatives of professional groups such as Colleges or Independent Practitioner Associations (IPAs)
• Representatives of regulatory agencies

*Guideline development teams can appear intimidating to consumers. Ideally there should be a minimum of two consumer representatives nominated by an established health consumer group and those representatives should have attended a consumers’ guideline training course. Maori and Pacific Island cultural views need to be incorporated within an evidence based guideline.
13.2 GUIDELINE DEVELOPMENT TEAM PROCESS

Diversity is an essential feature of a multidisciplinary guideline team. Early in the process members need to discuss and come to agreement on the following:

- Acceptance and tolerance of varying viewpoints
- Open discussion
- Declaration of any conflict of interest. These should be both real and perceived conflicts of interest by the members of the group.
- Agreement that evidence trumps opinion
- Agreement of shared workload
- Agreement of contribution by reading/sifting evidence and other printed material
- Commitment to attending the meetings, i.e. negotiate time off work, arrange backup
- Organising fees and reimbursement arrangements
- Define any areas of confidentiality
- Agreement not to publish information prepared for group or about the group without agreement of the group
- Media statements to be done in conjunction with, or at the direction of the group (e.g. particular guideline development team members could be given this task)
- Agreement that guideline development team members who are present in their capacity as representatives identify any controversial information that arises during the guideline development team process that they consider needs to be shared with their constituency. A representative in this position should let the group know what information is at issue and explain the reasons why this information needs to be discussed with a wider group. A representative needs to balance the need for confidentiality with consultation and accountability expectations.

13.3 DEFINE THE AIM AND TARGET AUDIENCE FOR THE GUIDELINE

Before proceeding, the team should clarify the purpose and focus of its task and identify the anticipated audience/s for the guideline. This should involve careful description of the following:

- The condition/situation at issue (the appropriate prescribing of HRT is not a condition)
- The type of care providers for whom the guidelines are intended
- The groups of consumers for whom the guidelines are relevant
- Description of consumers not covered by the guidelines (e.g. inclusion and exclusion criteria where applicable)
- Type of setting in which the guidelines will be used
- The specific interventions to be critiqued for the purposes of the guideline

It is essential that written information (e.g. leaflet, information sheet) based on the guideline, be developed for consumers as a way of informing them of the recommended best practice based on evidence.
13.4 INVOLVING THE STAKEHOLDERS

A draft of the guideline should be sent to all stakeholders and usually 4-6 weeks is given for comments to be received. This stage should occur before the final meeting of the guideline development team so that changes to the recommendations can be made as required. Endorsements can then be requested from the medical societies and health care organizations that were asked to comment on the guideline. Practicing clinicians should also be included.

13.5 PRE-TESTING THE GUIDELINE

Ideally the guideline should be pre-tested in a variety of settings, including both primary and secondary care. Final alterations to the guideline may need to be made in consultation with the wider guideline group.

13.6 TIMELINE FOR GUIDELINE DEVELOPMENT

At the beginning of the process, the guideline leader, in discussion with the group, should outline a reasonable timeline for the tasks and meetings. Each meeting usually takes one day. 3-4 meetings are likely to be sufficient.

First meeting:

- Introductions
- Clarification of the task and expectations of the group members.
- Group processes outlined including agreement on levels of evidence and grading of recommendations
- Declaration of conflicts of interest
- Background to topic outlined
- Suitability screen developed and discussed
- Group works on and agrees on clinical questions and searching
- Agreement on a template for the final guideline including topics to be covered in each chapter
- Assign responsibility for each chapter to a person/subgroup
- Tasks assigned for completion by next meeting (eg divide internal evidence and evidence table development between group members)

Second and third meeting:

- Searches and evidence tables completed
- Agreement on levels of evidence to be assigned
- Start to develop recommendations linked to each clinical question
- Start to develop algorithm (if appropriate)
- Assign to different team members the tasks of writing specific sections and finalising evidence tables and recommendations
• Start to work on the balance sheet
• Draft guideline sent to all stakeholders and practicing clinicians and feedback incorporated into the guideline

**Final Meeting:**

• By this stage the evidence tables should be fully completed
• Final discussion and agreement on the recommendations
• Agreement on the algorithm structure and process
• Discussion on what will go in the summary and what will be available in the full text
• Discussion of dissemination and implementation plan
• Agreement on balance sheet
• Plan pre-testing of the guideline in both primary and secondary care
REFERENCES

17. AGREE Collaboration. www.agreecollaboration.org
19. Booth, A. Becoming ADEPT: Applying Diagnosis Etiology Prognosis and Therapy methodological filters for retrieving the evidence. Distance Learning Programme,School of Health and Related Research, University of Sheffield, UK.
Appendix 1

Guidelines for the involvement of consumers in guideline development

Women's Health Action
May 1999

Contents

1. Introduction

2. Guiding Principles

3. Objectives and Benefits of Involving Consumers

4. Process

5. National Action Plan

6. Key Points for Involving Consumers
   6.1 who to involve
   6.2 consumer issues
   6.3 information and assistance
   6.4 involving consumer representatives
   6.5 selecting consumer representatives
   6.6 skills needed by consumer representatives
   6.7 guidelines for the payment of consumer representatives
1. Introduction

This booklet was initially developed as a discussion paper for workshop participants at the National Guidelines Conference in Wellington (6-7 June 1997) who were interested in the role of consumers in guideline development.

Suggestions made by participants were incorporated into an updated draft that was available for further discussion and comment. The paper has now been developed into a booklet to be used as a practical guide for involving consumers in the development, review and implementation of guidelines. Please direct any comments, suggestions or requests for copies to Women’s Health Action, PO Box 9947, Newmarket, Auckland ph (09) 520-5295 fax (09) 520-5731 or by e-mail to coneys@womens-health.org.nz

2. Guiding Principles

2.1 partnership and collaboration
- Treaty of Waitangi
- individual consumer-health professional contact
- consumer representatives and policy makers, planners, managers and health professionals

He tangata, he tangata, he tangata
It is people, it is people, it is people

2.2 democratic participation

2.3 equity and fairness

2.4 accountability

2.5 acceptability

2.6 to ensure the rights of consumers are upheld

2.7 to ensure consumer input is valued

2.8 wholistic approach
3. Objectives and Benefits of Involving Consumers

The following objectives are needed to achieve the guiding principles. The objectives also include benefits that can be obtained by involving consumers in guideline development:

3.1 partnership and collaboration
♦ guidelines will acknowledge and reflect bicultural values
♦ an informed and collaborative approach to decision making is more likely
♦ there will be increased opportunities for community groups to participate in decision-making on health matters of particular interest to them

3.2 democratic participation
♦ guidelines will acknowledge and reflect differing cultural values
♦ guidelines will reflect community values
♦ input into guidelines will take into account a range of viewpoints
♦ guidelines will facilitate increased community access to information about health and health services as well as enhancing the community’s understanding of health issues
♦ there would be a greater sense of community ownership
♦ decisions are more likely to be made with users of services rather than for them
♦ involving consumers is an effective way to balance an overly medical approach

3.3 equity and fairness
♦ there is more likely to be an opportunity to reach a shared understanding of the purpose and application of a guideline
♦ consumer input can help ensure the content and terms used in a guideline can be widely understood
♦ consumers can make sure the impact of decisions are taken into account

3.4 accountability
♦ guidelines will be credible, usable and of a high standard
♦ guidelines will be understandable and publicly available
♦ the allocation of health resources need to reflect community values and be spent on health services and treatments of demonstrated benefit

3.5 acceptability
♦ consumer involvement will ensure a consumer focus
♦ the guidelines are more likely to be acceptable to consumers
♦ the guidelines are more likely to use simple, clear language and avoid confusing jargon so the guideline can be used for any audience
♦ a greater range of formats is likely to be explored and information is more likely to be made available in a range of languages for those who have no English, or English as a second language

3.6 to ensure the rights of consumers are upheld
♦ guidelines will respect each person’s autonomy and their right to make decisions in consultation with others
guidelines will help to improve access for consumers to the most appropriate health services
- guidelines have a role in making sure consumers have sufficient information to make informed decisions
- consumers can use evidence-based guidelines to argue for increased availability of services or the modification or cessation of practises not of benefit

3.7 to ensure consumer input is valued
- there will be increased recognition of the value of consumer input
- consumer input is less likely to be marginalised

3.8 wholistic approach
- the guideline will take into consideration the whole person, their whanau, beliefs and circumstances including social, physical, emotional, spiritual and psychological needs
- there is likely to be a greater emphasis on self-help, wellness and ease rather than disease

4. Process

Using a Partnership Model with a Consumer-Centred Approach

The guideline development process needs to ensure consumers are included and steps are taken to achieve the principles and objectives outlined above. There also needs to be a commitment to working with consumers as equals. A partnership approach requires consumers to be involved from the beginning and for this involvement to be maintained through to the completion and implementation of the guideline. Guideline development needs to be evolutionary and inclusive.

Proposed steps:

1. The need for a guideline is identified (can be identified by anyone)

2. Someone needs to take responsibility for the process (a meeting of interested parties could determine this, for example)

3. Someone needs to check out if a similar guideline has already been developed. The NZ Guidelines Group website should provide relevant information and contacts.

4. A funding/resourcing plan needs to be determined to ensure consumers are able to participate in the process

5. A comprehensive consultation strategy is developed to ensure all relevant viewpoints are taken into account. Public participation is essential to ensure community ownership and acceptance of guidelines.
6. Those appropriate formats, for the different interest groups, are identified to maximise the likelihood of a guideline being used.

7. A draft using the standard format is circulated for comment

8. Comments from the draft are incorporated into an updated draft for further comment (this may need to be repeated several times)

9. A final draft is circulated prior to finalising the guideline along with a proposed distribution and implementation plan

10. Copies of the final guideline are widely distributed

11. The completed document is entered into electronic and other record systems for reference purposes. The guideline is also developed into other formats as appropriate.

12. At an agreed interval there will be a review of the guideline to take into account new information and the need for the guideline to be updated. There should also be agreement on when to carry out an evaluation of the usefulness of the guideline.

The step-by-step process outlined above should also be used for reviewing and updating existing guidelines.

5. National Action Plan

To further assist the process of developing useful guidelines, the following action is recommended:

• A national Guidelines Register is established, which keeps a record of all guidelines; whether the guideline is under development, has been completed or is being updated.

• A national guideline database is set up that includes a mailing list of people/groups wishing to be involved with guidelines

• A national guidelines clearinghouse role is created so people with shared interests are able to contact each other

• a designated person needs to be appointed to co-ordinate guideline development in order to avoid duplication and to keep interested parties well informed and up-to-date. (This person could be responsible for the national database, clearinghouse role and making sure that information about guidelines is easily accessible.)
6. Key Points for Involving Consumers

6.1 Who to involve?

Guideline development could include a number of categories of consumers:

- patients/users of health services
- caregivers and family/whanau members
- representatives from organised community groups (refer to section on the selection of consumer representatives)
- the public at large

Consideration also needs to be given to ensure there is Maori input, and that the views of other groups such as those from the Pacific Islands and Asian communities are represented. Other factors such as age, health condition or other relevant experience are also important.

6.2 Consumer Issues

Consumer groups have identified the following key issues in relation to health issues and service provision. It is important to consumers that there is:

1. access to quality services provided to a high standard
2. access to sufficient good quality information (making an informed choice requires knowledge of the options and what to expect from each)
3. consultation on an ongoing basis throughout the process
4. involvement and empowerment of consumers
5. accountability

6.3 Information and assistance

When involving consumer representatives, it is important to:

- disclose the level of funding that is available for the service under discussion or review
- keep consumer groups informed and maintain an ongoing dialogue with them
provide extra assistance, explanations and background information, particularly if the area under consideration is of a technical nature

• seek guidance from consumer groups about consultation options

• make meeting times in consultation with consumers, in recognition that many have other responsibilities and obligations including dependants to consider

• ensure there is sufficient funding to pay consumers and to cover additional expenses such as child care and transport costs

• ensure drafts and completed guidelines are widely distributed within the community as well as to health professionals and managers

6.4 Involving Consumer Representatives in Guideline Development

Advantages

• consumer representatives have a mandate to represent consumer views

• consumer representatives are accountable to the consumer groups they represent

• part of the mandate to be a representative includes the responsibility to report back

• consumer involvement could help shift the current medical model focus to a consumer model

• consumer involvement could broaden the focus from the restricted forms of predominately medical evidence presently being used in guideline development, to include holistic and self-help evidence as well

• consumers could influence the assumptions/baseline which is on a reduction of symptoms rather than on wellbeing and quality of life

• the focus is less likely to be on people just as numbers

Disadvantages

• there could be a consumer versus professional focus which could result in a consumer wish list and health professionals distancing themselves from the end result

• there is a lack of research on consumer issues and community values so guideline development relies on randomised controlled trials (RCTs) and other research with a strong scientific focus. The unresolved conflicts over what is the evidence and who defines the ‘best’ evidence could be alienating and off-putting for consumers.
6.5 Selecting consumer representatives

Consumer groups accept that representatives need certain skills and that there are particular requirements necessary to achieve an effective consultation process. There is a need, for example, for consumer representatives who are well informed and understand the concept of accountability.

A consumer representative:

- should be both a consumer/user of health services, and someone who is actively involved in community-based work on health issues with links and connections to a consumer group/s. They are more likely to be effective if they represent the views and perspectives of consumer groups.

- should be part of an established consumer health group. This link ensures accountability to the community via the group. The group can also provide support for the representative allowing the selected person to check out and discuss difficult issues that arise.

- represents a group or groups, and is not an individual participant. This provides for a degree of protection when particular positions attract criticism of a personal nature.

- comes with a mandate from the community, and has been through a process of selection within the community to qualify for this role.

Inappropriate choices for consumer representation

It is not appropriate:

- that those holding medical power, or operating as health professionals and health service providers take on the role of consumer representatives.

- to use professional social work services to act on behalf of the community. Such people do not necessarily have a mandate to represent community views and aspirations.

- to use an individual lay person as a consumer representative. Such a person acts in an individual capacity and is not usually accountable to anyone. They are often in a vulnerable position with only personal systems of support. They may not understand obscure medical jargon and may be easily intimidated. A lay person may not be up-to-date with how the health system works and what the current consumer concerns are.
6.6 Skills needed by consumer representatives

Consumer representatives should be selected for their ability to take on the task as well as their willingness to act as a representative. Skills identified as fundamental to being an effective representative include:

An ability to:

- analyse an issue and judge its effects on consumers and different sectors of the community
- present an argument rationally and convincingly
- see short and long term outcomes as well as directions leading from decisions
- negotiate on issues to achieve the best acceptable outcome
- strategise to achieve the desired outcomes from a committee or working party. To do this, representatives need to research the background of issues with which the committee will be dealing
- assess when consultation with the community group or groups is required. The representative may not be able to give instant answers on particular issues, but must be able to obtain the information.

A willingness to:

- act as a representative of a group and not as an individual, and to consult with the group when necessary.

6.8 Guidelines for the payment of consumer representatives

Introduction
A commitment to involving consumers in guideline development must include the provision of satisfactory payment and reimbursement arrangements. Health professionals and others involved with the development and implementation of guidelines are often able to participate as part of their paid employment.

The situation is very different for consumers who either have to take time off from their usual job or make arrangements for childcare. Consumer groups don't usually have the funds to provide the financial support needed for their representatives. Access to computers, email, fax and photocopying facilities is generally easily achieved by health professionals but can present difficulties for consumers. It is also important that the contribution and expertise provided by consumers is properly recognised.

A consumer group may be well placed to co-ordinate guideline development if funded to do so, or consumer representatives with children may require payments to cover child-care fees so they are able to participate.
**Working parties and committees**
A standard payment should be provided for attendance at meetings. In addition to this meeting fee there should be an agreed payment for additional work to be carried out outside of the meeting times. In situations where a meeting fee is being paid to other participants, a consumer representative should be paid at the same rate.

In situations where health professionals and others representing the health sector are contributing as a part of their work, the meeting fee for consumer representatives should be at least $250 for a full day.

**Lecturing and presenting**
Consumers who are preparing and carrying out a lecture or presentation need to be paid for the preparation time as well as for the presentation. Depending on what is involved the payment could be based on an hourly rate or on an amount that adequately compensates the consumer for the time and effort involved. An hourly rate should be at least $30. Alternatively, a global fee could be negotiated that takes into consideration the $30 an hour minimum rate.

**Facilitating and running workshops**
A consumer involved with the preparation and running of a workshop should be paid $30-$50 per hour. Alternatively, a global fee could be negotiated with this rate in mind so that the consumer facilitator is adequately compensated for the time and effort involved.

**Teleconferencing**
Consumers participating in meetings carried out by teleconference should be paid an equivalent rate of a meeting fee taking into account preparation and time spent on the call.

**Additional Expenses**
Although not necessarily a complete list of possible expenses that could be encountered, travel, childcare, phone/fax, photocopying and postage are common situations that consumers may need to be reimbursed for.

Particular consideration needs to be taken into account when a consumer is representing a geographically wide constituency so adequate remuneration is provided for toll calls.

A clear arrangement needs to be put in place at the beginning of a guideline development (or review) process so that claiming is straightforward and not awkward for consumers to initiate.
The GATE Notes: a Generic Appraisal Tool for Epidemiology

The GATE Notes were developed by the Effective Practice Institute, University of Auckland. You are welcome to copy them, if you acknowledge their origin. Please contact Professor Rod Jackson (rt.jackson@auckland.ac.nz) if you have any questions, comments or suggestions.
THE GATE NOTES:

- These guides incorporate most of the questions from the JAMA series of “Users’ Guides to the Medical Literature” (1-9), but they have been rearranged to more systematically link design and appraisal. It is recommended that you use the JAMA guides or an updated version of the key JAMA guides such as the EBM handbook by Sackett et al (10), as reading to complement the GATE guides.
- Each section of GATE starts with a brief explanation of the study type, then a checklist (2 pages) and a User guide for each checklist. It is useful to fill out as you go through a full set of appraisal questions.
- When you have completed each Section of the checklist, it is important to weigh up the overall quality of that aspect of the study.

AN OVERVIEW OF THE SUBTYPES OF EPIDEMIOLOGICAL STUDIES:

Epidemiological studies can be differentiated into major subtypes based on how the exposure and comparison subgroups in the study population are assigned (i.e. experimentally or non-experimentally) and based on the types of occurrence measures used (i.e. prevalence in cross-sectional studies and incidence in longitudinal studies). Some study designs are modifications of these major subtypes. Each study subtype can be derived using the Generic Appraisal Tool for Epidemiology (GATE) approach based on the 5 part PECOT diagram. A brief overview of each study subtype is given below.

There are 5 types of studies that are used in guideline development.

1. Randomised controlled trials. (RCTs): This is an experimental study where participants are randomly allocated to exposure(s) or comparison intervention (sometimes a placebo). Outcomes are typically measured over a period of time in RCTs, therefore most RCTs are longitudinal studies measuring incidence, however outcomes can also be measured cross-sectionally (i.e. prevalence measures) in RCTs. Screening studies investigate the effect of a screening test on a health-related outcome and should ideally be RCTs in which the test allocation is randomly allocated) but are sometimes cohort studies (see below) if the use of the test is ascertained rather than allocated by the investigator).

2. Cohort studies: If participants are assigned to exposure(s) and comparison groups based on the MEASUREMENT of these factors (rather than being randomly allocated), the study is non-experimental. These studies are often called observational studies, although outcomes are observed in all studies, both experimental and non-experimental, so the
term “non-experimental” is more appropriate than “observational”. **Cohort studies** can be considered as non-experimental versions of RCTs in which the exposure and comparison groups assignment is determined by measurement of these factors in the study participants and outcomes are measured over a follow-up period. Cohort studies are non-experimental longitudinal studies.

3. **Case-control studies** (non experimental) are “nested” inside cohorts and can be considered as efficient versions of cohort studies (not included in these notes)

4. **Prognostic studies** (non experimental) are cohort studies in which the objective is to investigate how well an exposure(s) predicts the occurrence of outcomes rather than whether or not the association is causal.

5. **Cross-sectional studies** (non experimental) are similar in design to cohort studies, except that outcomes are measured at one point in time; usually at the same time as the study population exposure and comparison groups are defined. **Diagnostic test studies** are cross-sectional studies that compare the accuracy of a diagnostic test with a reference standard.
1. RANDOMISED CONTROLLED TRIALS (Treatment studies)

(Relevant JAMA User's Guides, Numbers IIA & B: references (3,4)

Introduction:

The most valid study design for assessing the effectiveness (both the benefits and harms) of therapeutic or preventive interventions is the randomised controlled trial (RCT). This is an experiment in which the investigator controls the random allocation of participants or study communities in the study population to the interventions of interest (i.e. exposure or intervention subgroup/s) or a comparison subgroup (i.e. the control group).

Trials are considered the “purest” type of epidemiological study because the investigator has control over exposure allocation. If the investigator randomises individual participants or communities to intervention and comparison subgroups, it is possible to minimise differences in baseline characteristics between the groups that might influence the outcome of interest (i.e. it minimises confounding).

The comparison or control group may be allocated a placebo intervention, an alternative real intervention or no intervention at all.

If randomisation is successful and the groups are similar at baseline, the investigator can be more confident that observed differences in outcomes between the groups are related to the intervention rather than confounding factors.

Trials have a number of potential limitations compared with other designs. For practical and ethical reasons some important questions cannot be investigated using an experimental design. Moreover when trials are possible, they are often conducted in artificial environments and with highly motivated volunteers. This may limit their generalisability to populations of interest.
### GATE Checklist for Randomised Controlled Trials (Intervention: benefit or harm)

#### Study author, title, publication reference

<table>
<thead>
<tr>
<th>Key 5 part study question (PECOT). Was it focussed?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

#### SECTION 1: STUDY VALIDITY

**Appraised by:**

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>How well was this criterion addressed?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>✓ ? x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What were the key selection (inclusion &amp; exclusion) criteria?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were they well defined?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were they replicable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were inclusion &amp; exclusion criteria appropriate given study question?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposures &amp; Comparison</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What were the exposures (interventions) &amp; comparison?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well defined?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replicable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was assignment to exposure &amp; comparison groups randomised?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was randomisation concealed?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was randomisation successful:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>were exposure &amp; comparison groups similar at start of study?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were all participants analysed in groups to which randomised?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were participants, health workers, researchers blind to interventions?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apart from study interventions, were groups treated equally?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was compliance with interventions measured?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was it sufficient?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What outcome measures were used?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well defined?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replicable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How complete was follow up?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was it sufficient?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How many dropouts?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was outcome assessment blind?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Was follow up time sufficiently long to detect important effects on outcomes of interest?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY DESIGN:** How successfully do you think the study minimised bias? Very well = +, okay = ⊙, poorly = -
### SECTION 2: STUDY RESULTS: MAGNITUDE & PRECISION OF EFFECTS

What measures of occurrence (incidence / prevalence) & intervention effects (RR / RD / NNTs) were reported?

What measures of precision of effects were reported (CIs, p-values)?

#### THE NUMBERS TABLE: OCCURRENCE, EFFECT ESTIMATES & PRECISION

<table>
<thead>
<tr>
<th>Outcomes* &amp; Time (T)</th>
<th>Exposure event rate (EER=N_E/D_E/T) or mean*</th>
<th>Comparison event rate (CER=N_C/D_C/T) or mean*</th>
<th>Relative Risk* (RR = EER/CER) ± (95% CI)</th>
<th>Risk difference or mean difference (RD = CER-EER) ± (95% CI)</th>
<th>Number Needed to Treat* (NNT = 1/RD) ± (95% CI)</th>
</tr>
</thead>
</table>

* if outcomes continuous, can calculate means, mean differences, but not NNTs (don’t usually calculate relative means)

\( D_E = \text{Denominator (D) for exposure (intervention) group(s)}, D_C = \text{D for comparison (control) group}\)

\( N_E = \text{Numerator (N) for exposure group(s)}, N_C = N \text{ for comparison group}\)

#### QUALITY OF STUDY RESULTS: Useful, precise +/- or sufficient power?
- Very good = +
- Okay = ∅
- Poor = -

### SECTION 3: STUDY APPLICABILITY

#### Participants
- Was the source population for participants well described?
- Were participants representative of the source population?
- Can the relevance / similarity of the participants to a specific target group(s) be determined?

#### Exposures & Comparison
- Were the characteristics of the study setting described? e.g. rural, urban, inpatient, primary care
- Can the applicability of interventions be determined?
- Can the relevance of the comparison group management be determined?

#### Outcomes
- Were all important outcomes considered: benefits? harms? costs?
- Are likely benefits greater than potential harms & costs (or vice versa)? In what target group(s)?

#### QUALITY OF STUDY APPLICABILITY: (a) Was it possible to determine applicability? (b) Are findings applicable in your practice/setting?
**SECTION 1: STUDY VALIDITY**

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>How well was this criterion addressed?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td>What were the key selection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(inclusion &amp; exclusion) criteria?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were they well defined?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were they replicable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were inclusion &amp; exclusion criteria</td>
<td></td>
</tr>
<tr>
<td></td>
<td>appropriate given study question?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>List important selection criteria; e.g.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>age group, gender, risk profile,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>medical history. Usually in Methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>section. There should be sufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>information in the paper (or referenced)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to allow the reader to theoretically</td>
<td></td>
</tr>
<tr>
<td></td>
<td>select a similar population</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Are the participants a relevant group</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to apply the study intervention to?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e.g. diagnostic tests are not very</td>
<td></td>
</tr>
<tr>
<td></td>
<td>helpful in people with a very high</td>
<td></td>
</tr>
<tr>
<td></td>
<td>probability of disease).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examples include: dosage of drugs,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>description of surgical procedure,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>number of faecal occult blood tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in a screening study, management in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison group (e.g. check what care</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the comparison (placebo group)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>receive).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random allocation of interventions is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fundamental to this type of study.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The method of randomisation should be</td>
<td></td>
</tr>
<tr>
<td></td>
<td>described (e.g. using a computer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>algorithm). If the description of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>randomisation method is poor, or the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>process used is not truly random</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e.g. allocation by date, alternating</td>
<td></td>
</tr>
<tr>
<td></td>
<td>between one group and another) or can</td>
<td></td>
</tr>
<tr>
<td></td>
<td>otherwise be seen as flawed, the study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>should be given a lower quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rating.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This is an important quality issue as</td>
<td></td>
</tr>
<tr>
<td></td>
<td>it has been shown that RCTs that</td>
<td></td>
</tr>
<tr>
<td></td>
<td>failure to conceal allocation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>typically lead to an overestimate of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the treatment effect. The ideal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>concealment process would involve an</td>
<td></td>
</tr>
<tr>
<td></td>
<td>independent group that registered</td>
<td></td>
</tr>
<tr>
<td></td>
<td>each new participant, determined the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>allocation and then informed the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>caregivers of the allocation. This can</td>
<td></td>
</tr>
<tr>
<td></td>
<td>be done relatively simply by phone or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fax or by using an automatic web-based</td>
<td></td>
</tr>
<tr>
<td></td>
<td>system. This reduces the chance of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>care giver influencing allocation.</td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Was randomisation successful: were exposure &amp; comparison groups similar at start of study?</td>
<td>This can be judged by examining the similarity between baseline characteristics of the groups in each arm of the study. Successful randomisation will produce similar groups. The study should report significant differences in the composition of the study groups in relation to gender mix, age, stage of disease (if appropriate), social background, ethnic origin, or comorbid conditions. Of note, statistically significant differences are not always important in practice; consider if the differences are likely to be meaningful given the study questions.</td>
<td></td>
</tr>
<tr>
<td>Were all participants analysed in groups to which randomised?</td>
<td>It is rarely that all participants allocated to the intervention group receive the intervention throughout the trial, or that all those in the comparison group do not. Participants may refuse treatment, or contra-indications may arise. If the comparability of groups through randomisation is to be maintained, however, patient outcomes should be analysed according to the group to which they were originally allocated irrespective of the treatment they actually received. (This is known as intention to treat analysis.) If analysis were not on an intention to treat basis, the study validity could be compromised. Bias can be quantified by attempting both “on-treatment” and “intention-to-treat” analyses.</td>
<td></td>
</tr>
<tr>
<td>Were participants, health workers, researchers blind to interventions?</td>
<td>Blinding can be carried out on up to three levels. Single blinding is where participants are unaware of which intervention they are receiving; in double blind studies neither the care giver nor the patient know which intervention is being given; in triple blind studies neither patients, care givers, nor those conducting the analysis are aware of which participants received which intervention. Most studies described as double blind studies are usually triple blind. The higher the level of blinding, the lower the risk of bias in the study. Although the nature of the intervention may unblind the allocation in some cases (e.g., Surgical trials) the allocation to intervention &amp; comparison groups should be blind.</td>
<td></td>
</tr>
<tr>
<td>Apart from study interventions, were groups treated equally?</td>
<td>If some participants received additional interventions, even if of a minor nature or consisting of advice and counselling rather than a physical intervention, this can introduce confounding and may invalidate the results. If there is unequal intervention (apart from the study intervention) the study results should be interpreted with caution and given a low quality rating.</td>
<td></td>
</tr>
<tr>
<td>Was compliance with interventions measured? Was it sufficient?</td>
<td>Compliance is often a problem in studies involving ongoing interventions such as daily medication or behaviour change. Pill counts and blood levels of drugs are examples of objective methods of measuring compliance, although self-reports are more common but less reliable.</td>
<td></td>
</tr>
<tr>
<td>What outcome measures were used? Well defined? Replicable?</td>
<td>Criteria for assessing outcomes such as diagnostic algorithms should be well described or referenced. It should be theoretically possible for the reader to replicate the process.</td>
<td></td>
</tr>
<tr>
<td>How complete was the follow up? How many dropouts were there?</td>
<td>The number of participants who drop out of a study is a concern if the number is very high. Conventionally, a 20% drop out rate is regarded as acceptable, but this depends on the study question. Some regard should be paid to why participants dropped out, as well as how many. It should be noted that the drop out rate may be expected to be higher in studies conducted over a long period of time. A higher drop out rate will normally lead to downgrading, rather than rejection of a study.</td>
<td></td>
</tr>
<tr>
<td>Was outcome assessment blind?</td>
<td>Ideally the assessors who measure &amp; record outcomes should be blind to participant allocation. This is more important for assessing outcomes that are not clear cut &amp; where knowledge of the intervention may influence the diagnostic assessment.</td>
<td></td>
</tr>
<tr>
<td>Was follow up time sufficiently long to detect important effects on outcomes of interest?</td>
<td>This is specific to the study intervention and outcomes assessed</td>
<td></td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY DESIGN:** How successfully do you think the study minimised bias? Very well = +, okay = Ø, poorly = -
## SECTION 2: STUDY RESULTS: MAGNITUDE & PRECISION OF EFFECTS

### What measures of occurrence (incidence / prevalence) & intervention effects (RR /RD /NNTs) were reported?

Some studies do not provide the relevant number of participants (D) in the exposure and comparison groups, the number of outcomes (N), the event rates / proportions with outcomes (N/D) in each study group, or the relevant measures of effect (RR, etc). If they are not reported or cannot be calculated, it is not possible to ascertain the accuracy of the effect estimates such as relative risk (RR), risk difference (RD) or mean differences (if continuous measures of outcome are given) and numbers needed to treat (NNT) – see definitions below in the Numbers Table below.

### What measures of precision of effects were reported (CIs, p-values)?

Either confidence intervals or p values for the estimates of effect should be reported or be possible to calculate.

### THE NUMBERS TABLE: OCCURRENCE, EFFECT ESTIMATES & PRECISION

<table>
<thead>
<tr>
<th>Outcomes* &amp; Time (T)</th>
<th>Exposure event rate (EER=N_e/D_e/T) or mean*</th>
<th>Comparison event rate (CER=N_c/D_c/T) or mean*</th>
<th>Relative Risk* (RR = EER/CER) ± (95% CI)</th>
<th>Risk difference or mean difference (RD = CER-EER) ± (95% CI)</th>
<th>Number Needed to Treat* (NNT = 1/RD) ± (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>complete</td>
<td>complete</td>
<td>complete</td>
<td>complete</td>
<td>complete</td>
<td>complete</td>
</tr>
</tbody>
</table>

*If outcomes continuous, can calculate means, mean differences, but not NNTs (don't usually calculate relative means)

\[ D_e = \text{Denominator (D) for exposure (intervention) group(s)}, \quad D_c = \text{D for comparison (control) group} \]

\[ N_e = \text{Numerator (N) for exposure group(s)}, \quad N_c = \text{N for comparison group} \]

### Could useful effect estimates (e.g. RR, RDs or mean differences, NNTs) be calculated? For benefits & harm?

These numbers should be reported or able to be calculated in the Numbers Table (above). To be useful, they need to have some meaning in practice. For example a change of one point on a visual analogue scale of symptoms may have little meaning unless clearly linked to a symptom description.

### What was the magnitude and direction of the effect estimates?(RR, RD, mean differences, NNTs)

These numbers are the bottom line of every study. All other appraisal questions relate to the validity, precision and applicability of these numbers. The importance of these numbers in practice depends on the group to which they are applied (see Applicability - next section).

### Was the precision of the effect estimates sufficient?

If 95% confidence intervals are wide and include the no effect point (e.g. RR=1, RD=0) or p-values are >> 0.05, then the precision of the estimates is likely to be poor & insufficient

### If no statistically significant effects detected, was there sufficient power?

If an effect estimate is not significantly different from no effect and the confidence interval is wide, the study is probably not large enough to detect a real difference between treatment and comparison groups (i.e. a low power study). A non significant effect associated with a tight CI suggests there is no effect and that the study has adequate power. Look for a power calculation in the methods section.

### If multi-centred RCT - were effects homogeneous between sites?

In multi-site studies, confidence in the results should be increased if it can be shown that similar results were obtained at the different participating centres.

### QUALITY OF STUDY RESULTS: Useful, precise +/-or sufficient power? Very good = +, okay = ∅, poor = -
### SECTION 3: STUDY APPLICABILITY

<table>
<thead>
<tr>
<th>Participants</th>
<th>Was the source population for participants well described?</th>
<th>If the source population is not well described it is not easy to assess the generalisability of the study findings to a target group or whether the study participants are a typical or atypical subset of the source population.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Were participants representative of source population?</td>
<td>As above</td>
</tr>
<tr>
<td></td>
<td>Can the relevance / similarity of the participants to a specific target group(s) be determined?</td>
<td>As above</td>
</tr>
<tr>
<td>Exposures &amp; Comparison</td>
<td>Were the characteristics of the study setting well described? e.g. rural, urban, inpatient, primary care</td>
<td>This helps determine the applicability of the interventions</td>
</tr>
<tr>
<td></td>
<td>Can the applicability of interventions be determined?</td>
<td>These should be described in some detail in the paper or referenced. It should be theoretically possible for the reader to replicate the process.</td>
</tr>
<tr>
<td></td>
<td>Can the relevance of the comparison group management be determined?</td>
<td>It is important to determine whether the comparison group receive no interventions (e.g. placebo only) or whether they receive “usual care.” As usual care may differ in different settings, it is important to determine what usual care involves</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Were all important outcomes considered: benefits? harms? costs?</td>
<td>Many studies only report data on benefits of interventions. Decisions to intervene need to balance benefits, harms and costs.</td>
</tr>
<tr>
<td></td>
<td>Are likely benefits greater than potential harms &amp; costs (or vice versa)? In what target group(s)?</td>
<td>The benefits, harms and costs of interventions may differ between different groups of people due to severity, co-morbidities etc. Ideally studies should describe the overall balance of risks, benefits and costs in different subgroups.</td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY APPLICABILITY:**

(a) Was it possible to determine applicability? Very well = +, okay = ∅, poorly = -  
(b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -
STUDIES OF THE ACCURACY OF DIAGNOSTIC TESTS:
(Relevant JAMA Users' Guide Numbers IIIA & B: references (5,6))

Introduction:

The most valid study design for assessing the accuracy of diagnostic tests is a non-experimental cross-sectional study that compares a test’s classification of a diagnosis with a reference standard’s classification, in a relevant study population.

The conceptual starting point of a diagnostic test study is to apply the reference (or gold) standard to determine which study participants have the disease or condition (D_E) - equivalent to exposed subgroup in other studies described in this module - and which participants don’t have it (D_C) - equivalent to the comparison subgroup. In many diagnostic test studies information on test results rather than the reference standard are collected first, however applying the reference standard remains the conceptual starting point.

The outcome of interest in a diagnostic test study is the test result (N). This may initially appear counter-intuitive as the outcome of interest in most studies is the disease. In the simplest example illustrated in the PECOT diagram (page 12), the test result is either positive (N+) or negative (N-). If the test is positive in someone with the condition (i.e. reference standard positive) then we use the symbol N+\textsubscript{E}; if the test is positive in someone without the condition (i.e. reference standard negative) then we use the symbol N+\textsubscript{C}. Similarly we can derive test negative categories N-\textsubscript{E} and N-\textsubscript{C}.

The “Outcomes” square in the PECOT diagram (page 12) is equivalent to the 2x2 table often described in texts and studies about diagnostic tests, however we have turned it on its side. For some reason most 2x2 tables have the reference standard results across the top of the table and the test results down the side of the table. We suggest you use our table format because when you draw the PECOT diagram, it is more obvious where the 2x2 table comes from.

The most useful single measure of accuracy of a diagnostic test is the likelihood ratio (LR). The LR is equivalent to a relative risk in other epidemiological studies and is calculated in the same way. However it is possible to calculate LRs for different test result (e.g. for a positive or a negative test result) – see boxes below for definitions.
These numbers can also be used to calculate sensitivity and specificity, which are the more traditionally described characteristics of a diagnostic test study. While they provide useful information (see definitions in boxes below), the LR has the advantage of combining sensitivity and specificity in one number. Moreover, as long as you remember that it is equivalent to a relative risk, it is easy to derive the LR from the PECOT diagram.

If you know the LRs for a test and you have an idea of the average disease prevalence in the group of patients you would apply the test to (known as the pre-test probability), you can also use a simple tool, called a likelihood ratio nomogram (reference 6, page 705 or reference 11, page 79), to estimate the probability that the patient has the disease once you have received the test result (known as the post-test probability of disease).

For those readers who feel more comfortable with sensitivity and specificity, the LR for a positive test is the sensitivity/(1 – specificity) and the LR for a negative test is (1- sensitivity/specificity).

The likelihood ratio for a positive test (LR+ve) is the ratio of: i.) the likelihood of a positive test in people with disease to: ii) the likelihood of a positive test in people without disease.

Likelihood Ratio for positive test (LR+ve) = \[
\frac{\text{number of N+E outcomes}}{\text{number in DE}} \div \frac{\text{number of N+C outcomes}}{\text{number in DC}}
\]

The likelihood ratio for a negative test (LR-ve) is the ratio of: i.) the likelihood of a negative test in people with disease to: ii) the likelihood of a negative test in people without disease.

Likelihood Ratio for negative test (LR-ve) = \[
\frac{\text{number of N-E outcomes}}{\text{number in DE}} \div \frac{\text{number of N-C outcomes}}{\text{number in DC}}
\]
The sensitivity of a test is its ability to detect people who have disease; it is the proportion of all people with disease who are identified as positive by the test.

\[
\text{Sensitivity} = \frac{\text{number of } N^+ \text{ outcomes}}{\text{number in } D_E}
\]

The specificity of a test is its ability to detect people who do not have disease; it is the proportion of all people without disease who are identified as negative by the test.

\[
\text{Specificity} = \frac{\text{number of } N^- \text{ outcomes}}{\text{number in } D_C}
\]

The effectiveness of a diagnostic test in reducing the occurrence of a health problem (i.e. the effectiveness of screening with a diagnostic test) is best evaluated in a randomised controlled trial (see appraisal guide for experimental studies).
### GATE Checklist for Diagnostic Test Studies (cross-sectional)

**Study author, title, publication reference**

**Key 5 part study question (PECOT). Was it focussed?**

---

#### SECTION 1: STUDY VALIDITY

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>How well was this criterion addressed?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What were the key selection (inclusion &amp; exclusion) criteria?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were they well defined? Were they replicable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were selection criteria appropriate given study question?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did selection lead to an appropriate spectrum of participants (like those assessed in practice)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exposure/Comparison</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What was the reference standard of diagnosis? Was it clearly defined, independent &amp; valid?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was the reference standard applied regardless of test result?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was the reference standard assessed blind to test result?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What tests were used? Were they well defined? Replicable?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was the test applied regardless of the reference standard result?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was test assessment blind to reference standard result?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was the test validated in a second, independent group?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY DESIGN:** How successfully do you think the study minimised bias? Very well = +, okay = ∅, poorly = -
### SECTION 2: STUDY RESULTS: ACCURACY & PRECISION

What measures of test accuracy were reported (sensitivity, specificity, LRs)?  
What measures of precision were reported (CIs, p-values)?

### THE NUMBERS TABLE: LIKELIHOODS, LIKELIHOOD RATIO ESTIMATES & PRECISION

<table>
<thead>
<tr>
<th>TEST RESULT (N[O])</th>
<th>IF REFERENCE STANDARD + VE: likelihood of a specific test result (N[O]) = L+ve = (N[O]E / DE)*</th>
<th>IF REFERENCE STANDARD - VE: likelihood of a specific test result N(O) = L–ve = (N(O)C / DC)*</th>
<th>LIKELIHOOD RATIO LR = L+ve / L-ve (similar to RR)</th>
<th>± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ve</td>
<td>= sensitivity (a/a+c)</td>
<td>= 1 - specificity (b/b+d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ve</td>
<td>= 1 - sensitivity (c/a+c)</td>
<td>= specificity (d/b+d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N[O] represents the generic test result (e.g. +ve, -ve, or a level of a test)*

Could useful measures of test accuracy (i.e. likelihood ratios [LR]) be calculated?  
What was the magnitude of the LR estimates?  
Was the precision of the LR estimates sufficient?  
If no statistically significant associations detected, was there sufficient power?

### QUALITY OF STUDY RESULTS: Useful, precise +/or sufficient power? Very good = +, okay = ∅, poor = -

### SECTION 3: STUDY APPLICABILITY

| Participants | Was the source population for participants well described?  
|--------------|-------------------------------------------------------------|
|              | Were participants representative of source population?  
|              | Can the relevance of the participants to a specific target group(s) be determined? |

<table>
<thead>
<tr>
<th>Exposures &amp; Comparison</th>
<th>Were the characteristics of the study setting well described? e.g. rural, urban, inpatient, primary care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Can sensible estimates of individual patient’s pre-test probabilities be determined from the study? (or from elsewhere?)</td>
</tr>
</tbody>
</table>

| Outcomes | Is the test available, affordable and reproducible in the target settings?  
|-----------|----------------------------------------------------------------------------------------------------------------------------|
|           | Will resulting post-test probabilities affect management and help patients? For which target group(s)?  

### QUALITY OF STUDY APPLICABILITY: (a) Was it possible to determine applicability? Very well = +, okay = ∅, poorly = - (b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -
**SECTION 1: STUDY VALIDITY**

**Participants**
- What were the key selection (inclusion & exclusion) criteria? Were they well defined? Were they replicable?
  - List important selection criteria; e.g. age group, gender, risk profile, medical history. Usually in Methods section. There should be sufficient information in the paper (or referenced) to allow the reader to theoretically select a similar population.

- Were selection criteria appropriate given study question?
  - Are the participants a relevant group to apply the study intervention to? (e.g. diagnostic tests are not very helpful in people with a very high probability of disease).

- Did selection lead to an appropriate spectrum of participants (like those assessed in practice)
  - Studies including participants with the range of common presentations of the target disorder and with commonly confused diagnoses are far more informative than studies that only include the extreme ends of the spectrum (florid cases & asymptomatic volunteers only).

**What was the reference standard of diagnosis? Was it clearly defined, independent & valid?**
- The validity of the study requires that there is an accepted, valid and replicable reference (gold) standard of diagnosis. Readers should give careful and critical consideration to the authors’ choice of a reference standard. In addition, those applying and interpreting the reference standard should ideally be unaware of the result of the test to avoid conscious or unconscious bias. This is not always possible, and can lead to over or under-interpretation of the reference standard results.

**Exposure / Comparison**
- Was the reference standard applied regardless of test result?
  - Reference standards are often not applied to participants with negative tests, particularly if invasive. An alternative is to follow these participants for an extended period to identify any false negative cases.

- Was the reference standard assessed blind to test result?
  - see above, reduces under and over-interpretation of reference standard

**Outcomes**
- What tests were used? Were they well defined? Replicable?
  - The methods for undertaking tests should be well described or referenced. It should be theoretically possible for the reader to replicate the process.

- Was the test applied regardless of the reference standard result?
  - All participants who are assessed with the reference standard should be tested. Untested participants are equivalent to cases “lost to follow-up”
<table>
<thead>
<tr>
<th>Was test assessment blind to reference standard result?</th>
<th>see above, reduces under and over-interpretation of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the test validated in a second, independent group?</td>
<td>As diagnostic tests are predictors, not explainers, of diagnoses, it is possible that the findings in a participant group are related to the characteristics of those selected. Demonstration of test accuracy in a second participant group increases confidence in the findings.</td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY DESIGN:** How successfully do you think the study minimised bias? Very well = +, okay = ∅, poorly = -

**SECTION 2: STUDY RESULTS: ACCURACY & PRECISION**

What measures of test accuracy were reported (sensitivity, specificity, LR)?

Some studies do not provide the relevant number of participants (D) in the study population who were assessed using the reference standard, the numbers who were tested (N), the proportions with various test results (N/D) in each reference stand group, or the relevant measures of test accuracy. If they are not reported or cannot be calculated, it is not possible to ascertain the accuracy of the test(s) - see definitions below in the Numbers Table below.

What measures of precision were reported (CIs, p-values)?

Either confidence intervals or p values for sensitivity, specificity & LRs should be reported or be possible to calculate

**THE NUMBERS TABLE: LIKELIHOODS, LIKELIHOOD RATIO ESTIMATES & PRECISION**

<table>
<thead>
<tr>
<th>TEST RESULT (N[O])</th>
<th>IF REFERENCE STANDARD + VE: likelihood of a specific test result (N[O]) = L+ve = (N[O]E / DE)*</th>
<th>IF REFERENCE STANDARD - VE: likelihood of a specific test result (N[O]) = L–ve = (N[O]C / DC)*</th>
<th>LIKELIHOOD RATIO (similar to RR) LR = L+ve / L-ve ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ve</td>
<td>sensitivity (a/a+c) = 1-specificity (b/b+d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ve</td>
<td>1-sensitivity (c/a+c) = specificity (d/b+d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* N[O] represents the generic test result (e.g. +ve, -ve, or a level of a test)

Could useful measures of test accuracy (i.e. likelihood ratios [LR]) be calculated?

LRs should be reported or able to be calculated in the Numbers Table (above). If sensitivity & specificity are reported, it is possible to calculate LRs

What was the magnitude of the LR estimates?

These numbers are the bottom line of every study. All other appraisal questions relate to the validity, precision and applicability of these numbers. The importance of these numbers in practice depends on the group to which they are applied (see Applicability - next section).

Was the precision of the LR estimates sufficient?

If 95% confidence intervals are wide and include the no effect point (LR=1) or p-values are >> 0.05, then the precision of the estimates is likely to be poor & insufficient

If no statistically significant associations detected, was there sufficient power?

If an LR estimate is not significantly different from 1 and the confidence interval is wide, the study is probably not large enough to determine if the test is accurate (i.e. a low power study). A non significant LR associated with a tight CI suggests the test is not useful and that the study has adequate power. Look for a power calculation in the methods section.

**QUALITY OF STUDY RESULTS:** Useful, precise +/- sufficient power? Very good = +, okay = ∅, poor = -
### SECTION 3: STUDY APPLICABILITY

<table>
<thead>
<tr>
<th>Participants</th>
<th>Was the source population for participants well described?</th>
<th>If the source population is not well described it is not easy to assess the generalisability of the study findings to a target group or whether the study participants are a typical or atypical subset of the source population.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were participants representative of source population?</td>
<td>As above</td>
<td></td>
</tr>
<tr>
<td>Can the relevance of the participants to a specific target group(s) be determined?</td>
<td>As above</td>
<td></td>
</tr>
<tr>
<td>Were the characteristics of the study setting well described? e.g. rural, urban, inpatient, primary care</td>
<td>This helps determine the applicability of the test</td>
<td></td>
</tr>
<tr>
<td>Exposures &amp; Comparison</td>
<td>Can sensible estimates of individual patient’s pre-test probabilities be determined from the study? (or from elsewhere?)</td>
<td>The importance of a test depends to a large extent on the pre-test probability of the target condition (i.e. the prevalence of the condition) in the people to whom the test is applied in practice. This information is often difficult to find and readers often depend on the study to determine this.</td>
</tr>
<tr>
<td>Is the test available, affordable and reproducible in the target settings?</td>
<td>The reproducibility of a test may depend on the expertise of those performing and evaluating the test. Information on reproducibility and training in the study setting can help determine reproducibility in other settings.</td>
<td></td>
</tr>
<tr>
<td>Will resulting post-test probabilities affect management and help patients? For which target group(s)?</td>
<td>The post-test probabilities of the target condition (i.e. the probability of having the target condition if the test is positive or if the test is negative) depends on both the pre-test probability in the whole group tested and the test accuracy (LR). As pre-test probabilities are likely to differ between groups, the usefulness of a test will vary from group to group.</td>
<td></td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY APPLICABILITY:**

(a) Was it possible to determine applicability? Very well = +, okay = Ø, poorly = -
(b) Are findings applicable in your practice/setting? Very well = +, okay = Ø, poorly = -
Introduction:

- The structure of a non-experimental study investigating benefit, harm or causation is very similar to a controlled trial. The occurrence of health outcomes (prevalence or incidence) is measured in subgroups defined by specific exposures and comparisons, which may be interventions in some instances.

- The major difference between experimental and non-experimental studies is that the investigator controls the allocation of the exposures (or treatment) to participants in an experiment, whereas the investigator in a non-experimental study categorises participants into exposure and comparison subgroups after measuring factors (i.e. the exposures) that the study participants are exposed to. In other words, the non-experimental study investigator observes a “natural” experiment rather than conducting one.

- The main weakness of non-experimental studies is the potential for confounding. As exposure allocation is not controlled by the investigator, it is common to find differences between exposure and comparison subgroups other than the main exposures of interest, that also influence health outcomes. These differences are known as confounding factors causing a mixing of effects. For questions about the benefits or harm of therapy, experimental studies (particularly randomised controlled trials) are usually superior to non-experimental studies because of the large potential for confounding in the latter.

Longitudinal (or cohort) studies:

- Cohort studies are basically non-experimental versions of controlled trials and are undertaken to investigate the effects (both benefits and harms) of exposures. In a cohort study, participants are recruited into the study population, exposures are measured, and then participants are followed up over time to measure outcomes.

- As mentioned above, cohort studies are not the most appropriate study design for examining the effects of interventions, because the potential for confounding is typically greatest when people are selected or self-select the exposures of interest (particularly therapies or exposures requiring a conscious decision by the participant, such as taking leisure time physical activity). Nevertheless, cohort studies are often used to
investigate the effects of therapy because of other the shortcomings of experimental studies.

- Cohort studies can often be conducted in situations where controlled trials are not possible. In some situations a trial would be unethical (e.g. investigating the adverse effects of a dangerous exposure such as electromagnetic radiation or cigarette smoking). In addition, trials are often not feasible when the effect of exposure (e.g. cigarette smoking) takes many years to cause an outcome (e.g. lung cancer) or when the outcomes of interest are uncommon (e.g. asthma death) and very large numbers of study participants are required to identify sufficient outcomes.
- Non-experimental longitudinal studies are also the most appropriate design for investigating prognosis.

Cross-sectional studies:

- The cross-sectional study has an identical structure to the cohort study except that the exposures and outcomes are measured at the same time (i.e. cross-sectionally), whereas in a cohort study outcomes are typically measured after the exposure/s has been measured (i.e. longitudinally).
- Cross-sectional studies are the design of choice for assessing the prevalence of health-related outcomes in a target population. In such studies it is very important that the study population is representative of the target or source population of interest (i.e. the findings in the study population must be generalisable to the target population).
- Cross-sectional studies are also the design of choice for comparing diagnostic tests with a reference standard.
- Cross-sectional studies may be undertaken to investigate causal associations between exposures and outcomes, although they are not ideally suited for this purpose; especially if the outcomes of interest are acute events. As outcome measurements are made at one point in time in cross-sectional studies, many acute outcomes would be missed, particularly if they are either fatal (e.g. coronary death) or recovery occurs quickly and there are no lasting signs or symptoms of the event (e.g. asthma attacks).
- If the outcome of interest can affect the exposure of interest (e.g. a myocardial infarction may lower blood pressure), then it is not possible to validly investigate the association in a cross-sectional study, because the outcome (myocardial infarction) may be measured before the exposure (blood pressure) has been measured.
- It is therefore important to document whether the exposure was measured before the outcome occurred (i.e. check if the association is temporally correct).
As cohort studies and most cross-sectional studies are simply longitudinal and cross-sectional versions of the same study design, they are considered together in one appraisal guide.
GATE Checklist for Cohort & Cross-sectional Studies (causation or intervention, benefit or harm)

### Study author, title, publication reference

<table>
<thead>
<tr>
<th>Key 5 part study question (PECOT): Was it focussed?</th>
</tr>
</thead>
</table>

### Study Population

![Diagram of study population]

- **DE** = Denominator (D) for exposure (or intervention) group, **Dc** = D for comparison (control) group
- **NE** = Numerator (N) for exposure group, **NC** = N for comparison group

### SECTION 1: STUDY VALIDITY

<table>
<thead>
<tr>
<th>Evaluation criterion (NAXS = not applicable for cross-sectional studies)</th>
<th>How well was this criterion addressed?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td>What were the key selection (inclusion &amp; exclusion) criteria? Were they well defined? Were they replicable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were inclusion &amp; exclusion criteria appropriate given study question?</td>
<td></td>
</tr>
<tr>
<td><strong>Exposures &amp; Comparison</strong></td>
<td>What were the exposures (or interventions) &amp; comparison? Well defined? Replicable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was measurement of variables similar &amp; valid in all groups?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were exposure &amp; comparison groups similar at start of study except for study exposures?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>If not, were differences stratified / adjusted for in analyses?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were all participants analysed in groups to which initially assigned?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were participants, health workers, researchers blind to exposures?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apart from study exposures, were groups treated equally?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Were exposures remeasured during follow-up &amp; were there important changes? (NAXS)</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>What outcome measures were used? Well defined? Replicable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How complete was follow up? Was it sufficient? How many dropouts? (NAXS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was outcome assessment blind?</td>
<td></td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>Was follow up time sufficiently long to detect important effects on outcomes of interest? (NAXS)</td>
<td></td>
</tr>
</tbody>
</table>

### QUALITY OF STUDY DESIGN: How successfully do you think the study minimised bias? Very well = +, okay = ∅, poorly = -
SECTION 2: STUDY RESULTS: MAGNITUDE & PRECISION OF EFFECTS

What measures of occurrence (incidence / prevalence) & exposure effects (RR /RD /NNTs) were reported?

What measures of precision of effects were reported (CIs, p-values)?

THE NUMBERS TABLE: OCCURRENCE, EFFECT ESTIMATES & PRECISION

<table>
<thead>
<tr>
<th>Outcomes* &amp; Time (T)</th>
<th>Exposure event rate (EER=N_E/D_E/T) or mean*</th>
<th>Comparison event rate (CER=N_C/D_C/T) or mean*</th>
<th>Relative Risk* (RR = EER/CER) ± (95% CI)</th>
<th>Risk difference or mean difference (RD = CER-EER) ± (95% CI)</th>
<th>Number Needed to Treat* (NNT = 1/RD) ± (95% CI)</th>
</tr>
</thead>
</table>

* if outcomes continuous, can calculate means, mean differences, but not NNTs  (don’t usually calculate relative means)

\[ D_E = \text{Denominator (D) for exposure (intervention) group(s)}, D_C = \text{D for comparison (control) group} \]

\[ N_E = \text{Numerator (N) for exposure group(s)}, N_C = \text{N for comparison group} \]

Could useful effect estimates (e.g. RR, RDs or mean differences, NNTs) be calculated? For benefits & harm?

What was the magnitude and direction of the effect estimates?

Was the precision of the effect estimates sufficient?

If no statistically significant effects detected, was there sufficient power?

QUALITY OF STUDY RESULTS: Useful, precise +/-or sufficient power? Very good = +, okay = ∅, poor = -

SECTION 3: STUDY APPLICABILITY

Participants

Was the source population for participants well described?

Were participants representative of source population?

Can the relevance / similarity of the participants to a specific target group(s) be determined?

Exposures & Comparison

Were the characteristics of the study setting well described? e.g. rural, urban, inpatient, primary care

Can the applicability/relevance of exposures be determined?

Can the relevance of the comparison group be determined?

Outcomes

Were all important outcomes considered: benefits? harms? costs?

Are likely benefits greater than potential harms & costs (or vice versa)? In what target group(s)?

QUALITY OF STUDY APPLICABILITY: (a) Was it possible to determine applicability? Very well = +, okay = ∅, poorly = -  (b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -
STUDIES ABOUT PROGNOSIS

(Relevant JAMA Users’ Guide Number V: references (8))

Introduction:

• Prognosis describes the expected occurrence, or probability, of an outcome (either good or bad) in a person with a specified condition or set of characteristics. The standard prognostic study is a cohort study in which a group of people with a particular condition or set of characteristics is followed over a period of time. At the start of the period a range of factors that may influence outcomes are measured and outcomes are measured over the period.

• Factors demonstrated to predict outcomes in a prognostic study are known as prognostic factors. These factors are equivalent to the exposures and confounding factors in a cohort study. As prognostic factors do not have to be causal, confounders can be prognostic factors. Any factors that may have an important effect on the occurrence of outcomes should be measured and classified as potential prognostic factors.

• In prognostic studies it is particularly important that the study population is a well-described and representative sample from a relevant and recognisable group of people who have a specified condition or set of characteristics and are at a similar stage in the development of a disease or other health-related outcome. Sometimes the control group in a randomised trial is used to assess prognosis, however this may be quite inappropriate if the controls are a highly selected, unrepresentative subset of usual patients.
## GATE Checklist for Prognostic Studies

### Study Population

- **Source pop:**
  - **Participant selection:**

### Exposure

- **DE:** Denominator (D) for exposure (or intervention) group
- **DC:** D for comparison (control) group

### Outcome

- **NE:** Numerator (N) for exposure group
- **NC:** N for comparison group

### SECTION 1: STUDY VALIDITY

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>How well was this criterion addressed?</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What were the key selection (inclusion &amp; exclusion) criteria?</td>
<td>Well defined? Replicable?</td>
<td></td>
</tr>
<tr>
<td>Were inclusion &amp; exclusion criteria appropriate given study question?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were participants at a common point in the course of their disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exposures &amp; Comparison</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What were the prognostic groups?</td>
<td>Well defined? Replicable?</td>
<td></td>
</tr>
<tr>
<td>Was measurement of variables similar &amp; valid in all groups?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were different prognostic groups similar at start of study except for study prognostic factors?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If not, were differences stratified / adjusted for in analyses?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were all participants analysed in groups to which initially assigned?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were participants, health workers, researchers blind to prognostic factors?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were groups treated equally?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were prognostic factors remeasured during follow-up &amp; were there important changes?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What outcome measures were used?</td>
<td>Well defined? Replicable?</td>
<td></td>
</tr>
<tr>
<td>How complete was follow up? Was it sufficient? How many dropouts?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was outcome assessment blind?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was follow up time sufficiently long to detect important prognostic factors?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**QUALITY OF STUDY DESIGN:** How successfully do you think the study minimised bias? Very well = +, okay = ∅, poorly = -
SECTION 2: STUDY RESULTS: MAGNITUDE & PRECISION OF PROGNOSIS

What measures of prognosis (over what time) & differences between groups (e.g. RR /RD) were reported?

What measures of precision were reported (CIs, p-values)?

THE NUMBERS TABLE: PROGNOSIS & PRECISION

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Time of follow-up</th>
<th>Prognostic group E event rate (EER=NE/DE/T) or mean* ± (95% CI)</th>
<th>Prognostic group C event rate (CER=NC/DC/T) or mean* ± (95% CI)</th>
<th>Relative Risk* (RR = EER/CER) ± (95% CI)</th>
</tr>
</thead>
</table>

* if outcomes continuous, can calculate means & mean differences

Д = Denominator (D) for exposure group(s), DC = D for comparison group
N = Numerator (N) for exposure group(s), NC = N for comparison group

Could useful estimates of prognosis be calculated? (i.e. events/person/time)

What was the magnitude and direction of the prognostic estimates?

Was the precision of the prognostic estimates sufficient?

If no statistically significant estimates detected, was there sufficient power?

QUALITY OF STUDY RESULTS: Useful, precise +/or sufficient power? Very good = +, okay = ∅, poor = -

SECTION 3: STUDY APPLICABILITY

Participants

Was the source population for participants well described?

Were participants representative of source population?

Can the relevance / similarity of the participants to a specific target group(s) be determined?

Exposures & Comparison

Were the characteristics of the study setting well described? e.g. rural, urban, inpatient, primary care

Can the applicability/relevance of prognostic factors be determined?

Outcomes

Were all important outcomes considered: benefits? harms?

Would the prognostic information have an important impact on patient or practitioner decisions? In what target group(s)?

QUALITY OF STUDY APPLICABILITY: (a) Was it possible to determine applicability? Very well = +, okay = ∅, poorly = - (b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -
SYSTEMATIC REVIEWS OF STUDIES OF THERAPY OR OTHER INTERVENTIONS:

(Relevant JAMA Users' Guide Number VI: references (9))

Introduction:

- Few individual trials are large enough or include a wide enough range of relevant groups of people to definitively answer questions about therapy or prevention.

- Systematic reviews of trials attempt to summarise all relevant trials that have addressed a particular question, by following a set of systematic rules to ensure both the completeness of the review and the validity of the findings.

- By bringing together all relevant trials it is possible to determine more precise estimates of effect than that available in any single trial. Moreover if all trials of good quality show similar effect estimates, the reviewer can be more confident in the findings.

- All trials included in a systematic review should first be assessed using the critical appraisal guide for experimental studies. Ideally evidence-based decisions should be based on systematic reviews of evidence rather than individual studies.

- It is also possible to undertake systematic reviews of non-experimental studies.

- It is beyond the scope of this guide to describe the mathematical approach to synthesising data from different studies. This is covered in many papers and texts.
## GATE Checklist for Systematic Reviews of Randomised Controlled Trials

### Review author, title, reference

<table>
<thead>
<tr>
<th>Key 5 part Review question (PECOT). Was it focused?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### SECTION 1: REVIEW VALIDITY

#### Appraised by:

#### Evaluation criterion

<table>
<thead>
<tr>
<th>How well was this criterion addressed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality ( \sqrt{?\times} )</td>
</tr>
</tbody>
</table>

#### Participant studies

- What were the key (inclusion & exclusion) criteria for selecting studies? Were they well defined? Were they replicable?
- Were inclusion & exclusion criteria appropriate given study question?
- Was the search for studies comprehensive? Complete?
- How was the validity of individual studies assessed?

#### Exposures & Comparison

- What were the exposures (interventions) & comparison? Well defined? Replicable? Similar from study to study?
- Was assignment to groups randomised in all studies? Was randomisation concealed?
- Was randomisation successful in all studies? If not, how were potential confounders dealt with?
- Was intention-to-treat analyses used in all studies?

#### Outcomes

- What outcome measures were used? Well defined? Replicable? Similar in all studies?
- Was follow-up sufficient in all studies?

#### Time

- Was follow up time sufficiently long to detect important effects on outcomes of interest in all studies?

### QUALITY OF REVIEW DESIGN:

How successfully do you think the Review minimised bias? Very well = +, okay = \( \varnothing \), poorly = -

---

\[ D_E = \text{Denominator (D) for exposure (intervention) groups}, \quad D_C = \text{D for comparison (control) groups} \]

\[ N_E = \text{Numerator (N) for exposure groups}, \quad N_C = \text{N for comparison groups} \]
### SECTION 2: REVIEW RESULTS: MAGNITUDE & PRECISION OF EFFECTS

<table>
<thead>
<tr>
<th>What measures of occurrence (incidence / prevalence) &amp; exposure (intervention) effects (e.g. RR or ORs) were reported for each study?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What measures of precision of effects were reported (CIs, p-values)?</td>
</tr>
</tbody>
</table>

### THE NUMBERS TABLE: OCCURRENCE, EFFECT ESTIMATES & PRECISION

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Quality rating</th>
<th>Number of participants, brief description of selection criteria</th>
<th>Exposures &amp; Comparison</th>
<th>Outcomes &amp; time</th>
<th>Exposure event rate (EER=NE/DE/T) or mean*</th>
<th>Comparison event rate (CER=NC/DC/T) or mean*</th>
<th>Relative Risk* (RR = EER/CER) or Odds Ratio (±95% CI)</th>
</tr>
</thead>
</table>

* If outcomes continuous, can calculate means, mean differences.  
  - DE = Denominator (D) for exposure (intervention) group(s), DC = D for comparison (control) group.  
  - NE = Numerator (N) for exposure group(s), NC = N for comparison group  

Could useful summary effect estimates (e.g. RR, ORs or mean differences) be calculated? For benefits & harm?  
What was the magnitude and direction of the summary effect estimates?  
Was the precision of the summary effect estimates sufficient?  
If no statistically significant effects detected, was there sufficient power?  
Were the effect estimates consistent from study to study?  

**QUALITY OF REVIEW RESULTS:** Useful, precise +/- or sufficient power? Very good = +, okay = ∅, poor = -  

### SECTION 3: REVIEW APPLICABILITY

| Participants | Were source populations for participants described?  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Can the relevance / similarity of the participants in the Review to a specific target group(s) be determined?</td>
</tr>
</tbody>
</table>
| Exposures & Comparison | Were the characteristics of the study settings well described? e.g. rural, urban, inpatient, primary care  
|               | Can the applicability of exposures (interventions) be determined?  
|               | Can the relevance of the comparison groups management be determined?  
| Outcomes | Were all important outcomes considered: benefits? harms? costs?  
|               | Are likely benefits greater than potential harms & costs (or vice versa)? In what target group(s)? |

**QUALITY OF REVIEW APPLICABILITY:**  
(a) Was it possible to determine applicability? Very well = +, okay = ∅, poorly = -  
(b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -  

---

29
REFERENCES:


3. Guyatt GH, Sackett DL, Cook DJ. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993;270:2598-2601.

4. Guyatt GH, Sackett DL, Cook DJ. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? JAMA 1994;271:59-63.


Clinical question: Treatment 1 vs Treatment 2 or placebo or no treatment

<table>
<thead>
<tr>
<th>Study authors and year</th>
<th>Study Design</th>
<th>Participants</th>
<th>Exposure/Comparison</th>
<th>Outcomes</th>
<th>Results</th>
<th>Quality Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EER</td>
<td>CER</td>
<td>RR</td>
</tr>
</tbody>
</table>


# Considered Judgement Form

<table>
<thead>
<tr>
<th>Key question:</th>
<th>Evidence table ref:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Volume of evidence</strong>&lt;br&gt;Comment here on any issues concerning the quantity of evidence available on this topic and its methodological quality.</td>
<td></td>
</tr>
<tr>
<td><strong>2. Consistency</strong>&lt;br&gt;Comment here on the degree of consistency demonstrated by the available of evidence. Where there are conflicting results, indicate how the group formed a judgement as to the overall direction of the evidence</td>
<td></td>
</tr>
<tr>
<td><strong>3. Applicability</strong>&lt;br&gt;Comment here on the extent to which the evidence is directly applicable in the New Zealand setting. Comment here on how reasonable it is to generalise from the results of the studies used as evidence to the target population for this guideline.</td>
<td></td>
</tr>
<tr>
<td><strong>4. Clinical impact</strong>&lt;br&gt;Comment here on the potential clinical impact that the intervention in question might have – e.g. size of patient population; magnitude of effect; relative benefit over other management options; resource implications; balance of risk and benefit.</td>
<td></td>
</tr>
<tr>
<td><strong>5. Other factors</strong>&lt;br&gt;Indicate here any other factors that you took into account when assessing the evidence base.</td>
<td></td>
</tr>
</tbody>
</table>
6. Evidence statement

Please summarise the development group's synthesis of the evidence relating to this key question, taking all the above factors into account, and indicate the evidence level which applies.

<table>
<thead>
<tr>
<th>Evidence level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

7. Recommendation

What recommendation(s) does the guideline development group draw from this evidence? Please indicate the grade of recommendation(s) and any dissenting opinion within the group.

<table>
<thead>
<tr>
<th>Grade of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
APPRAISAL OF GUIDELINES FOR RESEARCH & EVALUATION
(AGREE) INSTRUMENT

The AGREE Collaboration

June 2001
COPYRIGHT AND REPRODUCTION

This document is the product of an international collaboration. It may be reproduced and used for educational purposes, quality assurance programmes and critical appraisal of clinical practice guidelines. It may not be used for commercial purposes or product marketing. Approved non-English language versions of the AGREE Instrument are being prepared, and must be used where available. Offers of assistance in translation into other languages are welcome, provided they conform to the protocol set out by the AGREE Collaboration.

Disclaimer
The AGREE Instrument is a generic tool designed primarily to help guideline developers and users assess the methodological quality of clinical practice guidelines. The authors do not take responsibility for the improper use of the AGREE Instrument.

© St George's Hospital Medical school, London, June 2001

ISBN


Funding:
The development of the AGREE Instrument was funded by a grant from the EU BIOMED2 Programme (BMH4-98-3669)

For further information about the instrument contact:

Françoise Cluzeau
Email: f.cluzeau@sghms.ac.uk

or

Jako Burgers
Email: j.burgers@hsv.kun.nl
INTRODUCTION

Purpose of the AGREE Instrument

The purpose of the Appraisal of Guidelines Research & Evaluation (AGREE) Instrument is to provide a framework for assessing the quality of clinical practice guidelines.

Clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances”\(^1\). Their purpose is "to make explicit recommendations with a definite intent to influence what clinicians do"\(^2\).

By quality of clinical practice guidelines we mean the confidence that the potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice. This process involves taking into account the benefits, harms and costs of the recommendations, as well as the practical issues attached to them. Therefore the assessment includes judgements about the methods used for developing the guidelines, the content of the final recommendations, and the factors linked to their uptake.

The AGREE Instrument assesses both the quality of the reporting, and the quality of some aspects of recommendations. It provides an assessment of the predicted validity of a guideline, that is the likelihood that it will achieve its intended outcome. It does not assess the impact of a guideline on patients’ outcomes.

Most of the criteria contained in the AGREE Instrument are based on theoretical assumptions rather than on empirical evidence. They have been developed through discussions between researchers from several countries who have extensive experience and knowledge of clinical guidelines. Thus the AGREE Instrument should be perceived as reflecting the current state of knowledge in the field.

Which guidelines can be appraised with the AGREE Instrument?

The AGREE Instrument is designed to assess guidelines developed by local, regional, national or international groups or affiliated governmental organisations. These include:

1. New guidelines
2. Existing guidelines
3. Updates of existing guidelines

The AGREE Instrument is generic and can be applied to guidelines in any disease area including those for diagnosis, health promotion, treatment or interventions. It is suitable for guidelines presented in paper or electronic format.


Who can use the AGREE Instrument?

The AGREE Instrument is intended to be used by the following groups:

i) By policy makers to help them decide which guidelines could be recommended for use in practice. In such instances, the instrument should be part of a formal assessment process.

ii) By guideline developers to follow a structured and rigorous development methodology and as a self-assessment tool to ensure that their guidelines are sound.

iii) By health care providers who wish to undertake their own assessment before adopting the recommendations

iv) By educators or teachers to help enhance critical appraisal skills amongst health professionals.

Key references

The following sources have been used for developing the AGREE Instrument criteria.


INSTRUCTIONS FOR USE
Please read the following instructions carefully
before using the AGREE Instrument

1. **Structure and content of the AGREE Instrument**
   AGREE consists of 23 key items organised in six domains. Each domain is intended to capture a separate dimension of guideline quality.

   **Scope and purpose** (items 1-3) is concerned with the overall aim of the guideline, the specific clinical questions and the target patient population.

   **Stakeholder involvement** (items 4-7) focuses on the extent to which the guideline represents the views of its intended users.

   **Rigour of development** (items 8-14) relates to the process used to gather and synthesise the evidence, the methods to formulate the recommendations and to update them.

   **Clarity and presentation** (items 15-18) deals with the language and format of the guideline.

   **Applicability** (items 19-21) pertains to the likely organisational, behavioural and costs implications of applying the guideline.

   **Editorial independence** (items 22-23) is concerned with the independence of the recommendations and acknowledgement of possible conflict of interest from the guideline development group.

2. **Documentation**
   Appraisers should attempt to identify all information about the guideline development process prior to appraisal. This information may be contained in the same document as the recommendations or it may be summarised in a separate technical report, in published papers or in policy reports (e.g. guideline programmes). We recommend that you read the guideline and its accompanying documentation fully before you start the appraisal.

3. **Number of appraisers**
   We recommend that each guideline is assessed by at least two appraisers and preferably four as this will increase the reliability of the assessment.

4. **Response scale**
   Each item is rated on a 4-point scale ranging from 4 "Strongly Agree" to 1 "Strongly Disagree", with two mid points: 3 "Agree" and 2 "Disagree". The scale measures the extent to which a criterion (item) has been fulfilled.

   - If you are confident that the criterion has been fully met then you should answer "Strongly Agree".

   - If you are confident that the criterion has not been fulfilled at all or if there is no information available then you should answer "Strongly Disagree".

   - If you are unsure that a criterion has been fulfilled, for example because the information is unclear or because only some of the recommendations fulfil the
criterion, then you should answer "Agree" or "Disagree", depending on the extent to which you think the issue has been addressed.

5. User Guide
We have provided additional information in the User Guide adjacent to each item. This information is intended to help you understand the issues and concepts addressed by the item. Please read this guidance carefully before giving your response.

6. Comments
There is a box for comments next to each item. You should use this box to explain the reasons for your responses. For example, you may "Strongly Disagree" because the information is not available, the item is not applicable, or the methodology described in the information provided is unsatisfactory. Space for further comments is provided at the end of the instrument.

7. Domain scores
Domain scores can be calculated by summing up all the scores of the individual items in a domain and by standardising the total as a percentage of the maximum possible score for that domain.

<table>
<thead>
<tr>
<th></th>
<th>Item1</th>
<th>Item2</th>
<th>Item 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraiser 1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Appraiser 2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Appraiser 3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Appraiser 4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>13</td>
<td>14</td>
<td>36</td>
</tr>
</tbody>
</table>

Maximum possible score = 4 (strongly agree) * 3 (items) * 4 (appraisers) = 48
The standardised domain score will be: (36/48) x 100 = 75%

Note:
The six domain scores are independent and should not be aggregated into a single quality score. Although the domain scores may be useful for comparing guidelines and will inform the decision as to whether or not to use or to recommend a guideline, it is not possible to set thresholds for the domain scores to mark a "good" or "bad" guideline.

8. Overall assessment
A section for overall assessment is included at the end of the instrument. This contains a series of options 'Strongly recommend', 'Recommend (with provisos or alterations)', 'Would not recommend' and 'Unsure'. The overall assessment requires the appraiser to make a judgement as to the quality of the guideline, taking each of the appraisal criteria into account.
1. The overall objective(s) of the guideline is(are) specifically described.

   Strongly Agree 4 3 2 1 Strongly Disagree

2. The clinical question(s) covered by the guideline is(are) specifically described.

   Strongly Agree 4 3 2 1 Strongly Disagree

3. The patients to whom the guideline is meant to apply are specifically described.

   Strongly Agree 4 3 2 1 Strongly Disagree
SCOPE AND PURPOSE

1. This deals with the potential health impact of a guideline on society and populations of patients. The overall objective(s) of the guideline should be described in detail and the expected health benefits from the guideline should be specific to the clinical problem. For example specific statements would be:
   - Preventing (long term) complications of patients with diabetes mellitus;
   - Lowering the risk of subsequent vascular events in patients with previous myocardial infarction;
   - Rational prescribing of antidepressants in a cost-effective way.

2. A detailed description of the clinical questions covered by the guideline should be provided, particularly for the key recommendations (see item 15). Following the examples provided in question 1:
   - How many times a year should the Hb1Ac be measured in patients with diabetes mellitus?
   - What should the daily aspirin dosage for patients with proven acute myocardial infarction be?
   - Are selective serotonin reuptake inhibitors (SSRIs) more cost-effective than tricyclic antidepressants (TCAs) in treatment of patients with depression?

3. There should be a clear description of the target population to be covered by a guideline. The age range, sex, severity, clinical description, comorbidity may be provided. For example:
   - A guideline on the management of diabetes mellitus only includes patients with non-insulin dependent diabetes mellitus and excludes patients with cardiovascular comorbidity.
   - A guideline on the management of depression only includes patients with major depression, according to the DSM-IV criteria, and excludes patients with psychotic symptoms and children.
   - A guideline on screening of breast cancer only includes women, aged between 50 and 70 years, with no history of cancer and with no family history of breast cancer.
4. The guideline development group includes individuals from all the relevant professional groups.

5. The patients’ views and preferences have been sought.

6. The target users of the guideline are clearly defined.

7. The guideline has been piloted among end users.
4. This item refers to the professionals who were involved at some stage of the development process. This may include members of the steering group, the research team involved in selecting and reviewing/rating the evidence and individuals involved in formulating the final recommendations. This item excludes individuals who have externally reviewed the guideline (see Item 11). Information about the composition, discipline and relevant expertise of the guideline development group should be provided.

5. Information about patients' experiences and expectations of health care should inform the development of clinical guidelines. There are various methods for ensuring that patients' perspectives inform guideline development. For example, the development group could involve patients' representatives, information could be obtained from patient interviews, literature reviews of patients' experiences could be considered by the group. There should be evidence that this process has taken place.

6. The target users should be clearly defined in the guideline, so they can immediately determine if the guideline is relevant to them. For example, the target users for a guideline on low back pain may include general practitioners, neurologists, orthopaedic surgeons, rheumatologists and physiotherapists.

7. A guideline should have been pre-tested for further validation amongst its intended end users prior to publication. For example, a guideline may have been piloted in one or several primary care practices or hospitals. This process should be documented.
RIGOUR OF DEVELOPMENT

8. Systematic methods were used to search for evidence.

Strongly Agree 4 3 2 1 Strongly Disagree

9. The criteria for selecting the evidence are clearly described.

Strongly Agree 4 3 2 1 Strongly Disagree

10. The methods used for formulating the recommendations are clearly described.

Strongly Agree 4 3 2 1 Strongly Disagree

11. The health benefits, side effects and risks have been considered in formulating the recommendations.

Strongly Agree 4 3 2 1 Strongly Disagree
8. Details of the strategy used to search for evidence should be provided including search terms used, sources consulted and dates of the literature covered. Sources may include electronic databases (e.g. MEDLINE, EMBASE, CINAHL), databases of systematic reviews (e.g. the Cochrane Library, DARE), handsearching journals, reviewing conference proceedings and other guidelines (e.g. the US National Guideline Clearinghouse, the German Guidelines Clearinghouse).

9. Criteria for including / excluding evidence identified by the search should be provided. These criteria should be explicitly described and reasons for including and excluding evidence should be clearly stated. For example, guideline authors may decide to only include evidence from randomised clinical trials and to exclude articles not written in English.

10. There should be a description of the methods used to formulate the recommendations and how final decisions were arrived at. Methods include for example, a voting system, formal consensus techniques (e.g. Delphi, Glaser techniques). Areas of disagreement and methods of resolving them should be specified.

11. The guideline should consider health benefits, side effects, and risks of the recommendations. For example, a guideline on the management of breast cancer may include a discussion on the overall effects on various final outcomes. These may include: survival, quality of life, adverse effects, and symptom management or a discussion comparing one treatment option to another. There should be evidence that these issues have been addressed.
12. There is an explicit link between the recommendations and the supporting evidence.

Strongly Agree | 4 | 3 | 2 | 1 | Strongly Disagree

13. The guideline has been externally reviewed by experts prior to its publication.

Strongly Agree | 4 | 3 | 2 | 1 | Strongly Disagree

14. A procedure for updating the guideline is provided.

Strongly Agree | 4 | 3 | 2 | 1 | Strongly Disagree

CLARITY AND PRESENTATION

15. The recommendations are specific and unambiguous.

Strongly Agree | 4 | 3 | 2 | 1 | Strongly Disagree
12. There should be an explicit link between the recommendations and the evidence on which they are based. Each recommendation should be linked with a list of references on which it is based.

13. A guideline should be reviewed externally before it is published. Reviewers should not have been involved in the development group and should include some experts in the clinical area and some methodological experts. Patients’ representatives may also be included. A description of the methodology used to conduct the external review should be presented, which may include a list of the reviewers and their affiliation.

14. Guidelines need to reflect current research. There should be a clear statement about the procedure for updating the guideline. For example, a timescale has been given, or a standing panel receives regularly updated literature searches and makes changes as required.

CLARITY AND PRESENTATION

15. A recommendation should provide a concrete and precise description of which management is appropriate in which situation and in what patient group, as permitted by the body of evidence.

- An example of a specific recommendation is: Antibiotics have to be prescribed in children of two years or older with acute otitis media if the complaints last longer than three days or if the complaints increase after the consultation despite adequate treatment with painkillers; in these cases amoxycillin should be given for 7 days (supplied with a dosage scheme).

- An example of a vague recommendation is: Antibiotics are indicated for cases with an abnormal or complicated course.

However, evidence is not always clear cut and there may be uncertainty about the best management. In this case the uncertainty should be stated in the guideline.
16. The different options for management of the condition are clearly presented.

Strongly Agree 4 3 2 1 Strongly Disagree

17. Key recommendations are easily identifiable

Strongly Agree 4 3 2 1 Strongly Disagree

18. The guideline is supported with tools for application.

Strongly Agree 4 3 2 1 Strongly Disagree

19. The potential organisational barriers in applying the recommendations have been discussed.

Strongly Agree 4 3 2 1 Strongly Disagree
16. A guideline should consider the different possible options for screening, prevention, diagnosis or treatment of the condition it covers. These possible options should be clearly presented in the guideline. For example, a recommendation on the management of depression may contain the following alternatives:
   a. Treatment with tricyclic antidepressants
   b. Treatment with SSRI
   c. Psychotherapy
   d. Combination of pharmacological and psychological therapy

17. Users should be able to find the most relevant recommendations easily. These recommendations answer the main clinical questions that have been covered by the guideline. They can be identified in different ways. For example, they can be summarised in a box, typed in bold, underlined or presented as flow charts or algorithms.

18. For a guideline to be effective it needs to be disseminated and implemented with additional materials. These may include for example, a summary document, or a quick reference guide, educational tools, patients' leaflets, computer support, and should be provided with the guideline.

19. Applying the recommendations may require changes in the current organisation of care within a service or a clinic which may be a barrier to using them in daily practice. Organisational changes that may be needed in order to apply the recommendations should be discussed. For example:
   i. A guideline on stroke may recommend that care should be co-ordinated through stroke units and stroke services.
   ii. A guideline on diabetes in primary care may require that patients are seen and followed up in diabetic clinics.
20. The potential cost implications of applying the recommendations have been considered.

Comments

Strongly Agree 4 3 2 1 Strongly Disagree

21. The guideline presents key review criteria for monitoring and/or audit purposes.

Comments

Strongly Agree 4 3 2 1 Strongly Disagree

EDITORIAL INDEPENDENCE

22. The guideline is editorially independent from the funding body.

Comments

Strongly Agree 4 3 2 1 Strongly Disagree

23. Conflicts of interest of guideline development members have been recorded.

Comments

Strongly Agree 4 3 2 1 Strongly Disagree
20. The recommendations may require additional resources in order to be applied. For example, there may be a need for more specialised staff, new equipment, expensive drug treatment. These may have cost implications for health care budgets. There should be a discussion of the potential impact on resources in the guideline.

21. Measuring the adherence to a guideline can enhance its use. This requires clearly defined review criteria that are derived from the key recommendations in the guideline. These should be presented. Examples of review criteria are:
- the HbA1c should be < 8.0%
- the level of diastolic blood pressure should be < 95 mmHg
- If complaints of acute otitis media lasts longer than three days amoxicillin should be prescribed

| EDITORIAL INDEPENDENCE |

22. Some guidelines are developed with external funding (e.g. Government funding, charity organisations, pharmaceutical companies). Support may be in the form of financial contribution for the whole development, or for parts of it, e.g. printing of the guidelines. There should be an explicit statement that the views or interests of the funding body have not influenced the final recommendations.

Please note: If it is stated that a guideline was developed without external funding, then you should answer 'Strongly Agree'.

23. There are circumstances when members of the development group may have conflicts of interests. For example, this would apply to a member of the development group whose research on the topic covered by the guideline is also funded by a pharmaceutical company. There should be an explicit statement that all group members have declared whether they have any conflict of interest.
Overall Assessment

Would you recommend these guidelines for use in practice?

- Strongly recommend
- Recommend (with provisos or alterations)
- Would not recommend
- Unsure
NZGG uses the AGREE Instrument to appraise guidelines. This template picks out the main topics identified by AGREE and those of NZGG guideline developers.

<table>
<thead>
<tr>
<th>Title Page for the Full Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Practice Evidence-based Guideline - Title and Date</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table of Contents</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Scope and Purpose of the Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>The objective of this guideline is to provide an evidence-based summary of the benefits, risks and contraindications to....</td>
</tr>
</tbody>
</table>

(Short direct statement about what the guideline is trying to achieve)

The consumer group to whom the guideline is meant to apply is described.

<table>
<thead>
<tr>
<th>About the Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
</tr>
</tbody>
</table>

Outline of the grading methodology used

Gaps between current practice and evidence

Guideline development process including description of the search methodology, the criteria for selecting the evidence, and the approach for formulating recommendations.

List of organisations involved in the peer review of the guideline

Endorsements and approvals – List the range of organisations that approve and or endorse the guideline. (Note – the organisations will want to see a copy of the final guideline before they sign off their endorsement or agreement |
– this can be quite time consuming). NZGG endorsement should also be sought.

Guideline team – provide the names and backgrounds of the people involved on the team

Declaration of Competing Interests / Conflicts of Interest – describe any funding, sponsorship, employment relationships, memberships or affiliations that may be perceived to have had some influence on group participants.

Acknowledgements – anything or anyone else that should be mentioned.

### Introduction

A general introduction to the issue covering current patterns of use or service in NZ and overseas and other related issues relevant to the delivery or effective treatment, care or service.

### Summary

Key Messages – dot point the main messages that arise from the guideline. Keep this clear, unambiguous and direct.

Summarise other detailed points under various headings.

The Summary of Evidence Table in the HRT guideline has been well received. It is a tick box matrix that clearly indicates those indications for and contraindications against the use of HRT in post menopausal women.

<table>
<thead>
<tr>
<th>Symptom / Risk Profile</th>
<th>Recommended</th>
<th>May be considered</th>
<th>Not recommended</th>
<th>Contraindicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot flushes</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal history of breast cancer</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous coronary disease</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

However, it may not be useful for all guideline topic areas.
Clinical Questions

Provide narrative evidence summary for each of the clinical questions asked and answered by the guideline. Health benefits, side effects and risks should be identified for each question.

Conclusions should be graded and formatted as follows

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>Grade of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>Opinion</td>
<td></td>
</tr>
</tbody>
</table>

There should be an explicit link between the recommendations and supporting evidence.

Balance Sheet

Inclusion of the balance sheet is optional. The cost implications of applying the recommendations should be made clear in the guideline.

Implementation

Suggestions for the dissemination and implementation of the guideline, the applicability of the guideline, and potential organizational barriers in applying the recommendations are identified (eg product x is not currently licensed for distribution in NZ)

Evidence Tables
**Alogorithms**

Standard Presentation for Algorithms Standardised graphical layout should be used purposes (see the example below). The algorithm should identify the management flow within the guideline. Key points giving further information are linked to the charts.

Instructions for Use
The Guideline Algorithm is created to facilitate the use of the practice guideline All links from the Algorithm to additional information or supporting evidence are underlined.

**References**

**Appendices**

**Glossary**

Note: where possible terminology should be standardised. Both within the guideline and between guidelines (eg defining RCT)
Template for Summary Guideline

Ideally this Summary should be able to be read at a glance and would fit onto 2 sides of an A4 sheet.

If this is not possible – then the most important messages and most frequently used/ most helpful advice should be clearly displayed on the outside pages.

Key Points and Recommendations

Algorithms

Standard Presentation for Algorithms Standardised graphical layout should be used purposes (see the example below). The algorithm should identify the management flow within the guideline. Key points giving further information are linked to the charts.

Instructions for Use
The Guideline Algorithm is created to facilitate the use of the practice guideline All links from the Algorithm to additional information or supporting evidence are underlined.

Other decision tools (eg if appropriate, prescribing advice)
Endorsements
See note for full guideline. The logos of endorsing or approving organisations are especially important and should be clearly noticed at a glance.

Team Members and Process

Brief outline of the Grading System used

Information on where to get access to the full guideline