TABLE OF CONTENTS

STAFF DETAILS.................................................................................................................. 5

OUTLINE AND LEARNING OBJECTIVES........................................................................... 6

1. INTRODUCTION ............................................................................................................. 8
   1.1 References.................................................................................................................. 11
   1.2 Further reading.......................................................................................................... 11

2. ASKING QUESTIONS ..................................................................................................... 12
   2.1 Learning objectives ................................................................................................. 12
   2.2 Core reading and key concepts ............................................................................... 12
   2.3 Worked examples .................................................................................................... 15
   2.4 Self-complete exercise ........................................................................................... 17
   2.5 Answers to self-complete exercise .......................................................................... 19
   2.6 Recommended Reading .......................................................................................... 19
   2.7 Further reading ....................................................................................................... 20
   2.8 References ............................................................................................................... 20
   2.9 Websites .................................................................................................................. 20

3. SEARCHING THE MEDICAL LITERATURE ................................................................. 21
   3.1 Learning objectives ................................................................................................. 21
   3.2 Core reading and key concepts ............................................................................... 21
   3.3 Worked example ..................................................................................................... 34
   3.4 Self-complete exercises .......................................................................................... 39
   3.5 Answers to self-complete exercise .......................................................................... 40
   3.6 Recommended reading ........................................................................................... 40
   3.7 Further reading ....................................................................................................... 41
   3.8 Websites .................................................................................................................. 41

4. ARTICLES ABOUT THERAPY .................................................................................... 43
   4.1 Learning objectives ................................................................................................. 43
   4.2 Core reading and key concepts ............................................................................... 43
   4.3 Worked example ..................................................................................................... 50
   4.4 Self-complete exercise ........................................................................................... 53
   4.5 Answers to self-complete exercises ........................................................................ 55
   4.6 Recommended reading ........................................................................................... 57
   4.7 References ............................................................................................................... 57
   4.8 Critical appraisal worksheets .................................................................................. 59

5. SYSTEMATIC REVIEWS .............................................................................................. 63
   5.1 Learning objectives ................................................................................................. 63
   5.2 Core reading and key concepts ............................................................................... 63
   5.3 Worked examples .................................................................................................... 72
   5.4 Self-complete exercises .......................................................................................... 79
   5.5 Answers to self-complete exercises ........................................................................ 84
   5.6 Recommended reading ........................................................................................... 87
   5.7 Further reading ....................................................................................................... 88
   5.8 References ............................................................................................................... 88
   5.9 Critical appraisal worksheets (systematic reviews) ............................................... 91

6. DIAGNOSTIC TESTS .................................................................................................... 96
STAFF DETAILS

Academic Coordinator
Dr Peter Greenberg
Physician, Department of General Medicine
Principal Fellow, Department of Medicine and Associate Professor, School of Population Health,
Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne
Project Director, Evidence Based Practice
Clinical Epidemiology & Health Care Evaluation Unit
The Royal Melbourne Hospital
Tel  9342 7459
Email:  Peter.Greenberg@mh.org.au

Administration
Ms Meg McKay
Academic Programs Officer (Health Practice)
School of Population Health, University of Melbourne
Room 427, Level 4, 207 Bouverie Street
Carlton  VIC  3053
Tel:  8344 9336
Email:  mckaym@unimelb.edu.au
OUTLINE AND LEARNING OBJECTIVES

Evidence-based Practice (EBP) has been described as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”¹. EBP integrates clinical expertise and patient values with the best available research evidence².

There are four basic steps to incorporating the best available research evidence in decision making: asking questions; accessing the best information; appraising that information for validity and relevance; and applying the information to patient care.

This EBP course aims to equip you with the skills needed to both identify and incorporate the best available research evidence into your pre-clinical and clinical training and subsequent practice. It is presented in Health Practice 2 and 3, and revised and practised during the clinical years.

In HP 2, there are 3 lectures on ‘Asking questions’, ‘Searching secondary databases’ and ‘Searching primary databases’, the latter 2 of which are followed by practical tutorials. In HP 3 there are 3 lectures on the critical appraisal of publications addressing ‘Risk and prognosis’, ‘Treatments and diagnostic tests’, and ‘Systematic reviews of treatments’, each of which is followed by a practical tutorial. This workbook complements the formal lectures and tutorials, and contains additional material for use during the later years of the medical course and afterwards.

Learning Objectives:

Knowledge to:
- frame specific clinical and research questions using the Patient group/Population, Intervention, Comparator, Outcome, Time (PICOT) framework.
- identify the types and strengths of evidence which address questions of cause, risk/prognosis, diagnosis, and treatment.
- understand concepts of ‘chance’, including p values and confidence intervals
- understand the principles of critical appraisal of research publications and of applying data from relevant, appraised publications to individual patients and to specific groups.
- understand the principles of how to integrate the results of questions, searches and appraisals with clinical experience

Skills to:
- formulate questions arising from clinical cases so that they can be answered using evidence
- conduct ‘secondary’ and ‘primary’ (MEDLINE) database searches, and to retrieve relevant publications which address the questions above
- critically appraise studies of disease risk factors, prognosis, diagnostic tests, treatments and systematic reviews of treatments
- consider the possibility that the results of studies might reflect chance, bias or confounding
- communicate the results of searches and appraisals, so that individuals and groups can make optimal decisions for themselves
Attitudes to:

- appreciate the benefits to patients when medical students and clinicians understand the principles of EBP.

EBP offers efficient tools to improve patient care and outcomes, and a framework for continual learning and professional development.

If you have questions or comments, please contact your lecturer, tutors, or staff listed under Staff Details in this workbook.

Peter Greenberg

Recommended for further reading and reference


1. INTRODUCTION

“Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?”

From “The Rock”. TS Eliot 1934

Decisions made during consultations between patients and clinicians have many determinants. They depend on the patient, the clinician involved at the time and the setting in which the interaction takes place. Practical constraints necessarily limit and shape the decisions made at particular places and times. Perceptions about the type and the quality of the evidence supporting the clinical decisions are also relevant.

Patient and clinician related factors include cultural beliefs, personal values, knowledge and experience. Constraints include policies, laws, community standards and costs. The relative importance of scientific evidence has grown over the last 150 years, and especially in the last 50 years. This partly reflects increasing expectations of patients (‘consumers’) for clinicians to know and use even more evidence in healthcare decisions. We should not be afraid of such changes in community expectations, but clinicians need tools to be better equipped to incorporate evidence into practice.

‘Evidence’ means different things to different people. To a fundamentalist it might mean what you know to be true. A legal definition of evidence is ‘material which tends to prove or disprove the existence of a fact, and which is admissible in court’. To a ‘post-modernist’, ‘evidence’ could mean something that you have experienced yourself. The ancient Greeks debated the value of rational evidence, which could be soundly argued according to logical rules, compared to empirical evidence, which was assumed to be valid because it was based on observations. Arguments about the relative values of different types of evidence remain unresolved, but the scientific basis of medicine is founded on empirical evidence, which comes from observations and experiments. Scientific evidence in clinical decisions incorporates basic, clinical, and epidemiological data obtained from observation and experiment.

This workbook is designed to help evolving clinicians adapt to a new way of providing evidence-based care. It is not always easy and may even be threatening at times. The patient before you in the clinic or hospital expects you to do your best for them. If your focus is on ensuring that the best quality evidence is available, if you listen to patients and take their views into consideration in formulating management plans, you are well on the path to “finding the knowledge lost in the information”.

What is evidence-based practice?

There is nothing new about applying empirical evidence to clinical situations. The first controlled trial is described in the Old Testament. The fate of Daniel and his colleagues, who were fed a diet of vegetables and water for 10 days, was compared to a control group of Nebuchadnezzar’s subjects, who consumed meat and wine. Maimonides, physician to Saladin the Great of Egypt and a 12th century philosopher, also supported a scientific approach: “He who puts his life in the hands of an empiricist who does not think scientifically, is like a mariner who places his trust in good luck.”
Scientific methodology in medical practice was firmly established by Pierre CA Louis (1787-1872), who established that blood-letting did not result in better outcomes for patients with pneumonia.

Although scientific evidence has been increasingly incorporated into clinical medicine, especially since the start of the nineteenth century, the term ‘evidence based medicine’ (EBM) was first used in Canada in 1990\textsuperscript{2}. It appeared in refereed publications soon after this\textsuperscript{3}.

Evidence based practice has been described as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”\textsuperscript{4}. It “requires the integration of best research evidence with clinical expertise and our patient’s unique values and circumstances”\textsuperscript{5}.

Straus and colleagues elaborate further on these elements\textsuperscript{5}:

**Clinical expertise**

“By clinical expertise we mean the ability to use our clinical skills and past experience to rapidly identify each patient’s unique health state and diagnosis, their individual risks and benefits of potential interventions, and their personal circumstances and expectations.”

**Best available research evidence**

“By best research evidence we mean valid and clinically relevant research, often from the basic sciences of medicine, but especially from patient-centred clinical research into the accuracy of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative and preventive regimens. New evidence from clinical research both invalidates previously accepted diagnostic tests and treatments and replaces them with new ones, that are more accurate, more efficacious, and safer.”

**Patient values**

“By patient values we mean the unique preferences, concerns and expectations each patient brings to a clinical encounter and which must be integrated into clinical decisions if they are to serve the patient.”

**Patient circumstances**

“By patient circumstances we mean their individual clinical state and the clinical setting.”

**The ‘steps’ of evidence-based practice (EBP)**

EBP requires the following steps\textsuperscript{5}:

- converting the information needs which arise from clinical encounters into questions.
- searching for the best evidence available, in order to answer these questions.
- critically appraising the evidence found. This involves examining the internal validity of the study, assessing the importance, or the size of the effect, and deciding on the relevance of the study to specific clinical situations.
- applying the information from the critical appraisal to the care of individual patients. This requires integrating the evidence with clinical expertise and with knowledge about individual patients, their values and circumstances. It also includes communicating the outcomes of the above steps to patients, to assist them to make optimal decisions for themselves.
- evaluating the four steps above, and learning ways to make them even more effective and efficient.
What can we expect in future?

Clinical practice is likely to continue to become even more enriched by an expanding body of research information. The quality of the evidence that is available to practising clinicians is likely to improve both as better research methods and the science of synthesis of research findings are developed. Clinicians are increasingly confronted with inadequate time and resources to find, evaluate and incorporate new knowledge into everyday clinical decision-making. Information technology is already providing ‘point-of-care’ access to data on effectiveness of care.

Most doctors want to be ‘better and faster’

Timely, useful evidence from biomedical literature should be an integral component of clinical decision-making. For example, if a treatment is shown to be superior, we need to know about it soon. It is easy for the latest research evidence to pass us by.

Although at some times we may lack the time, motivation, or all of the skills needed to efficiently and properly appraise and synthesise information, we need to understand the principles of how to do all of these. In contrast, we all need to know how to seek and find relevant information and how to involve patients in the clinical decisions that affect them. We also need to know the ‘short-cuts’ that are available for the other steps of EBP, so that we can incorporate evidence at those times when we do not wish or are unable to undertake all of the steps on our own.

Clinicians now face new challenges to ensure that all the knowledge that is available is used and that patients are fully involved, to the extent they wish to be, in making the best quality decisions. There are tools, which help us achieve these goals. If you develop and use the skills of EBP, you start the lifelong path to becoming and remaining a ‘better, faster doctor’. Enjoy yourself as you acquire and practise the skills needed to find “the knowledge lost amongst the information”!

Training as a clinician teaches you to listen to a patient’s story, to integrate the history with the findings on physical examination, to incorporate the results of investigations, and along with the patient, to define and prioritise the patients’ health ‘issues’. EBP helps you to access clinical ‘information’ and to appraise it, so that it becomes ‘knowledge’. Clinical experience, discussion with colleagues and others, and listening and responding to patients even has the capacity to transform ‘knowledge’ into ‘wisdom’.

“The most common criticism made at present by older practitioners is that young graduates have been taught a great deal about the mechanisms of disease, but very little about the practice of medicine - or, to put it bluntly, they are too “scientific” and do not know how to take care of patients.”

In 2006, the British Medical Journal sought views of its readers on the most important medical milestones. A panel of editors and advisers narrowed the field down from more than 70 to 15. One of these was ‘evidence based medicine’.
1.1 References


6. Peabody FW. The Care of the Patient. JAMA 1927; 88:877-882


1.2 Further reading


2. ASKING QUESTIONS

2.1 Learning objectives

- To understand the difference between ‘background’ and ‘foreground’ questions
- To understand the components of structured clinical questions.
- To be able to formulate the information needs arising from ‘issues’ generated during clinical encounters into potentially answerable questions.

2.2 Core reading and key concepts

Introduction

Almost every time we see a patient we need up to date and reliable information to answer questions arising about some aspects of cause, diagnosis, prognosis or treatment. Such needs depend on the extent of our ‘background’ knowledge and experience at the time. Additional information, often in the form of external evidence, which we might have to track down and then evaluate, may be required as well. For many clinicians, the effort required to both ask questions and seek answers is so formidable that, when coupled with limited time for reading, discussion and attending meetings and courses, some information needs are never fully realised.

As trainee clinicians progress through learning programs their information requirements change. There is a shift from fairly broad ‘background’ questions, which address general knowledge about conditions and other subjects, to be replaced by more patient and circumstance-specific ‘foreground’ questions.

The two components of ‘background’ questions are usually a question root (eg Who…?, What…?, Where…?, How…?, When…?, Why…?) and a condition (eg disorder, treatment or test). Some examples are: “What might cause this condition?”; “How do you diagnose…?”; “What is the prognosis for patients with…?”; “What are the principles of treatment for patients with…?”.

A structure for asking potentially answerable clinical ‘foreground’ questions is necessary. Such questions initially seem less intuitive than ‘background’ questions. This section is designed to give you some understanding of what is required to formulate potentially answerable ‘foreground’ clinical questions, the first step in evidence-based practice (EBP). In our experience, this is the hardest step faced in finding the best current evidence to address clinical problems. The good news is that the skills can be learned and that they do improve with practice. Clearly, not all potentially answerable clinical questions have answers, but some of those that do not have answers can be developed further into research questions to be answered later on.
Structure of a ‘good’ clinical question: ‘PICO’

1. **Patient or problem**
   How would you describe a group of patients similar to yours? What are the most important characteristics of the patient? This could include the primary problem, the disease, or the co-existing conditions. Sometimes the sex, age or ethnic background of patients might also be relevant to the diagnosis or treatment of a disease.

2. **Intervention, prognostic factor, or exposure**

3. **Comparison**
   What is the main alternative to compare with the intervention? Are you trying to decide between two drugs, a drug and no medication or placebo, or two diagnostic tests? Note that some clinical questions do not always require a specific comparison to be made.

4. **Outcomes**
   What do you hope to accomplish, measure, improve or affect? What are you trying to do for the patient? Prolong life? Relieve or eliminate symptoms? Reduce the number of adverse events? Improve functional performance or improve the results of tests?

The structure of a ‘foreground’ question related to therapy might look like this:

*In those patients with (Patient population) does this (Intervention) or (Comparison, if any) affect this (Outcome) over the next…..(Time)?*

The type of question and the type of study are helpful in focusing the question and determining the most appropriate type of clinical evidence.

**Type of question**

The most common types of questions related to clinical tasks:

- **Diagnosis**
  How to select and interpret diagnostic tests

- **Therapy**
  How to select treatments which offer patients more good than harm, and that are worth the efforts and costs of using them

- **Prognosis/Risk**
  How to estimate the patient's likely clinical course over time and how to anticipate the complications of diseases or the chance of developing them

- **Harm/Aetiology**
  How to identify causes for disease (including iatrogenic cause)
Type of study
There are many different types of studies. Some of these are shown below, in the form of a pyramid:

- Randomized Controlled Trials
- Cohort Studies
- Case Control Studies
- Case Series/Case Reports
- Opinions, letters
- Animal research
- In vitro research

The pyramid illustrates the evolution of the research literature. The base of the pyramid is where information usually starts, from an idea based on patho-physiological considerations, or from laboratory research. As these ideas are developed into possible ‘interventions’, such as drugs and diagnostic tools, they are tested in laboratory models, then in animals, and finally in humans. Testing in humans might begin with volunteers and go through several phases of clinical trials, before the drug or diagnostic tool is adopted for use within the general population. Randomised controlled trials are undertaken to further test the effectiveness and efficacy of a drug or therapy. *As you move up the pyramid the volume of published literature may decrease, but its relevance to clinical settings increases.*

The ‘best’ evidence for questions of therapy comes from randomized, controlled clinical trials and the systematic reviews derived from them. Randomised controlled trials are carefully planned projects that study the effect of a therapy or test on real patients. They include methodologies that reduce the potential for bias and that allow for comparison between intervention groups and control groups (no intervention). Evidence for questions of diagnosis is found in prospective studies, which compare tests with a reference or ‘gold standard’ test.

Evidence of aetiology or harm sometimes comes from randomised controlled trials but is usually found in cohort studies, which take a large population and follow patients with a specific condition or receive a particular treatment over time. These are compared to another group that has not been affected by the condition being studied. Case control studies are studies in which patients who already have a specific condition are compared with people who do not. In general, observational studies are less reliable than randomised controlled trials because the results are more likely to be biased. Furthermore, showing an association between two factors does not necessarily mean a causal relationship.
Case series and case reports consist of collections of reports on the treatment of individual patients or a report on a single patient. Because they are reports of cases and do not include control groups with which to compare outcomes, they have no statistical validity, but may serve to generate hypotheses, which can then be assessed scientifically.

The evidence pyramid also serves to introduce the concept of a hierarchy or ‘level’ of evidence, which is based on the type of study design. ‘Levels of evidence’ are best developed for studies related to questions addressing therapy (see Chapter 4.2). You may not always find the highest-level evidence to answer your clinical questions. In the absence of high-level evidence, you may need to consider moving down the pyramid to other types of studies.

Data derived from the evidence pyramid can be grouped in various ways to facilitate the dissemination and update of evidence into clinical practice:

Systematic reviews usually focus on a clinical topic and answer a specific question. Extensive literature searches are conducted to identify all relevant studies of sound methodology. The studies are reviewed, assessed and summarised according to predetermined criteria, to address the question posed in the review. Meta-analyses take systematic reviews a step further by using techniques to summarise the results of several studies into quantitative estimates of the combined result.

Clinical practice guidelines are systematically developed statements to assist practitioners and patients to make decisions about appropriate health care in specific circumstances. Guidelines review and evaluate the evidence and then make explicit recommendations for clinical practice in specified settings.

2.3 Worked examples
Start with the patient: a clinical problem/question arises from issues of care.

Joyce is an active 80-year old female. She lives in her own home and still drives a car. Several weeks ago she developed pain in her knee, making it hard for her to walk.

The doctor made the diagnosis of osteoarthritis, and gave her a short course of an oral non-steroidal anti-inflammatory drug (NSAID). She asks Joyce to see a physiotherapist to begin a course of physiotherapy. Joyce is wary of potential side effects of the tablets, but has had physiotherapy before for her back and is very sceptical about this suggestion. She wants to know if there is any proof that physiotherapy helps this problem.

You wonder whether there is any evidence to support the proposition that physiotherapy does any good for this problem also. Does it?
What information-seeking strategies can you think of to answer this question?

The Question
Consider your current strategies for keeping up to date with the medical literature?

Write them down:
1. ................................................................................................... ...................................
2. ................................................................................................... ...................................
3. ................................................................................................... ...................................

Some answers that might come to mind
1. Ask a keen (? young) physiotherapist
2. Ask a rheumatology consultant or registrar
3. Thumb through the back issues of the Medical Journal of Australia (MJA) in your medical library, hoping that this topic was recently included.
4. Check the rheumatology section of the ‘Oxford Textbook of Medicine’
5. Glance at your lecture notes

Are there easier ways to get information from the published literature that might appeal to someone who wants to be a ‘better faster doctor (or medical student)’?

Can you phrase this problem as an answerable clinical question?
This is the first and most important step on the road to using search strategies for a literature search (eg. using MEDLINE) and for critical appraisal.

The Question
Construct the clinical question.
HINT: Phrase your question in four parts;

Population     What is the group of patients that you will apply the information to? (young or old, gender)
Intervention   What is the intervention? (treatment, diagnostic test)
Comparison     What do you want to compare this intervention to? (nothing?, placebo?, other?)
Outcome.       What is the result you are looking for? (less pain, fewer falls, fewer knee replacements, improvement in knee X Ray, side-effects of medication, diagnostic test accuracy)

Choose the option which you think is the best question:

A. Is physiotherapy good for a painful knee?
B. Can physiotherapy help an elderly 80-year old female?
C. In people with knee osteoarthritis does physiotherapy improve pain and mobility?
A. *Is physiotherapy good for a painful knee?*
   This question is not specific enough. We still want to know the particular condition and the outcome expected.

B. *Can physiotherapy help an elderly 80-year old female?*
   We still want to know what the problem is and what the treatment can do for the patient (the outcome).

C. *In people with knee osteoarthritis does physiotherapy improve pain and mobility?*
   We now know the patient problem, the intervention, and the desired outcome.

| Note | Although we may not use all of this information to construct a search, we may need to use more of it when it comes to evaluating publications retrieved. |

The elements of the ‘answerable’ clinical question are then used to build a search strategy, in order to find the most relevant and highest quality evidence to inform a decision.

### 2.4 Self-complete exercise

**Clinical Problem**

A 65-year old retired teacher has osteoarthritis in both knees
- she still has pain in spite of paracetamol 4 gm daily
- you wonder about a trial of a non-steroidal anti-inflammatory drugs (NSAID) instead

You are aware of an article², which says that paracetamol and one of the NSAIDs are equally effective in relieving pain. You wonder if one NSAID has a better side-effect profile than another. In particular, you want to know whether there is any information about the relative gastro-intestinal risks of different NSAIDs. This might help you decide which NSAID to try, without putting your patient at undue risk.

**Can you phrase this as an answerable clinical question?**

This is the first and most important step on the path leading to a literature search and critical appraisal.

**HINT:** Phrase your question in three or four parts:

- **Population** What is the group of patients that you will apply the information to?
- **Intervention** What is the intervention?
- **Comparison** What are you comparing (i.e. if you want to make a comparison)?
- **Outcome** What is the result you are seeking?
**CLINICAL CASE/SCENARIO**
What are the important features of the case?

**PATIENT OR PROBLEM**
How would you describe a group of similar patients? What are the most important characteristics of the patient?

**INTERVENTION, EXPOSURE, PROGNOSTIC FACTOR**
What main intervention are you considering? What do you want to do with this patient?

**COMPARISON**
What is the main alternative being considered (if any)?

**OUTCOME**
What are you trying to accomplish, measure, improve or affect?

**QUESTION**
The clinical question is:
2.5 Answers to self-complete exercise

CLINICAL CASE/SCENARIO:
What are the important features of the case?

65-year old retired teacher with osteoarthritis in both knees who currently has inadequate pain relief on Paracetamol 4 gm daily, requests a trial of NSAIDs. Does one NSAID have a better gastro-intestinal side effect profile than another?

PATIENT OR PROBLEM
How would you describe a group of similar patients? What are the most important characteristics of the patient?

has osteoarthritis

INTERVENTION, EXPOSURE, PROGNOSTIC FACTOR
What main intervention are you considering? What do you want to do with this patient?

prescribe a non-steroidal anti-inflammatory drugs (NSAID)

COMPARISON
What is the main alternative being considered?

another non-steroidal anti-inflammatory drug

OUTCOME
What are you trying to accomplish, measure, improve or affect?

gastro-intestinal side-effects

The clinical question is:

Does one NSAID have fewer (gastro-intestinal) side-effects than another?

2.6 Recommended Reading

2.7 **Further reading**


2.8 **References**


2.9 **Websites**

- ‘Focusing clinical questions’ from the Centre for Evidence-Based Medicine Oxford(UK) [http://www.cebm.net/focus_quest.asp](http://www.cebm.net/focus_quest.asp)

- An interactive tool from the Centre for Evidence Based Medicine Toronto [http://www.cebm.utoronto.ca/practise/formulate](http://www.cebm.utoronto.ca/practise/formulate)

- A ‘downloadable’ tool for ‘hand-holds’ for storing clinical questions from the Centre for Evidence Based Medicine Toronto [http://www.cebm.utoronto.ca/palm/cqlogbook](http://www.cebm.utoronto.ca/palm/cqlogbook)
3. SEARCHING THE MEDICAL LITERATURE

3.1 Learning objectives

To understand:

• how to use the components of a (‘foreground’) structured question to devise a ‘search strategy’
• the principles of searching ‘primary’ databases like MEDLINE, and:
  o the difference between a ‘text word search and a ‘thesaurus’ search, using MeSH (Medical Subject Heading) terms
  o the main search operations
  o the use of ‘filters’
  o the use of ‘limits’
• the principles of searching ‘secondary’ databases
• the principles of searching with ‘EndNote’

3.2 Core reading and key concepts

When overwhelmed by information we sometimes find it difficult to locate specific material which adds value to clinical decision-making. In order to keep up with the medical literature, some clinicians have a random approach, some read journals systematically, and some rely on serendipity. Others set time aside to read at least the tables of contents of the latest journals and then to peruse particularly interesting articles.

As you know, one solution for the problem of keeping up to date with the ever-changing medical literature is called ‘problem-based learning’. When confronted by a clinical question and unsure of the current best answer, you need to know where to look and how to find answers. You need sensitive (ie retrieve relevant citations), specific (ie don’t retrieve irrelevant citations) and efficient (ie save time) search strategies to help you find answers to your questions, to “find the knowledge lost amongst the information”.

3.2.1 Where do you find the best evidence?

Textbooks, especially electronic versions which are frequently updated, remain a key source of background information to help you learn the basics of diseases and therapeutics.

Repositories of clinically relevant, primary sources of research evidence are held in huge databases, which are accessed electronically. MEDLINE is the best-known example, but there are many others, such as CINAHL and EMBASE. The Cochrane library has several components, including a primary database, the Cochrane Central Register of Controlled Trials, now called “Clinical Trials”.

Increasingly however, knowledge of common clinical problems is summarised and stored in smaller, more selective electronic repositories. These secondary sources of evidence, which may include material that has been systematically reviewed and/or pre-appraised, are based on the ‘primary’ research publications. They are increasingly finding favour as trusted and easily accessible sources of valuable and up to date information, particularly for common conditions. Busy clinicians are increasingly accessing these electronically, at the point of patient care at the bedside or in the clinic and office, in their quest to become “better, faster doctors”.
Clinicians need to know how to use computers to access information from these sources and to recognise when they may need to seek assistance, especially from librarians in academic and clinical settings.

This chapter provides background information about MEDLINE. You will learn data acquisition skills through a clinical issue, which generates a structured question as the basis for developing a MEDLINE search strategy to find clinically relevant information. You will also be introduced to the ‘secondary’ databases. A list of resources to help you find evidence-based information is provided at the end of this chapter.

Subsequent chapters in this Workbook address critical appraisal of the information retrieved, and the other steps of evidence-based practice.

### Note

Depending on the type of clinical question being asked, it is often more efficient to search first the secondary databases such as ‘Cochrane Reviews’, ‘Clinical Evidence’ or the ‘ACP Journal Club’. If more information is required, you can then search primary databases like MEDLINE. In this chapter, however, searching with MEDLINE is discussed first.

#### 3.2.2 ‘Primary’ sources of evidence: Developing a search strategy (eg MEDLINE)

**The key stages in search preparation are:**
- asking a specific search question
- breaking the question into parts
- choosing the best search terms
- using the chosen terms for searches

**Developing a search strategy is the process of:**
- formulating the question
- choosing the appropriate database
- selecting the best Medical Subject Headings (MeSH) to describe your terms
- combining the terms or sets
- limiting citations retrieved to those deemed most appropriate

**Reliable and efficient searching requires an understanding of:**
- medical subject headings or ‘MeSH’ terms
- ‘explode’ and ‘focus’ functions for MeSH searches
- ‘subheadings’ for MeSH terms
- ‘text words’ including ‘truncation’
- ‘boolean’ logic (eg AND, OR, and NOT)
- how to limit searches
The strategy for searching a ‘primary’ database like MEDLINE is outlined below:

3.2.3 The MEDLINE database

MEDLINE is produced by the National Library of Medicine (NLM) in the United States [http://www.nlm.nih.gov/nlmhome.html](http://www.nlm.nih.gov/nlmhome.html). Although broad in its coverage and the most well known and widely used database within the medical field, it is important to remember that MEDLINE has a significant American bias. It contains about 50% American material.

3.2.4 Access to the MEDLINE database


Students at the University of Melbourne can, after logging in, also access ‘PubMed’ through ‘Supersearch’. This route provides more links to full text publications than through the standard route to ‘PubMed’, described above.

2) Access to MEDLINE can also be purchased, usually with additional electronic resources as well, including additional links to free full text articles, through commercial providers like ‘Institute for Scientific Information (ISI) Web of Science’ and ‘OVID’. The University of
Melbourne subscribes to ‘ISI Web of Science’ which, after logging in, students and staff can access through ‘Supersearch’. The Victorian Department of Human Services, through the ‘Clinicians Health Channel’, and many of the teaching hospital libraries subscribe to ‘OVID’, which students and staff can access while working within Victorian public hospitals.

In the section of this chapter, the principles of searching MEDLINE through ‘PubMed’ are outlined. These are also some brief comments on searching through ‘ISI Web of Science’ and ‘OVID’.

Most of us find it easier to learn how to use one of these 3 portals to MEDLINE in the first instance, possibly ‘PubMed’, given its universal availability. Once the principles are mastered it is easy to apply them to other portals to MEDLINE. On the other hand, University students might prefer to learn to use ‘ISI Web of Science’ in the first instance, which is easily accessed.

In practice, some clinicians use ‘ISI Web Science’ from the University, ‘OVID’ from the ‘Clinicians Health Channel’ or teaching hospitals and ‘PubMed’ when these resources are unavailable. Other clinicians use ‘PubMed’ exclusively for searching, and only use the other portals as a source of additional full text publications, once these have been located.

Each of these 3 portals to MEDLINE provides excellent ‘online’ tutorials on their homepages, with details of how to search using them.

To give a few facts and figures:

- MEDLINE contains over 15 million citations to articles published, mainly since 1966
- 500,000 citations are added to MEDLINE each year
- MEDLINE is not a full text database, so searches are conducted within the title and abstract, and key words, but not the full text
- MEDLINE covers over 4,800 different journal titles, originating from over 70 different countries
- more than 70% of citations contain abstracts

Each record in the database is broken down into separate sections or fields (eg author, title, abstract). To aid retrieval of information the database is fully indexed using over 23,000 standard or thesaurus terms. These are called Medical Subject Headings or MeSH terms for short.

3.2.5 Medical Subject Headings (MeSH terms)

Searching with MeSH terms helps to prevent important material being missed (ie it increases the sensitivity of a search) by avoiding the need to search for additional words or phrases as well, in the text of titles and abstracts of records making up the database. For example, the problems associated with singular and plural forms of terms, differences between British and American spellings, and the use of different synonyms by different authors. Searching with MeSH terms can also increase the search specificity (ie reduce the proportion of irrelevant citations), and hence improve the overall efficiency of searches.

The MEDLINE database is indexed by information content specialists. They read each article carefully and then assign terms to describe its content, as specifically as possible. The terms used come from a standardised list of vocabulary and definitions called MeSH terms or Medical Subject Headings. Each citation is allocated about 10-20 MeSH terms.

The principle underlying MeSH terms is that all articles about the same concept are assigned the same standardised subject heading. For example, one author may use the term diabetes type 1, another may use insulin-dependent diabetes, and a third might use diabetes mellitus, sudden-
onset. Citations for all of these articles are indexed under *Diabetes Mellitus, Insulin-Dependent*, the MeSH term for this disease. Searching using the text phrase (ie phrase within quotation marks) “insulin dependent diabetes” in ‘PubMed’ retrieves only about one quarter of the citations retrieved when searching with the MeSH term *Diabetes Mellitus, Insulin-Dependent* or when searching the terms *insulin AND dependent AND diabetes*, as the latter search enables mapping to MeSH terms, which phrase searching does not (see 3.2.6).

**How to find MeSH terms and explore MeSH trees**

MeSH terms and trees can be displayed in MEDLINE and in the Cochrane Library databases. You can check the appropriate MeSH term and also see its relationship to broader (ie a more ‘proximal’ branch of the MeSH tree) and more specific (ie more ‘distal’ branch of the MeSH tree) MeSH terms.

‘PubMed’ has a ‘MeSH Database’, which is accessed from the left panel on the ‘PubMed’ home page. In ‘ISI Web of Science’ choose ‘General Search’ then ‘MeSH HEADING’ then ‘thesaurus’, and insert terms in the MeSH Thesaurus box. In ‘OVID’, choose ‘Search Tools’ and enter a term using ‘Permutted Index’ for MeSH terms and ‘Tree’ for MeSH trees. MeSH trees are displayed in the Cochrane databases by choosing ‘MeSH Search’ from the right side panel on the home page.

Three MeSH trees, which include the MeSH ‘Diabetes Mellitus, Type 1’ from the ‘PubMed MeSH database’, are pasted below:
Why bother learning about MeSH terms and exploring MeSH trees?
One reason is that broad MeSH terms ‘branch’ into successively narrower terms. MEDLINE indexers are requested to index publications to the most ‘specific’ level possible. This means that you will often find fewer and only more ‘general’ citations indexed when using broader (ie ‘proximal’) MeSH terms for searching, unless you take steps to include the more ‘distal’ branches of the MeSH trees, which include the more specific terms as well. In the example above, for example, a MeSH search only using the MeSH term ‘Diabetes Mellitus’ would not include citations allocated the MeSH term ‘Diabetes Mellitus, Type 1’, which might be your main interest! The good news is that you can take steps to include more distal MeSH terms as well (see ‘Explode’ below).

‘Mapping’ of text words to MeSH terms
Another reason for understanding MeSH terms is to understand how a particular search will be or was processed. With ‘PubMed’, ‘ISI Web of Science’ and ‘OVID’, the default is direct ‘mapping’ of text words to the relevant MeSH terms. Mapping is not always possible, as MeSH terms are only allocated to some of the contents of publications. MeSH words may not yet be developed for new and evolving topics. You can also search with MEDLINE using the appropriate MeSH term with or without text words if you wish, or search with text words only (see 3.2.6).

How to confirm the process of a MEDLINE search
It is easy, and often very useful to see exactly how a search was processed, for example to confirm if MeSH terms were indeed used in that particular search. You can see how a search was processed in ‘PubMed’ by choosing ‘Details’, from the home page, at the conclusion of a search a search. You can also delete elements from the search and re-search with more specific searches or different terms if you wish. In ‘ISI Web of Science’ the search details are displayed in ‘Search Results—Summary’, at the top of the page, after searching. The ‘Search History’ column in ‘OVID’ shows how a search was processed, and if and which MeSH terms were used.

‘Major’ (‘focus’) of publications
Indexers at the USA National Library of Medicine also identify and specify those citations of publications where the MeSH terms used indicate that the term applied is a ‘major’ focus of the article being indexed. In this case the term is referred to as being ‘majored’. MeSH terms are asterisked when they are ‘majored’. In MEDLINE you can select ‘majored’ publications when searching, to make searches more specific.

How to retrieve citations where the MeSH term is a major focus of the publication
With MeSH searches in ‘Pub Med’, check the box alongside: ‘Restrict Search to Major Topic headings only’, in ‘ISI Web of Science’ MEDLINE check the box beside ‘Major Topic headings only’ and in the ‘OVID’ ‘Mapping Display’ screen, check the box beside ‘Focus’.

| Note | As ‘majoring’ is a judgement made by the indexer, who is required to identify the 3 or 4 major topics of the article, it is often best to begin with a broader search strategy by not restricting to majored citations for that MeSH term. If you get too many citations, you can always go back and re-enter the subject heading and choose only the ‘majored’ citations. |

Explode
This allows you to expand your search by including additional and more specific citations from MeSH terms from the distal branches of the MeSH trees. This function is known as ‘exploding’ the broader MeSH term. One problem of ‘exploding’ is that if the topic of your search is very specific, ‘exploding’ a broad (specific) MeSH term may also retrieve a great deal of irrelevant citations as well. Therefore, it is sometimes best to increase the search specificity at the cost of search
sensitivity, by searching with a specific MeSH term on its own, without ‘exploding’; the risk is losing relevant citations indexed more specifically (see ‘Why bother about MeSH terms and exploring MeSH trees?’ above.

How to retrieve citations which include all the ‘distal’ branches of the MeSH tree

With ‘MeSH’ searches in ‘Pub Med’ (see the ‘MeSH Database’ screen), ‘Explode’ is the default setting, but you can choose ‘Do Not Explode This Term’ for a more specific search, which includes only the chosen MeSH heading. Check the box alongside ‘Do Not Explode this term’, in ‘ISI Web of Science’ MEDLINE, and in the ‘OVID’ ‘Mapping Display’ screen, check the box beside ‘Explode’.

‘Majoring’ (or ‘Focusing’) and ‘Exploding’

Note

‘Majoring’ (or ‘Focusing’) and ‘Exploding’ are completely different concepts. For example it is possible and often, but not always, useful for the reasons outlined above, to both ‘Focus’ and ‘Explode’ a MeSH term before searching.

Subheadings

For more specific searching, MeSH terms may be combined with particular subheadings, but you must select just one MeSH term at a time in this case. There are over 80 sub-headings, which describe particular aspects of, or particular ways of looking at, the MeSH term concerned. For example: ‘adverse effects’; ‘diagnosis’; ‘therapy’. Different subheadings apply to different MeSH terms: where a sub-heading cannot be applied to a particular MeSH term it is because it would not make sense to do so. Subheadings can help by narrowing a search to those more specific aspects of the MeSH term, which you consider to be more relevant at the time.

How to search using the MeSH term subheadings for particular MeSH terms

You can select more than one subheading by checking in the boxes next to the terms. In ‘PubMed’ all subheading are selected by checking the MeSH term itself. When searching with MeSH terms, after choosing the subheadings of relevance, choose an option from the pull-down menu at the top and bottom of the MeSH page, called ‘Send to’, eg ‘Search Box with AND’.

In ‘ISI Web of Science’ MEDLINE choose ‘General Search’, then ‘thesaurus’ then ‘View qualifier hierarchy’ from the ‘MeSH Theraurus’ screen, and select subheadings of interest. In ‘OVID’ MEDLINE you are asked to include all or particular subheadings for MeSH terms in the course of the search process.

The 38 Subheadings for the MeSH term ‘Diabetes Mellitus, Type 1’ from the ‘PubMed MeSH database’, are pasted on the next page. The subheading ‘therapy’ has been selected.
Note also;

- in the left-hand side of the screen above, how you would locate the ‘PubMed’ MeSH database from the ‘Entrez PubMed’ home page, as discussed in ‘How to find MeSH terms and explore MeSH trees’ above
- in the middle of the screen above, the box to check to ‘Restrict Search to Major Topic headings only’, and the box to ‘Do Not Explode’ this term, as discussed above

### 3.2.6 Searches using text words

In contrast to MeSH searches, in ‘text word’ searches you search for the **exact word(s) or phrase(s) used in the title and/or abstract or key words of the publication.** This does NOT necessarily guarantee that the word or phrase is the main topic of the publication.

Text words are useful in finding information on:

- topics for which there is no allocated MeSH heading, such as one that might be very new in the published literature
- subjects that have very specific names, such as the ‘GUSTO’ trials
- When searching with text words, you need to be sure that you retrieve as many citations as possible on the topic, even though automatic ‘mapping’ might include MeSH terms in the search as well, when this is possible. (see ‘How to confirm the process of a MEDLINE search’ above)
**Truncating text words:** text words can be *truncated* to include various forms of a word. For example, if you are looking for references that use the word *empirical*, you might also want to consider looking for the word *empirc*.

| Note | NB: Truncation removes the automatic MeSH term assignment and searching in MEDLINE. This is one reason why it is often important to check the details of the process of a MEDLINE search afterwards. See ‘How to confirm the process of a MEDLINE search’ above. |

**How to truncate text words**
In ‘PubMed’ an asterisk * is used for truncation eg empiric*. In ‘OVID’, colon ‘:’ or dollar ‘$’ symbols are used to truncate: eg empiric$ or empiric: and ‘ISI Web of Science’ MEDLINE uses an asterisk, like ‘PubMed’.

### 3.2.7 Combining terms (AND  OR  NOT)

After retrieving at least two separate sets of citations that discuss each of the key elements of your topic, you usually need to combine them, to identify the relevant articles among the separate sets. Sets can be combined in the following ways:

**AND** all terms must appear in each citation.
This will decrease the number of citations to review. Use ‘and’ to include all the important concepts in the citations retrieved.

**OR** either of the selected terms appears in each citation

In ‘PubMed’, searches are combined by typing the appropriate Boolean term (eg AND [and] OR [or]) between the terms. As AND [and] is the default, you can also just leave a space between the search components in the search box. Alternatively, you can also combine terms by clicking ‘History’ (ie below the search box and then combining individual searches (eg by typing ‘#1 AND #2’ to combine search #1 with search #2). ‘ISI Web of Science’ MEDLINE also uses ‘SEARCH HISTORY’ for combinations. The ‘COMBINE’ function in ‘OVID’ creates a new set on the Main Search page.

| Note | Only use the ‘Boolean’ operator ‘NOT’ when you are experienced in searching, as it is easy to lose relevant citations with this term. |

### 3.2.8 MEDLINE searching using ‘limits’

In MEDLINE, a number of ‘limits’ can be set.
These include limits for:
- **Human**, to only retrieve citations about human subjects. If you limit by ‘Age’ you do not need to use this limit.
- **Age** group
- **Language** of publication. Foreign language articles with English abstracts are eliminated by limiting a search to ‘English’.
- **Type of publication** eg Reviews, Practice Guidelines, Editorials, Clinical Trials
- **Abstracts** including citations with abstracts or summaries of the article.
- **Latest Update** includes citations from the most recent update of the database.
In ‘PubMed’ select limits by clicking on ‘Limits’, below the search box, on the ‘PubMed’ home page. With ‘ISI Web of Science’ MEDLINE, limits are called ‘Refine your results’ and are located on the ‘search Results- -Summary page, after a search. With ‘ISI Web of Science’ MEDLINE, to filter or reduce results on the Summary page, enter a topic query in the ‘search within results’ box, and then click ‘Search’: this search will return only those records in the original set that contain the topic term(s) you entered. In ‘OVID’ limits are listed directly below the box labelled ‘Enter Keyword or phrase’: clicking on ‘More Limits’ expands this list.

**Note**

It is usually best not to apply limits until the final stage of a search.

### 3.2.9 Searching with Filters

Inbuilt filters can be used to identify all articles indexed in MEDLINE, which have a particular set of methodological characteristics. Specific search filters can also be designed.

**Note**

‘PubMed Clinical Queries’, which is accessed from the left side panel on the ‘Entrez PubMed’ home page, is a very useful resource to start searching for questions of diagnosis, treatment, prognosis and cause, for systematic reviews clinical prediction guides and medical genetic searches.

In ‘OVID’ choose ‘More Limits’ after searching for citations, and then choose a particular filter from ‘Clinical Queries’.

### 3.2.10 Displaying the results

Once a search strategy has retrieved a reasonable set of citations, you can begin to review the results.

In ‘PubMed’, the ‘Display’ button on the home page enables a choice from the dropdown menu for displays other than in ‘Summary’ form. Choose sorting options by clicking on ‘Sort by’ and on ‘Send to’ to email the page.

### 3.2.11 Searches using ‘EndNote’

The University of Melbourne has a site licence for this software package. CD’s or passwords can be obtained from the Brownless Library for loading ‘EndNote’ on to your own computer. In addition to searching, ‘EndNote’ stores and manages references in files called ‘libraries’ for future perusal and for formatting in bibliographies, in the citation style of your choice. The University Library homepage has a link to further details about ‘EndNote’, which includes guides to its use and links to web-based tutorials. ‘EndNote’ libraries have an excellent ‘search’ facility for finding terms within them. This makes them very useful to store the references retrieved.

MEDLINE searching can be undertaken by searching MEDLINE through ‘PubMed’, ‘ISI Web of Science’ MEDLINE or ‘OVID’ first, as outlined above, and by then downloading and storing the
citations in your chosen ‘EndNote library’. Alternatively searches can be undertaken from ‘EndNote’ directly, after choosing from a broad variety of search fields.

When searching with ‘Endnote’, it is easy to include multiple search terms, to search in a large range of ‘fields’ from the ‘pull-down’ menu and to store relevant citations for future reference in your own ‘EndNote library’.

<table>
<thead>
<tr>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is strongly recommended that you learn to use ‘EndNote’ for searching, and also for storage of citations for your future reference.</td>
</tr>
</tbody>
</table>

### 3.2.12 ‘Troubleshooting’ searches in ‘primary’ databases

If you’re having difficulty finding enough relevant citations, consider the following:

- review the MeSH headings you have found to be assigned to publications that are of most interest and relevance to your search question. You can then, if you wish, construct an additional search strategy by including MeSH headings that you may not have considered. This is called a ‘cascade search’, and helps to ensure that you have identified the relevant MeSH headings.

- after choosing the most relevant citations, in ‘PubMed’ click on ‘Related Articles’ to retrieve additional citations. With ‘ISI Web of Science’ MEDLINE, ‘click on ‘PubMed Related Articles’, and in ‘OVID’ click on ‘Find Similar’.

- seek assistance from more experienced ‘searchers’, especially the librarians in teaching institutions, who are information experts!

### 3.2.13 ‘Secondary’ sources of evidence

When searching for answers to questions, it is often more efficient to search within these databases first, and to only search the ‘primary’ sources if the search is unsuccessful. One reason is that citations are selected for inclusion in the ‘secondary’ databases, so that there are fewer citations than in primary databases like MEDLINE; depending on the process of selecting citations, searches may be much more specific.

Several examples of ‘secondary’ databases are presented, but there are many others. In order to use ‘secondary’ databases properly, it is critical to know how the database has the potential to make you a ‘better and faster doctor’, and the process for selection and inclusion of citations. These potentials and processes vary greatly. For example:

- ‘Cochrane Reviews’ are high quality systematic reviews (see Chapter 5)
- ‘ACP Journal Club’ has an explicit process for selecting high quality ‘primary’ publications from selected journals, presents data in short, structured abstracts, and includes commentaries from practising physicians
- ‘Clinical Evidence’ contains summaries of the effectiveness of therapies for common conditions arising in general practice
- **BMJ updates** ([http://bmjupdates.com/index.asp](http://bmjupdates.com/index.asp)) comprises publications chosen from ~150 journals, selected according to their quality, and then rated for clinical relevance and newsworthiness by at least 3 practitioners for each discipline for which the article might be pertinent.
- **Bandolier** ([http://www.jr2.ox.ac.uk/bandolier/index.html](http://www.jr2.ox.ac.uk/bandolier/index.html)), amongst other functions does monthly searches of PubMed and the Cochrane for systematic reviews and meta-analyses published in the recent past, and publishes brief critical appraisals with comments on some of these.
Cochrane Library

This is available at no cost throughout Australia. It can be accessed from Australia through the National Institute of Clinical Studies (NICS) website [http://www.nicsl.com.au](http://www.nicsl.com.au). This also includes a ‘Cochrane User’s Guide’ from the ‘Quick Links’ box. The Library is also available through the ‘Clinicians Health Channel’, from your teaching hospital libraries, and directly from [http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME](http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME).

The Cochrane Library includes primary (The Cochrane Central Register of Controlled Trials, now called ‘Clinical Trials’) and secondary components, which can be searched separately or together. Some of the other components are indicated below:

- **Cochrane Reviews** are regularly updated, high quality systematic reviews, maintained by the various Cochrane Review Groups. This is a very useful secondary database for clinicians.
- **Database of Abstracts of Reviews of Effectiveness** (DARE), provides access to abstracts of systematic reviews, and is maintained by the National Centre for Reviews and Disseminations at the University of York. DARE includes structured abstracts of systematic reviews which have been critically appraised by reviewers at the NHS Centre for Reviews and Dissemination at the University of York, England. Systematic reviews are identified from a variety of sources and assessed according to a set of quality criteria. Structured abstracts describing the methodology, results and conclusions are produced. Comments on the overall quality of reviews and their implications for health care are also included. This is another secondary database.
- **Cochrane Central Register of Controlled Trials** (Clinical Trials) is a primary database of controlled trials. Most of these are also in MEDLINE, but are easy to find when grouped together in the Cochrane database.

<table>
<thead>
<tr>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Cochrane Library is different to MEDLINE in two important ways:</strong></td>
</tr>
</tbody>
</table>
| • **the default is to search all of the text** of publications, although you can choose to restrict the search to the title or abstract. Clicking on ‘History’ enables you to combines searches, as in ‘PubMed’.
| • **not all citations are allocated MeSH terms yet.** This means that you may need to search with text words as well as MeSH terms, when this might be important. |

The principles of using the Cochrane Library are otherwise similar to those for MEDLINE. The website has excellent guides and ‘Help’ sections.

ACP Journal Club

This secondary database will be available through ‘Supersearch’ at the University of Melbourne from late 2007. It can also be accessed online through your teaching hospitals via their library resources and through the ‘Clinicians’ Health Channel’. From the ‘Clinicians Health Channel’, for example, choose ‘A to Z of Journals’, then ‘Annals of Internal Medicine’, then ‘ACP Online’, then ‘Journals’. The effort is well worthwhile!

ACP Journal Club’s general purpose is to select published articles according to explicit criteria that warrant immediate attention by physicians attempting to keep pace with important advances in the treatment, prevention, diagnosis, cause, prognosis, or economics of the conditions managed by internists. These include a broad spectrum of general medical conditions. Included articles are summarised in structured abstracts. There are also comments by clinical experts. More than 100 journals are regularly reviewed for ACP Journal Club. A short list of key journals is included in...
each issue on the Table of Contents, and a full list is available on request. The summaries and comments are easily searched across a range of categories.

For more information see ACP-ASIM Online: http://www.acpj.org

Clinical Evidence
This very useful ‘secondary’ database is produced by the British Medical Journal (BMJ) group. It is accessed from your teaching hospitals via the ‘Clinicians’ Health Channel’ or through their library resources, but is not currently available through ‘Supersearch’ at the University of Melbourne. It is a regularly updated database of evidence summaries, which are relevant to primary and secondary care. The editorial process involves choosing questions of clinical significance, searching systematic reviews and clinical trials, and then summarising the data and presenting it in a structured format, with both detailed text and shorter summaries. It comprises ~25 sections, each of which has ~ 10 questions.

For more information see http://www.clinicalevidence.com, although the database is presented here in a different format than that in the ‘OVID’ interface, which is provided by the ‘Clinicians Health Channel’.

3.2.14 Searches using e-updated textbooks and ‘point of care’ resources
Electronically updated subscription textbooks like ‘American College of Physicians (ACP) Medicine’, as well as ‘point of care’ resources like ACP ‘Physicians’ Information and Education Resources (PIER)’ are often useful for supplementary searching of ‘primary’ databases for foreground questions. These will be available through the University of Melbourne Library ‘SuperSearch’ from late 2007. Some of these resources include direct links to ‘PubMed’ citations, which indicate MeSH terms allocated, which can then be used, if needed, for additional MeSH searches. These resources are also a major source of information for ‘background’ questions (see 2.2 above). Another electronically updated subscription textbook ‘UpToDate’ (see http://www.uptodate.com) is available through some of the teaching hospital libraries.

3.2.15 Searches using Search Engines eg ‘Trip’ and ‘Google’
At times you may retrieve useful citations from MEDLINE, and many other sources as well, through search engines like ‘Trip’, ‘Google’ and ‘Google Scholar’.

The ‘Trip’ search engine (see http://www.tripdatabase.com/index.html) searches a broad range of evidence based resources including synopses, guidelines and core medical journals.

Accessing ‘Google’ or ‘Google Scholar’ from ‘Supersearch’ through the University of Melbourne, provides direct access to many full text MEDLINE publications. Some clinicians start searching with ‘Google’ or ‘Google Scholar’, while many prefer to save these to supplement less satisfactory searches. This latter sequence is recommended at present, as ‘Google’ currently accesses only a portion of MEDLINE. Also, at this point of time, there is a bias in ‘Google’ for frequently cited and less recent publications, although this might change in the future.

3.2.16 Search Strategy
There is no all-inclusive generic search strategy! The particular search strategy, including the sequence you actually use on a particular occasion, depends on several factors. These include the type of question being asked, the time available, the purpose of the search, the thoroughness of
the search, the detail required and the availability of data sources, as well as your own experience in searching particular resources.

In this chapter the ‘primary’ databases have been presented before the ‘secondary’ databases, but in general the following strategy is recommended, especially for questions of therapy:

1. Search the appropriate ‘secondary’ database (eg ‘Clinical Evidence’; ‘Cochrane Systematic Reviews’; ‘ACP Journal Club’ etc).
   If necessary after this:
2. Use search filters (eg search ‘Clinical Queries’ in ‘PubMed’)
   If necessary after this:
3. Search in MEDLINE (‘PubMed’, ‘ISI Web of Science’ or ‘OVID’). Check how the search was processed by the database and, when appropriate, confirm that the search was conducted with MeSH terms.
   If necessary after this:
4. Define MeSH terms and use these to search in MEDLINE, using appropriate subheadings.
   If necessary after this:
5. Review and repeat the search using different terms and/or also search with e-updated textbooks and/or ‘point of care’ resources (see 3.2.14 above) and/or search-engines like ‘Trip’ and ‘Google’ (see 3.2.15 above) and/or search with other ‘primary’ databases and/or check ‘Related Articles’.

### 3.3 Worked example

The worked example is a clinical exercise. The purpose of the exercise is to practise how we can use MEDLINE to enhance patient care, by seeking appropriate data efficiently.

Consider your current strategies available for literature searching and keeping up to date:

Write them down:

1. ........................................................................................................................................................................................................................................

2. ........................................................................................................................................................................................................................................

3. ........................................................................................................................................................................................................................................

Do you know any easier ways to get published medical information, which might appeal to someone who wants to be a “better faster doctor”?

Now that you are sitting in front of a computer, how could you find this information using MEDLINE?
Getting started
For access to MEDLINE we will use ‘PubMed’. This is easily accessed through a web-browser such as ‘Google’ (http://www.google.com), by searching for Entrez PubMed. The web address is http://www.ncbi.nlm.nih.gov/entrez/query.fcgi. You might prefer to access ‘PubMed’ through ‘Supersearch’ via the University library, which means quick access to more full text publications. In this session we will:
• examine a clinical scenario
• refine a clinical problem by asking a potentially ‘answerable’ question
• search for citations which address the question posed

Remember Joyce? (see 2.3)

The structured question that arose from her request for information was:
**In people with knee osteoarthritis does physiotherapy improve pain and mobility?**

Question: how could you find information to answer this question using ‘PubMed’?
Answer: by using the components of the structured question as search terms.

Try searching in the ‘PubMed’ search box for ‘knee osteoarthritis’, then for ‘physiotherapy’, then for ‘pain’ and then for ‘mobility’.

1. Type in ‘knee osteoarthritis’ and then click ‘SEARCH’
   How many citations are retrieved? ____

2. Now type in ‘physiotherapy’ and then click ‘SEARCH’
   How many citations are retrieved? ____

3. Now type in ‘pain’ and then click ‘SEARCH’
   How many citations are retrieved? ____

4. Now type in ‘mobility’ and then click ‘SEARCH’
   How many citations are retrieved? ____

   **Note** You can quickly view the result of a series of searches by clicking on ‘History’, below the ‘PubMed’ search box. Try it!

Joyce is concerned about both ‘pain’ and ‘mobility’ as ‘outcomes’ from physiotherapy, so

5. Now type in ‘pain OR mobility’, to retrieve citations containing both of these terms and then click ‘SEARCH’
   How many citations are retrieved? ____

Next we will:
• Combine the search terms to try to find citations which answer Joyce’s question
• Limit the search
• Check the search ‘Details’ to see how the search terms used have been processed.
• Explore some of the MeSH terms used, and review concepts for MeSH and text word searches.
6. Now type in ‘knee osteoarthritis’ AND ‘physiotherapy’ AND ‘(pain OR mobility)’, to retrieve citations containing all of these concepts and then click ‘SEARCH’

How many citations are retrieved? 

We now have ~ 200 citations, which is a large number to peruse, so let’s limit the search.

7. Now click on ‘Limits’ to limit the last search by choosing ‘Human’, ‘English’ and ‘All Adult’. Click on ‘History’ to check the progress of the search again, and to confirm the limits you have set.

How many citations are retrieved? 

The number of citations is reduced, but there are still many to peruse. As this is a question relating to ‘therapy’, Joyce as well as you might consider limiting the search further, by choosing ‘Randomized Controlled Trial’ under ‘Type of Article’.

8. Now click on ‘Limits’ to limit the last search by choosing ‘Human’, ‘English’ and ‘All Adult’ and ‘Randomized Controlled Trial’ under ‘Type of Article’.

How many citations are retrieved? 

9. Now peruse the list of ~ 70 citations retrieved. You can change the ‘Display’, by choosing from the menu at the bottom left of the search results page, to check abstracts of citations, rather than the default ‘Summary’. You can check which particular MeSH terms were used for each citation by displaying ‘MEDLINE’ instead.

A MEDLINE search using ‘PubMed’ has been undertaken to answer Joyce’s question. You could check relevant titles and abstracts, and full text publications when appropriate. The next step is to pass on this information to Joyce (see Chapter 8), so that she can decide if she wishes to have physiotherapy.

Having found a relevant article we must firstly appraise it for validity (closeness to the truth) and relevance (usefulness in practice), and then decide on the clinical significance of the result, using critical appraisal worksheets as guides. Evaluating medical literature is a complex undertaking. You will find that the answers to the questions of validity are not always clearly stated, so you may have to use your own judgment about the importance and significance of each question. Finally we must use this information in conjunction with our patients’ preferences to decide on the next step in treatment.

The next chapter will get you started on the road to Critical Appraisal of articles about therapy. You could come back to this article after you have learned these skills and critically appraise this article so that you can apply the knowledge to help Joyce to decide about physiotherapy for her troublesome knee?

Below are some additional exercises, relevant to Joyce’s question, which are designed to help you understand more about the differences between MeSH term and text word searches, how to search with either of these and more about the processes involved with both MeSH term and text word searches.
10. Return to your last search, the one undertaken in task 8 above. You can find this by clicking ‘History’, and by then clicking on the number of citations, on the right-hand side of the screen, as in the pasted screen below:
11. Check the details of the last search (see 10 above), by clicking ‘Details’, below the search box. Note that the search has indeed included MeSH terms for the key elements, except for ‘mobility’. The screen is displayed below:


How many citations are retrieved?
Check in ‘History again, and compare this to the number of citations retrieved when MeSH terms were used as well as text terms (task 6 above)!


Search the ‘MeSH Database’ for ‘knee osteoarthritis’. Note that the MeSH term is ‘Osteoarthritis, Knee’, the subheadings you could choose to restrict a MeSH search to, the options to ‘Restrict Search to Major Topic Headings only’ and NOT to ‘Explode’ the MeSH term. From the MeSH trees at the bottom of the screen, note that the term Osteoarthritis, Knee’ is a end of a MeSH tree, so that no that more terms could be included by ‘Exploding’ anyway (see. 3.2.5 above).
Search the ‘MeSH Database’ for ‘physiotherapy’, and click on ‘Physical Therapy Modalities’. This is clearly the appropriate MeSH term to choose for this search. As above, note the subheadings you could choose to restrict the MeSH search to, and the options to ‘Restrict Search to Major Topic Headings only’, and NOT to ‘Explode’ the MeSH term. Also note that in contrast to the MeSH term ‘Osteoarthritis, Knee’, as the MeSH term ‘Physical Therapy Modalities’ is NOT at the (distal) end of the MeSH tree, if you choose NOT to explode, the retrieved citations will ONLY apply to this term, no the more distal, and clearly relevant terms for a search to the answer to Joyce’s question.

Note

You can quickly confirm exactly how a search was undertaken in ‘PubMed’ by clicking on ‘Details’ at the conclusion of a search.

3.4 Self-complete exercises

3.4.1
You are asked at 5:15pm by the resident on the surgical ward to see a patient with obstructive airways disease, who is scheduled for resection of a cancer of the sigmoid colon (partial colectomy) the next day.
You believe that the patient is at high risk of post-operative chest infection. You consider that peri-operative chest physiotherapy could reduce the risk of post-operative chest infection. Rather than just arranging physiotherapy, you wonder if there are published data confirming that it is worthwhile.

Your task is to:
- re-phrase this clinical problem as a structured question
- undertake a MEDLINE search to identify a relevant randomised controlled trials for critical appraisal

3.4.2
You are a general practitioner member of a Department of Human Services working party established to recommend measures to prevent illness amongst elderly people during an anticipated heat wave next summer.

Your task is to
- re-phrase the clinical problem as a structured question
- identify publications from MEDLINE which describe risk factors (see Chapter 7) for the development of heat stroke amongst the elderly.
3.5 Answers to self-complete exercise

3.5.1

Patients: those patients having colectomy
Intervention: physiotherapy
Comparison: preferably those not having physiotherapy
Outcomes: eg death; chest infection; time to getting out of bed; time to discharge; patient satisfaction; post-operative complications

If you used ‘physiotherapy’ and ‘colectomy’ as search terms in ‘PubMed’, you find that none of the few randomised trials address the question. In contrast, searching with ‘physiotherapy’ and ‘abdominal surgery’ leads to ~600 citations.

Checking in ‘Details’ shows that this search included some text words and the following MeSH terms: ‘abdomen’ (‘surgery’ as a subheading); ‘surgical procedures, operative’; ‘surgery; ‘physical therapy modalities’.

Limiting this search to ‘Randomised, Controlled Trials’ and ‘English’ results in ~50 citations, many of which are relevant to this question, including Fagevik Olsen M, Hahn I, Nordgren S, Lonroth H, Lundholm K. Randomised controlled trial of prophylactic chest physiotherapy in major abdominal surgery. British Journal of Surgery 1997;84:535-1538. The critical appraisal of this article is addressed in the next chapter.

3.5.2

Patients: elderly persons
Intervention (exposure): high ambient temperatures
Comparison: usual ambient temperature
Outcomes: heat stroke, heat exhaustion; falls; dehydration

Using ‘heat stroke’ and ‘heat exhaustion’ as search terms in ‘PubMed’ yields ~ 2000 citations, with ~ 170 after limiting citations to those published in English and to those >65 years, and ~10 with further limiting to ‘Reviews’.

3.6 Recommended reading

3.7 Further reading


- Giustini G. How Google is changing medicine. BMJ 2005; 331:1487-1488

- Steinbrook R. Searching for the right search-Reaching the medical literature. N Eng J Med 2006; 354:4-7


3.8 Websites

- The ‘Clinicians Health Channel’ has ‘Help cards’ and ‘Self directed learning’ links on its home page: there are guides to the ‘OVID’ portal to MEDLINE, to importing references into ‘EndNote’, and to the ‘Cochrane Library’. The ‘Clinicians Health Channel’ can be accessed through your teaching hospital computers. Ask the librarians in your hospitals to show you how. http://www.health.vic.gov.au/clinicians


- ‘ISI Web of Science’ MEDLINE has tutorials included (see ‘SuperSearch’ at http://www.lib.unimelb.edu.au


- Centre for Evidence-Based Medicine, Oxford (UK). An outstanding collection of tools and resources http://www.cebm.net/index.asp
EBP Workbook 07


- ‘Bandolier’ has a ‘Learning Zone’ and ‘EBM Glossary’ (as well as many critically appraised topics) [http://www.jr2.ox.ac.uk/Bandolier](http://www.jr2.ox.ac.uk/Bandolier)

- University of Alberta Evidence Based Medicine toolkit. [http://www.med.ualberta.ca/ebm/ebm.htm](http://www.med.ualberta.ca/ebm/ebm.htm)

- Canadian Centre for Health Evidence. This contains the ‘Users’ Guides to the Medical Literature’. [http://www.cche.net/usersguides/main.asp](http://www.cche.net/usersguides/main.asp)

- The Australian Centre for Evidence Based practice (Flinders University, Adelaide) has resources and links. [http://www.acebcp.org.au/intro.htm](http://www.acebcp.org.au/intro.htm)

- ‘Netting the Evidence’, from the University of Sheffield(UK) has resources and links. [http://www.shef.ac.uk/scharr/ir/netting](http://www.shef.ac.uk/scharr/ir/netting)

- ‘Duke University Medical Center Library Online’. There are ‘online’ tutorials containing self-test exercises. [http://www.mclibrary.duke.edu/training](http://www.mclibrary.duke.edu/training)

- Centre for Evidence-Based Medicine, University of Toronto. This website has on-line tutorials containing self-test exercises [http://www.cebm.utoronto.ca](http://www.cebm.utoronto.ca)

- ‘EPIQ’ (Effective Practice, Informatics & Quality Improvement) is produced by Professor Rod Jackson (University of Auckland). It has very useful materials including notes and tools. [http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm](http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm)
4. ARTICLES ABOUT THERAPY

4.1 Learning objectives

- To understand the need for medical students and practising clinicians to be able to critically appraise articles about health care interventions.
- To evaluate publications about therapy in order to detect fatal flaws in the methods or study design, which threaten the validity and relevance of the findings of the study.
- To understand the importance of clinical significance and the different measures of treatment effect and effect size.
- To calculate the Number Needed to Treat (NNT) from trial results reporting binary outcomes.
- To apply the results of particular studies to individual patients.

4.2 Core reading and key concepts

Introduction

Being able to critically evaluate the findings of research reports about therapy is fundamental to the practice of evidence-based practice. Pre-appraised articles about health care interventions are becoming increasingly available but medical students and clinicians also need to make independent judgements about the validity of findings from trial reports.

Clinicians often rely on opinion leaders to guide clinical practice. Expert opinion clearly has an important role in influencing clinical practice but relying on expert opinion alone has limitations. There is evidence to suggest that expert opinion and textbook recommendations may be inconsistent with, or lag behind evidence from clinical trials. Colleagues may provide clinical expertise and judgement but some of their beliefs may not be evidence based. In the current clinical environment providing effective care relies on clinicians being able to locate and critically appraise research reports that address their information needs as they arise during day-to-day practice.

In order to be able to locate and appraise relevant research reports, clinicians need to understand how to choose the most appropriate study design to answer the question at hand, how to evaluate the internal validity of the study, how to assess the clinical significance of the study, and how to apply the results of the study to the care of individual patients.

Critical appraisal guides have been developed to assist with the critical appraisal process. An extensive series of Users’ Guides have been published in JAMA (see http://www.cche.net/usersguides/main.asp). These guides have adopted a similar format to assess the validity of studies designed to assess a diverse range of clinical questions. Three basic questions are addressed:

1. Are the results of the study valid?
2. What are the results?
3. Will the results help me in caring for my patients?

The full critical appraisal guide for articles about therapy is summarised at the end of this section. Some important concepts relevant to interpreting the results of studies are also outlined below.
The evidence hierarchy

When making decisions about clinical interventions different types of research or sources of information are given different weight. ‘Levels of evidence’ have been developed to rank the validity of evidence. Levels are allocated solely on the basis of the study design and are graded according to the likelihood of bias affecting studies with this design. The ‘level’ of evidence is necessary, but insufficient to decide the overall quality of a study. Most of the rankings for ‘level of evidence’ developed follow the hierarchy set out below, with minor variations.

1. Systematic reviews and meta-analyses of randomised controlled trials
2. Randomised controlled trials with definitive results (a result with confidence intervals that do not overlap the threshold clinically significant effect)
3. Randomised controlled trials with non-definitive results (point estimate that suggests a clinically significant effect, but with confidence intervals overlapping the threshold for this effect
4. Cohort studies
5. Case-control studies
6. Cross sectional surveys
7. Case reports/case series
8. Expert opinion (based on physiology, laboratory research, first principles, or experience)

A high quality systematic review with definitive results and homogeneity in the results of studies from individual randomised trials included, represents the highest level of evidence. However, regardless of the type and source of reviews or studies, all evidence requires critical appraisal. For example, a well-designed large cohort study might provide more valid evidence than a randomised-controlled study with serious methodological flaws.

Equivalence (‘noninferiority’) trials

In this workbook only ‘superiority’ trials are considered, where the issue is whether one treatment is better than another. The process is to reject the null hypothesis of equal efficacy. An increasing number of ‘equivalence’ trials, however, are now being undertaken, where the objective is to establish equivalence or ‘noninferiority’ within specified limits. Special considerations apply to equivalence trials, which are not considered further in this workbook.

Tests of statistical significance and random error

Random error may potentially influence the results of all studies. Statistical tests can be applied to assess the possible role of chance in the study results. Clinical trials are conducted on samples of patients rather than all of the possible patients, and statistical tests are used to help make inferences about the ‘truth’.

If a difference in outcome is observed in a clinical trial comparing two groups (e.g. treatment and control groups), this difference could either represent a true difference or it may be due to random variation. A type I statistical error arises when we conclude that a difference, that is due to chance alone, represents a true difference. Inferential statistics are used to determine the likelihood of a type I error. The p value quantifies the probability that an observed difference between two study groups might have arisen due to chance alone, assuming that there is in fact no difference between the groups. Traditionally p < 0.05 is used as a cut off for determining ‘statistical significance’.

While statistical tests are important, other factors must also be taken into account when interpreting results. A highly statistically significant result (e.g. p< 0.0001) could be misleading if the study methods were susceptible to bias or systematic error. Systematic error may arise from any process that causes observations to deviate systematically from the true values.
Clinical significance

It is often possible for studies (particularly large studies) to produce statistically significant results that may be of doubtful clinical significance. When assessing the clinical significance of trial results both the types of outcome measures being reported and the size of the effect should be taken into account. Some investigators make the distinction between patient-oriented outcomes ('clinical') and disease-oriented ('process') outcomes. Examples of patient oriented outcomes are morbidity, mortality, symptom reduction, quality of life and other end points of importance to patients. Examples of disease-oriented outcomes include lung function tests, glycosylated haemoglobin level, left ventricular function etc. The effect size is another important consideration. In some circumstances this is easy to judge, for example, a cancer treatment that prolongs life by an average of 1 to 2 weeks would not be considered clinically significant but a treatment that prolonged life by an average of 12 months would be considered significant. For some outcome measures the Number Needed to Treat (NNT) is a useful way to express the absolute outcomes. This is discussed under measures of treatment effect.

Bias and validity

The validity of a study is the extent to which its design and conduct are likely to prevent systematic errors, or bias. More methodologically rigorous studies may be more likely to yield results that are closer to the ‘truth’. There are four main types of systematic errors in trials examining healthcare interventions: selection bias, performance bias, attrition bias and detection bias.

Table 4.1: Steps involved in the conduct of randomised controlled trials and potential sources of bias

<table>
<thead>
<tr>
<th>STUDY PROCESS</th>
<th>SOURCES OF BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation of subjects to intervention and control groups</td>
<td>Selection bias: systematic differences in comparison groups</td>
</tr>
<tr>
<td>Implementation of study interventions</td>
<td>Performance bias: systematic differences in care provided apart from the intervention being studied.</td>
</tr>
<tr>
<td>Follow up of participants</td>
<td>Attrition bias: systematic differences in withdrawals from the trial.</td>
</tr>
<tr>
<td>Evaluation of outcomes</td>
<td>Detection bias: systematic differences in outcome assessment</td>
</tr>
</tbody>
</table>

Selection bias

Randomised controlled studies will be susceptible to allocation and selection bias if the randomisation process is inadequate. In particular inadequate allocation concealment is an important source of bias. With appropriate methods, random allocation removes the potential for bias in the assignment of patients to control and intervention groups. However when methods such as alternation or hospital UR numbers are used to allocate patients, it is possible for investigators to subvert the process. When assessing a potential participant’s eligibility for a trial, investigators recruiting participants should remain unaware of the next assignment in the sequence until the decision about eligibility has been finalised. This can be achieved if the assignment schedule (generated by an appropriate randomisation method, such as computer generated random numbers) is administered by someone who is not responsible for recruiting subjects. A central trial office or pharmacy department might be used for this purpose.
Performance bias
Performance bias may influence the results of a study if there are systematic differences in the care provided to the comparison groups other than the intervention being investigated. Double blinding is often used to prevent this type of bias. In this situation neither the patients nor the investigators are aware of which treatment the participant is receiving. Double blinding also protects against placebo effects. For research with subjective outcome measures such as pain, blinding is particularly important\textsuperscript{12,13}.

Attrition bias
Losses to follow up or dropouts, which are systematically different between comparison groups, may lead to attrition bias. Participant losses may occur because participants are withdrawn from the study, dropout or because of protocol violations. Often these details are poorly reported in trial results.

Detection bias
Where outcome assessment differs systematically between comparison groups, detection bias may occur. Outcome assessment should be blinded, where possible.

Selective reporting of outcomes
For studies that use multiple outcome measures selective reporting of ‘positive’ outcome measures may introduce bias. For example in a study of pulmonary rehabilitation for COPD, several outcomes may be examined, e.g. Lung function tests, exercise tests, quality of life, utilisation of health care services. The investigators may choose to report only those outcome measures that are positive in published results.

Precision and confidence intervals
Precision is another measure of the likelihood of random errors. It is reflected in the confidence interval around the estimate of effect. Confidence intervals can also be used to examine statistical significance and provide us with more information than p values. In particular, confidence intervals do more than indicate whether the results might be a chance effect. They also show, allowing for the role of chance, how small or how large the true effect size could be, and how ‘precise’ the data are\textsuperscript{14}.

The exact effect size (treatment effect observed in a particular study) is called the point estimate. This is the best estimate from the study of the true effect size. The precision or stability of this estimate may be expressed using confidence intervals. By convention, 95\% confidence intervals are usually reported in research studies but in some circumstances other ranges might be used such as 90\% or 99\% confidence intervals.

The narrower the confidence interval the more certain one can be about the size of the true effect. If an unbiased study reports a 95\% confidence interval this means that there is a 95\% chance that the true effect lies within the confidence interval.

If the value corresponding to no effect is not included in the 95\% confidence intervals then the results are statistically significant at the 0.05 level of significance. If the confidence intervals include the null value (the value suggesting no difference between the two groups) then the results are not statistically significant at the 0.05 level of significance.

Studies with large numbers of participants will usually give more precise estimates of effectiveness and hence, narrower confidence intervals\textsuperscript{6,14}.
**Negative studies**

Confidence intervals also help us to interpret the results of studies that appear to be negative. When the results of a study are negative this may either be a true negative or a false negative result. A false negative result is also termed a type II statistical error and may arise if the study has insufficient statistical power to detect a true difference between groups. Examining the upper end of the confidence interval can be helpful when interpreting the results of negative studies. A trial is definitely negative if the smallest benefit of clinical importance lies above the upper boundary of the confidence interval. On the other hand, if clinically important benefits lie within the confidence interval, the trial has not excluded the possibility that the treatment may be effective\(^5\).

![Figure 4.1 How to interpret confidence intervals](image)

This figure displays the range of possible results from a hypothetical study. The study is randomised controlled with an intervention and control group. Subjects in the study are at risk for ischaemic heart disease and the intervention group receive a new therapy aimed at reducing death from ischaemic heart disease. The control group receive standard treatment. A relative risk of less than 1 suggests that the new treatment has a protective effect but for this study, a relative risk of less than 0.6 is considered clinically significant.

**Measures of treatment effect**

The results of studies with binary outcomes (yes/no) may be expressed in several ways. The relative benefit of an active treatment over a control is often expressed as the relative risk, the relative risk reduction, or the odds ratio. For applying the results of trials to clinical practice however, the Number Needed to Treat (NNT) is more meaningful. Some important measures of treatment effect are described below.
Consider a hypothetical parallel group randomised study with an active treatment group and a placebo control group. The subjects are followed up for a set period of time and the object of the study is to prevent an adverse event (e.g. death) and the response to treatment is measured as a binary outcome (event vs no event). The event rate in the placebo group is \( X \) and the event rate in the treatment group is \( Y \).

**Absolute risk reduction** (ARR) is the difference in the proportion of subjects with the outcome of interest in each group, in this case \( \text{ARR} = X - Y \).

**Relative risk** (RR) is defined as the probability of an event in the active treatment group divided by the probability of an event in the control group. In this case the \( \text{RR} = Y / X \). A relative risk of 1 is the null value or no difference.

**Relative risk reduction** (RRR) is related to the relative risk and can be calculated by subtracting the relative risk from 1. The relative risk reduction can be thought of as a standardised measure of the absolute risk reduction and can be expressed as the absolute risk reduction divided by the probability of an event in the control group. In this case \( \text{RRR} = 100 \times (X - Y) / X \).

**Number needed to treat** (NNT) is the number of patients you need to treat to prevent one additional bad outcome. The number needed to treat is the reciprocal of the absolute risk reduction (\( \text{NNT} = 1 / \text{ARR} \)). The 95% confidence intervals for the NNT can be constructed by simply inverting and exchanging the limits of a 95% confidence interval for the absolute risk reduction\(^1\). Calculating confidence intervals for negative studies (when the confidence interval for ARR includes 0) requires alternative methods but these are beyond the scope of this course\(^1\).

**Applying the results of studies to patients with different baseline risks**

Relative measures of treatment effect can be limited when applied to clinical decision-making. Relative measures can provide a useful estimate of treatment effect for broad groups of patients, when the benefit of a specific treatment (e.g. the relative risk reduction) remains roughly the same for patients with different baseline risks. Often it is useful however, to consider the baseline (control) risk of an event before making recommendations about treatment. For example for patients at high risk of an adverse event a relative risk reduction of 50%, which is statistically significant, is likely to be clinically significant. On the other hand for patients with a low baseline risk this level of risk reduction may not be sufficient to justify the possible harms (side effects) and costs of treatment\(^1\).

The Number Needed To Treat (NNT) can be useful for comparing the likely benefits of treatment with a particular therapy across groups of patients with varying levels of baseline risk for the adverse event(s). The simplest method is to compare the baseline risk of an individual patient with that of the typical patient in the published study. If the baseline risk for a particular individual is a factor of \( f \) times the baseline risk of a typical study patient and the relative risk remains constant, the absolute risk reduction for the patient is scaled according to the same factor \( f \). The corresponding number needed to treat will be the estimated number needed treat from the study results divided by the factor \( f \)\(^1\).

**Example**


This study examines the risk of stroke after myocardial infarction. The experimental group were treated with pravastatin and the control group received placebo. There was a 32% risk reduction in
the risk of stroke in the experimental group corresponding to a number needed to treat of 83. The results are outlined in table 5.2. The possible effect of extrapolating the findings of the study to a lower risk population (for example those without a history of AMI) is illustrated in the table (bottom line).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experimental event rate (Y)</th>
<th>Relative risk reduction RRR = (X – Y)/X</th>
<th>Absolute risk reduction ARR = X – Y</th>
<th>Number needed to treat NNT = 1/ARR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke (after AMI)</td>
<td>3.7%</td>
<td>2.5%</td>
<td>32%</td>
<td>1.2%</td>
<td>83</td>
</tr>
<tr>
<td>Stroke (risk at baseline reduced by half)</td>
<td>1.9%</td>
<td>1.3%</td>
<td>32%</td>
<td>0.6%</td>
<td>167</td>
</tr>
</tbody>
</table>

* Event rate = cumulative incidence of major event outcomes during study period

**External validity/generalisability**

The results of randomised controlled studies may not be generalisable to all patients outside the clinical trial setting. Randomised controlled trials usually address the question of whether a treatment can work (efficacy) but may not tell us about whether the treatment will be effective when offered to the broad range of patients seen in day to day clinical practice. Randomised controlled studies tend to enrol only a small proportion of the potential population of patients with the disease of interest. Often extensive inclusion and exclusion criteria are used in order to limit sources of heterogeneity that may influence the study results.

Applicability is closely related to the concepts of generalisability and external validity but it also encompasses broader issues relating to the overall impact of the therapy on individual patients.

Factors that may affect the applicability of trial results to individual patients:
- Differences in the pathophysiological basis for disease in different populations.
- Between population differences in the response to treatment (e.g. differences in drug metabolism or immune responses)
- Differences in patient compliance
- Differences in baseline risk for adverse events being targeted by the treatment.
- Presence of comorbid conditions that may alter the potential benefits and risks of treatment.
- Feasibility of implementing treatments in different clinical settings (e.g. factors such as technical requirements for safe and effective administration of therapy, availability of trained staff and cost may influence decisions to offer therapies in different settings.)

**Critical appraisal guides**

Numerous critical appraisal guides have now been published. The guide summarised below has been taken from Guyatt et al 1993.
Are the results of this study valid?
Was the assignment of patients to treatment randomised?
Were all patients who entered the trial properly accounted for and attributed at its conclusion?
Was follow up complete?
Were patients analysed in the groups to which they were randomised?
Were patients, health workers, and study personnel ‘blind’ to treatment?
Were the groups similar at the start of the trial?
Aside from the experimental intervention, were the groups treated equally?

What were the results?
How large was the treatment effect?
How precise was the estimate of the treatment effect?

Will the results help me in caring for my patients?
Can the results be applied to my patient care?
Were all clinically important outcomes considered?
Are the likely treatment benefits worth the potential harms and costs

4.3 Worked example

Introduction
You are the resident admitting a patient, Mr PT, scheduled for an elective Nissen fundoplication operation the following day. The PT is a 55-year old man who is overweight and stopped smoking a month ago. You believe that he is at high risk of developing a postoperative chest infection.

You are aware of a broad discussion about the potential role of prophylactic chest physiotherapy in reducing the risk of such events.

You wonder “Will prophylactic physiotherapy reduce the risk of post-op complications in my patient?”

You conduct a MEDLINE search and you find the following article:


Critically appraise this article using the worksheet provided.
Abstract

**Question:** In patients having major abdominal surgery, can prophylactic chest physiotherapy reduce pulmonary complications after surgery?

**Design:** Randomised controlled trial.

**Setting:** University hospital in Goteborg, Sweden.

**Patients:** 368 patients who were 19 to 92 years of age (mean age 53 y, 57% women) and were having elective open abdominal surgery. Patients were considered to be high risk if they were > 50 years of age and had >= 1 of the following risk factors: current or previous (quit within the past 12 months) history of smoking, body mass index > 30 kg/m², pulmonary disease requiring medication, or another medical condition causing reduced ventilatory function. 79 patients (21%) were high risk.

**Intervention:** Patients were allocated to prophylactic chest physiotherapy (n = 174) or usual treatment (n = 194). Patients in the physiotherapy group received information and training the day before surgery from a physiotherapist. Training involved breathing exercises with pursed lips and hourly huffing and coughing (30 deep breaths with huffing and coughing between every 10th breath every daylight hour after the operation). The importance of changing position in bed and getting out of bed soon after surgery was emphasised. High-risk patients used positive respiratory pressure masks for respiratory resistance training during the breathing cycle. Patients in the usual-treatment group did not receive physiotherapy training or information before the operation but did receive chest physiotherapy with the mask if pulmonary complications developed after surgery.

**Main outcome measures:** Pulmonary complications defined as oxygen saturation < 92% or 2 of the following: temperature > 38.2°C, pathological lung auscultation, or radiologically confirmed pneumonia or atelectasis.

**Main results:** Fewer patients who received physiotherapy had pulmonary complications after surgery than did patients who received usual treatment (P < 0.001); this reduction was also seen in both high-risk and low-risk patients (P < 0.001 for both) (Table). Greater oxygen saturation was maintained in the physiotherapy group during the first 3 days after surgery (P < 0.05). The groups did not differ for body temperature or need for oxygen support.

**Conclusion:** Prophylactic chest physiotherapy done before and after major abdominal surgery reduced pulmonary complications.

**Are the results of the study valid?**

**Was the assignment of patients to treatments randomised?**

The authors state “to avoid patient interference a cluster randomisation was performed to alternate months”. The use of alternation would usually be considered inadequate because assignment to groups is not concealed. This may introduce selection bias (such bias may occur in either direction).

**Were all patients who entered the trial properly accounted for and attributed at its conclusion?**

Yes.

**Was follow-up complete?**

Four patients did not complete the study and the reasons were described.
**Were patients analysed in the groups to which they were randomised?**
Although not specifically stated in the article the analysis appears to be intention to treat.

**Were patients, health workers, and study personnel ‘blind’ to treatment?**
The study was not blinded. Although blinding to this type of intervention might be difficult it should be possible to have a blinded assessment of outcome. Lack of blinding means the study is susceptible to performance and detection bias.

**Were the groups similar at the start of the trial?**
Yes.

**Aside from the experimental intervention, were the groups treated equally?**
The treatment group received both pre operative and postoperative prophylactic physiotherapy. The control group did not receive any information or training. Postoperative physiotherapy was withheld unless the patient developed a pulmonary complication. Because the investigators were not blinded the possibility of performance bias exists.

**What are the results?**
For prevention of pulmonary complications (oxygen saturation of less than 92% or 2 of the following: temperature greater than 38.2°C, pathological lung auscultation, and radiological evidence of pneumonia or atelectasis) the results are summarised below.

**Patient groups: All patients**
Physiotherapy: 6% (10/172)
Usual treatment: 27% (52/192)

**Patient groups: High-risk patients**
Physiotherapy: 15% (6/40)
Usual treatment: 51% (20/39)

**Patient groups: Low-risk patients**
Physiotherapy: 3% (4/132)
Usual treatment: 21% (32/153)

**How large was the treatment effect?**

<table>
<thead>
<tr>
<th>Table 4.3: Clinically useful measures of the effects of treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>All patients</td>
</tr>
<tr>
<td>High risk patients</td>
</tr>
<tr>
<td>Low risk patients</td>
</tr>
</tbody>
</table>
Therefore, for high risk patients 3 patients need to be treated to prevent 1 pulmonary complication arising post operatively.

**How precise was the estimate of the treatment effect**
The confidence intervals presented above have been taken from an ACP Journal Club review article but were not presented in the original article. The confidence intervals are narrow and hence the estimates of effectiveness are reasonably precise. For all patients, there is a 95% chance that the true NNT lies between 3 and 7.

**Will the results help me in caring for my patient?**
Can the results be applied to my patient?

Mr PT is similar, both clinically and in demographic terms to the patients in the study. You also need to consider whether there are physiotherapists, trained in the methods described in the study, available at your institution.

**Were all clinically important outcomes considered?**
Pneumonia is clearly an important clinical outcome, hypoxia and temperature changes alone are essentially surrogate markers. Interestingly despite the apparent benefits of the intervention, length of stay did not differ between the 2 groups.

**Are the likely treatment benefits worth the potential harms and costs?**
There were no adverse events reported, but these would be expected to be minimal with this type of intervention. Cost would be an important consideration, but this was not discussed.

### 4.4 Self-complete exercise

You are the resident assessing a patient with an infective exacerbation of chronic obstructive pulmonary disease (COPD) in the emergency department.

Mrs DW is a 72-year old woman with known COPD. She regularly attends the outpatient department but has not previously required admission for COPD. Her past medical history includes hypertension and late onset diabetes (managed with dietary modifications alone). She stopped smoking two years ago, having previously smoked approximately a packet of cigarettes per day for 40 years. Her usual medications include salbutamol 200mg MDI 2 puffs 6 hourly, ‘Atrovent MDI’ 2 puffs 6 hourly and ‘Diltiazem Slow Release’ 240mg oral daily. She has no history of asthma and her previous lung function tests demonstrate severe fixed airflow obstruction with a baseline FEV₁ of 0.9 Litres.

She now presents with a week of increasing shortness of breath and productive cough. After taking a history and examining her you conclude that she has an infective exacerbation of COPD. There is no evidence of pneumonia on her chest x-ray but she is hypoxic and requires admission for oxygen therapy and treatment of airways disease. You commence treatment with regular nebulised bronchodilators, intranasal oxygen (2 litres/minute), oral antibiotics and intravenous hydrocortisone 250mg IV QID.

You are aware that systemic corticosteroids are usually used for the management of exacerbations of COPD but you have heard colleagues cast doubt on the value of this treatment in the past. You decide to find out what evidence there is available to support the use of systemic corticosteroids for exacerbations of COPD and in particular whether this treatment should be offered to your patient.
You conduct a MEDLINE search and you find the following article:


### Abstract

**BACKGROUND AND METHODS**

Although their clinical efficacy is unclear and they may cause serious adverse effects, systemic glucocorticoids are a standard treatment for patients hospitalised with exacerbations of chronic obstructive pulmonary disease (COPD). We conducted a double blind, randomised trial of systemic glucocorticoids (given for two or eight weeks) or placebo in addition to other therapies, for exacerbations of COPD. Most other care was standardized over the six-month period of follow-up. The primary end point was treatment failure, defined as death from any cause or the need for intubation and mechanical ventilation, readmission to the hospital for COPD, or intensification of drug therapy.

**RESULTS**

Of 1840 potential study participants at 25 Veterans Affairs medical centers, 271 were eligible for participation and were enrolled; 80 received an eight-week course of glucocorticoid therapy, 80 received a two-week course, and 111 received placebo. About half the potential participants were ineligible because they had received systemic glucocorticoids in the previous 30 days. Rates of treatment failure were significantly higher in the placebo group than in the two glucocorticoid groups combined at 30 days (33 percent vs. 23 percent, \(P=0.04\)) and at 90 days (48 percent vs. 37 percent, \(P=0.04\)). Systemic glucocorticoids (in both groups combined) were associated with a shorter initial hospital stay (8.5 days, vs. 9.7 days for placebo, \(P=0.03\)) and with a forced expiratory volume in one second that was about 0.10 litre higher than that in the placebo group by the first day after enrolment. Significant treatment benefits were no longer evident at six months. The eight-week regimen of therapy was not superior to the two-week regimen. The patients who received glucocorticoid therapy were more likely to have hyperglycemia requiring therapy than those who received placebo (15 percent vs. 4 percent, \(P=0.002\)).

**CONCLUSIONS**

Treatment with systemic glucocorticoids results in moderate improvement in clinical outcomes among patients hospitalised for exacerbations of COPD. The maximal benefit is obtained during the first two weeks of therapy. Hyperglycemia of sufficient severity to warrant treatment is the most frequent complication.
4.5 Answers to self-complete exercises

Are the results of this study valid?

Was the assignment of patients to treatment randomised?
Yes the study was randomised. Allocation concealment appears to be adequate (study medications were administered by pharmacy department blind to treatment allocation).

Were all patients who entered the trial properly accounted for and attributed at its conclusion?
No, see below.

Was follow up complete?
Follow up details were not well outlined in the article. Study drugs were discontinued for reasons other than the primary end point in 25 patients and follow up data was not complete for 6 of these. Presumably these are the only losses to follow up in the study but this is not explicitly stated in the article.

Were patients analysed in the groups to which they were randomised?
Yes the analysis was by intention to treat.

Were patients, health workers, and study personnel ‘blind’ to treatment?
Yes the study was double blind and placebos were used.

Were the groups similar at the start of the trial?
Refer to table 1 in the article. As the authors point out there were small differences in the total pack years of cigarette smoking, prior use of systemic glucocorticoids, and the prevalence of diabetes mellitus. These differences are likely to be of limited clinical importance.

Aside from the experimental intervention, were the groups treated equally?
Yes, again the study was double blinded using placebo medications. There was no reason to suspect ‘unmasking’ of treatment allocation for whatever reason during the study period.

What were the results?
For the primary outcome measure (death, intubation, readmission for COPD, or intensification of therapy) the results for 30 and 90 days are outlined below:
Table 4.4: Clinically useful measures of the effects of treatment

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experimental event rate (Y)</th>
<th>Relative risk reduction RRR = (X – Y)/X</th>
<th>Absolute risk reduction ARR = X – Y</th>
<th>Number needed to treat (NNT = 1/ARR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 days</td>
<td>33%</td>
<td>23%</td>
<td>31% (95%CI: -2 to 53)</td>
<td>10%</td>
<td>10 (95% CI: 5 to infinity)</td>
</tr>
<tr>
<td>90 days</td>
<td>48%</td>
<td>37%</td>
<td>23% (-3 to 42)</td>
<td>11%</td>
<td>9 ( 4 to infinity)</td>
</tr>
</tbody>
</table>

Confidence intervals for RRR and NNT were not provided in the article. They are taken from an ACP Journal Club review\(^{21}\). The investigators used survival analysis to analyse their results and the point estimate for the RRR above may underestimate the survival analysis reported in the article.

**How precise was the estimate of the treatment effect?**
The confidence intervals presented above are relatively wide but these may be an overestimate. Nonetheless they do suggest that the true effect size could be a lot smaller than the point estimate.

**Will the results help me in caring for my patients?**

*Can the results be applied to my patient care?*
You work at a tertiary institution probably similar to the ones in the study. In general, your patient appears to fit the profile of many of the patients included in this study except that the majority (99%) of participants in the study were males. It is worth noting that some studies suggest that females attending the emergency department with COPD have a better prognosis than males\(^{22}\). If the baseline risk of adverse events is lower in female subjects then the NNT to prevent them would be greater compared with males.

It is worth noting that lower doses of corticosteroids are used to treat exacerbations of COPD in Australia than were used in this study. Evidence from asthma studies, however, suggests that lower doses are likely to be sufficient. Oral administration might also be adequate but there has been little research to evaluate this.

*Were all clinically important outcomes considered?*
Most clinically important outcomes were considered for this study, but quality of life would be another outcome of interest to patients.

*Are the likely treatment benefits worth the potential harms and costs?*
The main side effect was hyperglycaemia requiring treatment. The rate of hyperglycaemia was 15% in the glucocorticoid treated group compared with 4% in the placebo group. This suggests that for every 9 patients treated with glucocorticoids, 1 will develop hyperglycaemia requiring treatment. However, if short lived and uncomplicated this side effect is probably of small clinical importance when weighed against the benefits of treatment. Although not significant in this study other studies have shown an elevated risk of secondary infection, myopathies and other complications with systemic corticosteroids\(^{22}\).
4.6 Recommended reading

- See [http://www.cche.net/usersguides/main.asp](http://www.cche.net/usersguides/main.asp) and read “Therapy or Prevention" OR see 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? Evidence-Based Medicine Working Group. JAMA 1993; 270:2598-2601 and 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? Evidence-Based Medicine Working Group. JAMA 1994; 271:59-63

- ‘EPIQ’ (Effective Practice, Informatics & Quality Improvement) is produced by Professor Rod Jackson (University of Auckland). It has very useful materials including notes and tools. [http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm](http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm)

- Cautionary tales in the clinical interpretation of therapeutic trials reports
  Scott IA, Greenberg, PB. Internal Medicine Journal 2005; 35:611-621

4.7 References


   http://www.cochrane.dk/cochrane/handbook/hbook.htm


18. Dans AL,Dans LF et al. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. JAMA 1998;279:545-549

19. Guyatt G, Sackett DL et al. Users' guides to medical literature: II. How to use an article about therapy or prevention. A. Are the results of the studies valid? JAMA. 1993; 270:2598-2601


4.8 Critical appraisal worksheets

4.8.1 Example adapted from Centre for Evidence-Based Medicine, Mount Sinai Hospital
http://www.cebm.utoronto.ca/teach/materials/therapy.htm

Therapy worksheet

<table>
<thead>
<tr>
<th>Are the results of this study valid?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the assignment of patients to treatments randomised? Was the randomisation list concealed?</td>
</tr>
<tr>
<td>Was follow-up of patients sufficiently long and complete?</td>
</tr>
<tr>
<td>Were all patients analysed in the groups to which they were randomised?</td>
</tr>
</tbody>
</table>
### EBP Workbook 07

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were patients and clinicians kept 'blind' to treatment?</td>
<td></td>
</tr>
<tr>
<td>Were the groups treated equally, apart from the experimental treatment?</td>
<td></td>
</tr>
<tr>
<td>Were the groups similar at the start of the trial?</td>
<td></td>
</tr>
</tbody>
</table>

**What were the results?**

How large was the treatment effect?

________________________________________________________________________
________________________________________________________________________

How precise was the estimate of the treatment effect?

________________________________________________________________________
________________________________________________________________________
Your calculations

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experimental event rate (Y)</th>
<th>Relative risk reduction ( RRR = \frac{(X - Y)}{X} )</th>
<th>Absolute risk reduction ( ARR = X - Y )</th>
<th>Number needed to treat ( NNT = \frac{1}{ARR} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Will the results help me in caring for my patient?**

<table>
<thead>
<tr>
<th>Can these results be applied to my patient care?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is your patient so different from those in the study that its results cannot apply?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the treatment feasible in your setting?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Were all clinically important outcomes considered?</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Are the likely treatment benefits worth the potential harms and costs?</td>
</tr>
</tbody>
</table>

**Additional Notes**

4.8.2 Example adapted from Critical Appraisal Skills Programme (CASP)
See [http://www.phru.nhs.uk/casp/appraisa.htm](http://www.phru.nhs.uk/casp/appraisa.htm)
5. **SYSTEMATIC REVIEWS**

5.1 **Learning objectives**

- To understand the role of systematic reviews in evidence based practice decision making
- To understand the steps involved in undertaking a systematic review
- To understand the limitations of systematic reviews and meta-analyses
- To be able to critically appraise systematic reviews of controlled trials.

5.2 **Core reading and key concepts**

**Introduction**

You are a general practitioner, and are consulted by a 45-year-old woman. She has had rhinorrhoea for 10 days and she complains that for the last 7 days she has had throbbing pain over the right maxilla that is only partially relieved with paracetamol. She also describes purulent post-nasal discharge. On examination she has tenderness over the right maxillary sinus. There is no relevant past history apart from allergy to co-trimoxazole.

After completing your history and examination you conclude that she has acute sinusitis, which may either be viral or related to a secondary bacterial infection. You prescribe cefaclor: 375mg orally twice daily for 5 days. You ask her to return if there has been no improvement after completing this course of antibiotics.

You often see patients with acute sinusitis in your practice and over lunch you discuss your management practices with your colleagues. You often prescribe cephalosporins because of concerns about possible resistance to amoxycillin. However, a colleague of yours comments that she does not always prescribe antibiotics for uncomplicated acute sinusitis in the first instance, but sees patients again a few days later if symptoms have not improved. Another colleague comments that he usually prescribes amoxycillin for patients with sinusitis. At the end of the day you decide to examine the evidence regarding the use of antibiotics for the treatment of acute sinusitis. You want to know if antibiotics have been shown to improve the clinical course of acute sinusitis and if newer generation antibiotics have been shown to be superior to agents such as amoxycillin.

You decide to search MEDLINE, but where do you begin? You decide to search with a search filter (see 3.2.9) in ‘PubMed Clinical Queries’ using the terms ‘acute’ AND ‘sinusitis’ AND ‘antibiotics’. With a ‘sensitive’ search ~100 citations result, and with a ‘specific’ search there are ~170 citations.

You look at the titles and relevant abstracts and notice that there have been quite a few randomised studies that report different results. Of note, some of the relatively early studies reported a benefit from antibiotics compared with placebo but others did not. You also notice several trials that have compared amoxycillin or penicillin with other agents.

You do not have time to retrieve all the studies and examine them in detail, and you wonder what to make of the fact that different studies have reported disparate findings. A review article might be helpful.

In the following section we will discuss the role of *review articles* in evidence-based practice. In particular the advantages and limitations of systematic reviews will be outlined. Some of the detail...
below is included for completeness and the sections in italics may be skipped without loss of continuity.

**Systematic reviews and meta-analyses**

*According to the Cochrane Collaboration Handbook, a systematic review is an overview of primary studies that contains an explicit statement of objectives, materials and methods and has been conducted according to explicit and reproducible methodology*\(^6\).

The two key features of systematic reviews are that the methods used are transparent and that studies are included in the review on the basis of objective criteria (independent of the results of the study). A good systematic review will be based on a well-structured focused clinical question. The validity of information included in the review is usually assessed by evaluating the methodological quality of individual trials according to specific criteria. Systematic reviews summarise the available literature using qualitative methods but may also include numerical quantification, in which case it is called a meta-analysis\(^7\).

Meta-analysis is defined as a statistical synthesis of numerical results of several trials that all examine the same question\(^8\). Thus a meta-analysis is not synonymous with a systematic review but is frequently used to give a statistically valid summary of the data.

**Why do we need systematic reviews and meta-analyses?**

Clinicians often use textbooks and narrative review articles to keep up to date with the medical literature. However these articles may not always provide an unbiased and comprehensive summary of the publications\(^9\). Such articles may include a mixture of personal opinion and evidence or be based on a biased or incomplete selection of primary studies. While traditional narrative reviews can be useful in certain circumstances, for many clinical questions, a systematic review is the best source of evidence-based information.

There are many features of systematic reviews that make them useful tools for summarising and evaluating research evidence\(^10\). The most fundamental advantage of a systematic review is that bias in identifying, selecting and summarising the evidence is minimised. Because the methods used are both explicit and systematic the conclusions should be reliable and closer to the ‘truth’ than previous standard methods of reviewing the literature\(^10\).

For those reviews where meta-analysis can be performed, the quantitative synthesis of data from different studies can provide additional information. The advantage of meta-analysis is that for average estimates there is a gain in statistical power. This is particularly important where individual studies may be too small to detect modest but clinically important effects (insufficient statistical power). Combining the results of individual trials provides a greater ‘sample size’ and therefore power increases. For the same reasons the precision of the estimate is likely to be greater\(^11\).

Another important feature of systematic reviews is that the generalisability and consistency of the results can be explored\(^10\). If a review demonstrates similar effects across a wide range of settings or study designs we can be reassured about the robustness of the findings and their generalisability to other settings. Where inconsistencies are noted between different studies the sources of variation can be explored.
Steps involved in undertaking a systematic review

1. Formulation of a structured question specifying the qualities of primary studies including populations, interventions and outcomes to be considered for the review.
2. Extensive search of the literature including non-English language or unpublished sources where appropriate. The search strategy should be documented.
3. Selection of studies for inclusion in the review using at least 2 independent reviewers and a transparent process for resolving disagreements.
4. Assessment of the methodological quality of included studies using pre-specified criteria and 2 independent reviewers.
5. Abstraction of relevant data
6. Data synthesis and discussion of findings.

* Data on the level of agreement required between reviewers for steps 3 to 5 should be provided.

Assessing agreement between reviewers

The level of agreement between different reviewers assessing studies for inclusion in a review or evaluating the quality of included studies should be reported in reports of systematic reviews. Kappa statistics are often used to present data on inter-observer agreement. A Kappa statistic is a measure of proportional agreement that has been corrected for chance (also sometimes called Cohen’s κ statistic). A detailed understanding is beyond the scope of this course. A κ (kappa) of 1 represents perfect agreement. A value of 0 indicates no agreement better than chance and negative values indicate worse than chance agreement. There are no absolute criteria used to interpret kappa statistics but the following may be used as a guide^12:

<table>
<thead>
<tr>
<th>Value of κ</th>
<th>Strength of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.2</td>
<td>Poor</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41 - 0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61 - 0.80</td>
<td>Good</td>
</tr>
<tr>
<td>0.81 - 1.00</td>
<td>Very good</td>
</tr>
</tbody>
</table>

Sources of bias in systematic reviews

Most systematic reviews and meta-analyses are retrospective studies and may be influenced by several sources of bias. Critical appraisal guides have been developed to help readers evaluate the validity of the findings of systematic reviews. Reviewers can introduce bias by using inappropriate study selection criteria, inadequately assessing the quality of included studies, selectively presenting outcomes or by drawing conclusions that are not consistent with the study results. Missing trials or missing data are other potential sources of bias.

Bias related to inclusion criteria

Bias can occur if study results are used to select studies for inclusion. The criteria should relate closely to a focused clinical question.

Bias related to identification of studies

Inadequate literature search

Literature searches limited to databases such as MEDLINE are likely to miss relevant published reports. For randomised clinical trials, a search strategy that is limited to MEDLINE has been shown to identify only 51% of known published studies. Some studies are not indexed on MEDLINE and even those that are indexed on MEDLINE may be missed by MEDLINE searches due to the limitations of the indexing process. Even with a highly sensitive search strategy using
MEDLINE, only approximately three-quarters of all published studies indexed on MEDLINE are likely to be identified. The authors of review articles should describe the methods used to identify primary studies. The following sources of information should be considered:

- Electronic databases including MEDLINE, EMBASE, the Cochrane Controlled Trial Register, Current Contents or other relevant health information databases.
- Searching the bibliographies of previous review articles.
- Contact with authors of primary studies and experts in the field to identify data missed by the search (published or unpublished data)
- Contact with pharmaceutical companies, where relevant.
- Hand searching of relevant journals or abstracts of international meetings may also be important for some reviews.

**Publication bias**

This type of selection bias has been found to occur commonly in systematic reviews and meta-analyses. Publication bias refers to the fact that studies with positive or significant results are more likely to be submitted, published or published more quickly than studies with ‘negative’ or non-significant results. For example, a review of studies submitted to an Australian ethics committee found clinical trials, with significant results, were 3 times more likely to be published than studies with null results. Interestingly, studies with indefinite outcomes were even less likely to be published than studies with a non-significant study outcome (p > or = 0.10).

There are several methods that reviewers can use to assess the possibility of publication bias. A detailed discussion is beyond the scope of this course. For those interested, references are provided. Nonetheless, in most reviews, reviewers should assess the possibility of publication bias and where relevant perform a sensitivity analysis to assess the potential impact of missing studies on the review findings. An example of publication bias is provided below.

**Publication bias: an example using funnel plots**

Funnel plots are one method that can be used to examine the possibility of publication bias. A funnel plot is a simple scatter plot with the treatment effect (a measure of effect size) on one axis plotted against some measure of study size on the other axis. The precision of the estimate of treatment effect will increase as the sample size of the individual studies increases, giving rise to the appearance of an inverted funnel. The estimates of treatment effect from smaller (and less precise studies) will tend to scatter at the bottom of the graph and the degree of spread becomes less with studies of larger sample size. A pooled estimate of the treatment effect (usually the most precise estimate of treatment effect) will often be placed at the top of the funnel. If bias is not present the plot should resemble a symmetrical inverted funnel.

An example of a funnel plot is provided in Figure 5.1 taken from Midgley et al. This is taken from a meta-analysis of randomised controlled trials examining the effect of reduced dietary sodium intake on blood pressure in both normotensive and hypertensive individuals. The reviewers of this study concluded that dietary sodium restriction for older hypertensive individuals might be considered but the evidence in the normotensive population does not support current recommendations for universal dietary sodium restriction. The reviewers constructed a funnel plot and the results suggest that publication bias may be present, although other reasons for funnel plot asymmetry are possible.
Figure 5.1 Funnel plot to explore publication bias. The vertical axis represents outcome systolic blood pressure minus baseline systolic blood pressure in the high-sodium group subtracted from the low-sodium group; horizontal lines are at zero and at the fixed-effects mean. The upper and lower dotted lines are twice the median variance divided by the effective sample size. Data from all trials were included in this analysis. From Midgley: JAMA 1996; 275:1590-1597.

Citation bias
This type of bias may occur when the reviewers use searches of the bibliographies of other review articles to identify primary studies. In some fields there may be a tendency for authors of narrative reviews to cite positive studies more often than negative ones or vice versa depending on the beliefs of the authors. This may mean that some studies are missed by bibliographic searches.  

Language bias
Basing reviews on studies that are published in English only can introduce bias. Some evidence suggests that studies conducted in non-English speaking countries are more likely to be published in English language journals if the results are ‘positive’ and in local journals if ‘negative’. Such negative studies would be missed by a search limited to articles published in English.

Bias related to selective inclusion or reporting of outcomes
The results of a review may be influenced by whether evidence on all the important outcomes is included in the report. For example if reviewers omit important data about adverse events readers may be left with a biased impression of the relative risks and benefits of the intervention.

Bias related to the inclusion of poor quality primary studies
Systematic reviews should include an assessment of the validity of the primary studies included in the review. Combining the results from primary studies that are poor quality or biased is likely to produce unreliable results. Differences in the methodological quality of studies included in reviews may also explain differences in study results. Some reviewers use quality scales to assess the validity of included studies; others report validity by assessing a set number of individual criteria. There are advantages and limitations to both of these approaches and no absolutely correct method. Readers of review articles should make
judgements about whether the criteria used to assess the validity of included studies are appropriate. For controlled studies the criteria outlined in the section on therapeutics (in this workbook) should be considered.

**Quantitative synthesis of study results (meta-analysis)**

If appropriate, systematic reviews may include a meta-analysis of the results of primary studies included in the review. Meta-analysis involves 2 steps. Firstly an appropriate summary statistic is calculated for each of the individual studies in the review. Secondly these summary values are combined into a weighted average\(^{20,21}\). Often studies are weighted according to their size, so that larger studies receive more weight. Different statistical methods are available for combining different types of data. A detailed description is beyond the scope of this course.

*Broadly speaking there are 2 types of models used for meta-analysis*\(^{21,22}\):

- **Fixed effects models** consider only the within-study variability in the calculation of the common effect. The assumption is that all the studies are studying the same true effect and that variability is due to random error.
- **Random effects models** take into account both between study and within study variability in the calculation of the common effect. This model assumes that the true effect differs among studies and therefore must be represented by a distribution of values instead of single value. The random effects model will usually produce wider confidence intervals. It should be used if there is evidence of statistical heterogeneity between the studies included in the review (see section below).

The results of meta-analysis are often displayed using a Forest plot, an example is provided in Figure 5.2 taken from a review of lung cancer screening interventions\(^{23}\).
Understanding heterogeneity

Heterogeneity refers to the diversity that exists between different studies included in a review. Differences between studies might include different patient populations, variations in interventions (eg different pharmaceutical preparations or classes of medications), or variations in methodological quality. Statistical heterogeneity refers to the incompatibility in the quantitative results of different studies. The statistical question that arises is whether there is greater variation between the results of the trials than is compatible with chance. The Q statistic, interpreted using a chi-squared distribution, is generally used as a test of homogeneity. As stated, this is a test of homogeneity and therefore a p value of less than 0.05 indicates that there is significant statistical heterogeneity between the results of different studies. This suggests that a single common treatment effect is unlikely. Tests of homogeneity have been shown to have low statistical power and may fail to detect heterogeneity, therefore a more liberal significance level may be used to such as 0.1 rather than 0.05. A failure to demonstrate statistical heterogeneity does not necessarily rule out clinically important heterogeneity.

The clinical importance/significance of statistical heterogeneity can be examined by examining the size of the difference in outcome between the 2 most disparate studies. Furthermore examining the boundaries of the confidence intervals of the different studies can be useful. If the difference between the boundaries (for example the upper boundary of one study and the lower boundary of
another study) of the confidence intervals for disparate studies is greater than 5% this suggests clinically significant heterogeneity\textsuperscript{25}. Readers of review articles can quickly estimate whether there might be clinically important heterogeneity between different studies included in the review by examining the Forest plot and determining whether the confidence intervals from individual studies tend to overlap or not.

If significant heterogeneity is present then reviewers should first consider whether it is appropriate to combine the results of different studies. Such combined estimates may not make clinical or biological sense. If a meta-analysis is performed with heterogeneous (statistically significant heterogeneity) data, the random effects model should be used. The reasons for heterogeneity should be explored and explained where possible\textsuperscript{22}.

**Sensitivity analyses**

Sensitivity analyses can be performed to evaluate the robustness of the results of the meta-analysis. Sensitivity analyses examine how the results vary under different assumptions about how the review or meta-analysis is conducted\textsuperscript{26}. For example sensitivity analyses may include the following:

- Changing inclusion criteria.
- Excluding poor quality studies.
- Re-analysing the data using alternative statistical approaches, for example comparing fixed effects with random effects models.
- Re-analysing data to include imputed values for missing data (for example if publication bias is suspected).

**Subgroup analyses**

If data from different subgroups of patients are available from each study, subgroup analyses might be possible. This may provide further insights into reasons for heterogeneity. In most cases, meta-analysis is retrospective and subgroup analyses are therefore post-hoc and need to be interpreted cautiously\textsuperscript{22}. The following criteria have been suggested for determining whether the results of subgroup analysis are likely to be reliable\textsuperscript{27}:

- Was the hypothesis made prior to undertaking the study/review?
- Was it one of only a few hypotheses tested?
- Are the findings consistent across studies?
- Are differences biologically plausible?
- Is there a big difference in treatment effect between subgroups?
- Is the difference in treatment effect highly statistically significant?

Findings from subgroup analyses that satisfy these criteria may carry more weight. In general, if these criteria are not met, subgroup analyses should be viewed as hypothesis generating. Subgroup analyses derived from comparisons between studies (comparing patients in one study with patients in another) should be viewed sceptically\textsuperscript{27}.

*Meta-regression can be used to examine gradients in treatment effects. In this way differences in outcomes can be further examined in relation to differences in the attributes of studies. Factors such as age or duration of follow up can be explored. Where graded associations exist, statistical tests for trend can be used. (A detailed description of the statistical methods is beyond the scope of this course). For those interested in further reading a reference is provided\textsuperscript{28}.*
Outcome measures for systematic reviews

For reviews of controlled or randomised-controlled studies examining treatments, dichotomous outcomes (yes/no) will often be presented with measures such as relative risk or absolute risk reduction. For such outcomes it is useful if the reviewers also provide summary measures, such as NNT. As discussed in the section on therapy, presenting results as NNT helps readers understand and communicate the practical importance of the results. It should be remembered, however, that although NNT is a clinically useful estimate of the average treatment effect in patients at the average risk in the included trials, it may not be directly relevant to an individual patient. In calculating the NNT it may help to use the control event rates from individual studies included in the review or from cohort studies that have included patients similar to those of interest to the reader.

For some types of continuous data such as functional status using different instruments, the results of each study may be summarised using the effect size. This is a way of determining the average effect of the intervention. The effect size is the difference in outcomes between the intervention and control groups divided by the standard deviation. In this way the results of each study are summarised in terms of the number of standard deviations of difference between the intervention and control groups. The combined effect of all studies can be calculated from a weighted average of the effect sizes. Ideally such results should then be presented in a way that can be understood by readers, perhaps by translating the summary effect size back into natural units.

Issues related to the interpretation of the results of systematic reviews

Some reviews may find inconclusive evidence. In these circumstances it is important not to confuse ‘no evidence of effect’ with ‘evidence of no effect’. Often the findings of reviews will neither confirm nor refute the presence of treatment benefit, particularly if there are a small number of underpowered studies. Evidence of no effect however, can only be concluded if the confidence intervals from a pooled analysis clearly exclude a clinically important difference between intervention and control groups (particularly if the studies are high quality and homogeneous). For a discussion of confidence intervals see 4.2. Even reviews with inconclusive results are helpful. They provide a summary of the available literature and often highlight how future research should be focused.

Although systematic reviews of randomised controlled studies are usually considered the highest level of evidence in the evidence hierarchy, some reviews provide higher-level evidence than others. For example smaller meta-analyses (those with a limited number of small trials or less than 200 outcome events) are likely to be less reliable than larger meta-analyses. Such small meta-analyses are still useful for summarizing the available evidence and generating hypotheses for future research. The results of a meta-analysis of randomised controlled trials in which the treatment effects from individual studies are consistent (homogeneous), and the lower limit of the confidence interval for the treatment effect exceeds the minimally clinically important difference, carry greater weight than those meta-analyses with significant heterogeneity (wide disparity between individual studies) or in those where the treatment effect overlaps the minimally clinically important benefit.

Cochrane Reviews

The Cochrane Collaboration is an international organisation that aims to produce evidence based systematic reviews that are accessible, clinically useful, quality controlled and periodically updated. Cochrane systematic reviews are published in the Cochrane Library. The Cochrane Library can be accessed in Australia from http://www.nicsl.com.au. In addition to original reviews, the Cochrane Library contains a database of abstracts of reviews (DARE) from other sources that have been quality assessed and summarised. An important feature of Cochrane Reviews is that
they are periodically updated and may be revised to include new research findings or in response to valid criticisms. Each new addition of the Cochrane Library (updated quarterly) contains previously published Cochrane reviews in addition to updates and new protocols.

**Critical appraisal of systematic reviews**

Several critical appraisal guides have been published\(^8,26,27\). We have based our systematic review critical appraisal worksheet (at the end of this chapter) on the Users’ guides’ series\(^27\) with some modifications taken from the Centre for Evidence-Based Medicine, Mount Sinai Hospital website: [http://www.cebm.utoronto.ca/teach/materials/therapy.htm](http://www.cebm.utoronto.ca/teach/materials/therapy.htm)

The following criteria are suggested for appraising review articles\(^27\):

**Are the results of the study valid?**

**Primary guides:**
- Did the overview address a focused clinical question?
- Were the criteria used to select articles for inclusion appropriate?

**Secondary guides:**
- Is it unlikely that important, relevant studies were missed?
- Was the validity of the included studies appraised?
- Were assessments of studies reproducible?
- Were the results similar from study to study?

**What are the results?**
- What are the overall results?
- How precise were the results?

**Will the results help me in caring for my patients?**
- Can the results be applied to my patient care?
- Were all clinically important outcomes considered?
- Are the benefits worth the harms and costs?

### 5.3 Worked examples

Recall the scenario discussed in the introduction to this section. To help you decide if antibiotics are indicated for the management of your patient (and if so whether newer generation antibiotics are more effective), you search for a systematic review.

You decide to search the Cochrane Database of Abstracts of Reviews of Effectiveness and the Cochrane Database of Systematic Reviews. You use the term acute sinusitis and find 7 articles; the following 3 are the most relevant:


You examine the article by de Ferranti et al in more detail.

Abstract

Objectives: To examine whether antibiotics are indicated in treating uncomplicated acute sinusitis and, if so, whether newer and more expensive antibiotics with broad spectra of antimicrobial activity are more effective than amoxycillin or folate inhibitors.

Design: Meta-analysis of randomised trials.

Setting: Outpatient clinics.

Subjects: 2717 patients with acute sinusitis or acute exacerbation of chronic sinusitis from 27 trials.

Interventions: Any antibiotic versus placebo; amoxycillin or folate inhibitors versus newer, more expensive antibiotics.

Main outcome measurements: Clinical failures and cures.

Results: Compared with placebo, antibiotics decreased the incidence of clinical failures by half (risk ratio 0.54 (95% confidence interval 0.37 to 0.79)). Risk of clinical failure among 1553 randomised patients was not meaningfully decreased with more expensive antibiotics as compared with amoxycillin (risk ratio 0.86 (0.62 to 1.19); risk difference 0.9 fewer failures per 100 patients (1.4 more failures to 3.1 fewer failures per 100 patients)). The results were similar for other antibiotics versus folate inhibitors (risk ratio 1.01 (0.52 to 1.97), but data were sparse (n=410) and of low quality.

Conclusions: Amoxycillin and folate inhibitors are essentially as effective as more expensive antibiotics for the initial treatment of uncomplicated acute sinusitis. Small differences in efficacy may exist, but are unlikely to be clinically important.

To obtain the full text of this article go to the BMJ website at, http://www.bmj.com. Click on Search/Archive, then under Word(s) in Title or Abstract type: ‘acute sinusitis, meta-analysis’ and click Search. This article will appear in the search results along with a few other articles.

A critical appraisal of this article is outlined on the following page.
# Worked example: Systematic review (of therapy) worksheet

## Citation


## Are the results of this systematic review valid?

<table>
<thead>
<tr>
<th>Did the systematic review address a focused clinical question?</th>
<th>Although there is no single statement outlining all the important elements of the question for this review in the introduction, the clinical question is clear from the abstract and text and includes the following components:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong> – Patients with acute uncomplicated sinusitis, or acute exacerbation of chronic sinusitis treated in the outpatient setting</td>
<td><strong>Intervention</strong> – Antibiotics</td>
</tr>
<tr>
<td><strong>Comparison</strong> – compared with different classes of antibiotics or placebo</td>
<td><strong>Outcomes</strong> – Clinical cure, improvement, and treatment failure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Were the criteria used to select articles for inclusion appropriate?</th>
<th>Yes, the following inclusion criteria applied:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What were the inclusion criteria?</strong></td>
<td>1) trial compared amoxycillin or a folate inhibitor with another antibiotic, generally one with broad spectrum of activity, including cephalosporins, penicillins with beta-lactamase inhibitors, tetracyclines, quinolines, and macrolides</td>
</tr>
<tr>
<td></td>
<td>2) studies comparing antibiotic treatment with a placebo control group were also included</td>
</tr>
<tr>
<td></td>
<td>3) Only studies with random assignment to treatment arms were included</td>
</tr>
<tr>
<td></td>
<td>4) Trials evaluating acute sinusitis or an exacerbation of chronic sinusitis</td>
</tr>
</tbody>
</table>

Exclusions:
- studies that compared doses of non-antimicrobial drugs
- trials of subacute or chronic sinusitis (mean duration of symptoms > 3 weeks)

These criteria are appropriate to the clinical question being addressed in the review.

| Is it likely that important, relevant studies were missed? | The search included medical databases and hand searching of conference proceedings. Bibliographies were also searched. Searches of databases such as EMBASE were not undertaken and therefore the reviewers could have missed studies published outside |

---

School of Population Health

74
There was no mention of attempts to find possible missing published or unpublished data for example by contacting experts in the field or pharmaceutical companies.

Publication bias may be source of bias for this review. In the meta-analysis analysing placebo controlled trials there are several small ‘positive’ studies. It is possible that there are also some smaller ‘negative’ studies that were conducted but not published. The authors have not attempted to determine whether publication bias might be a factor in this review or determine the potential impact on the results by performing a sensitivity analysis and it is therefore difficult to determine whether this might be an important source of bias in this review.

The authors discuss the possibility of publication bias in relation to the trials comparing different antibiotic classes but not in relation to the placebo controlled studies (see discussion section).

<table>
<thead>
<tr>
<th>Was the validity of the included studies appraised?</th>
<th>Yes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>If validity was assessed, explain how.</td>
<td>The reviewers used 2 methods to assess quality, a subject specific set of criteria and the Jadad scale for assessing the methodological quality of clinical trials.</td>
</tr>
<tr>
<td></td>
<td>The subject specific criteria were:文化的 blinded versus unblinded design文化的 criteria for diagnosis of sinusitis文化的 clinical outcomes文化的 loss of subjects to follow up文化的 use of decongestants</td>
</tr>
<tr>
<td></td>
<td>The Jadad quality scale is a validated scale that assesses the internal validity of randomised controlled trials. The details are provided at the end of this section.</td>
</tr>
<tr>
<td>Were assessments of studies reproducible?</td>
<td>The authors state that data extraction (of results) was performed by 2 reviewers independently. They do not say whether the selection of studies for inclusion in the review or quality assessments was undertaken by 2 independent reviewers or whether agreement between reviewers was assessed.</td>
</tr>
<tr>
<td></td>
<td>Without a comment about reproducibility, it is difficult to assess the reliability of the reviewers’ data extraction and the quality assessment process.</td>
</tr>
</tbody>
</table>
Were the results similar from study to study? (homogeneous)

Did the authors assess the possibility of statistical heterogeneity?

For the placebo controlled studies the authors comment that there was no statistically significant heterogeneity. They did note, however that there was a suggestion that one trial which included patients simply on the basis of sinusitis-like symptoms without further diagnostic information, had the highest rate of cure in the placebo group (85% at 10 days) and showed no benefit from antibiotics. Those trials with more tightly defined patient populations and lower spontaneous improvement rates showed a clear benefit from antibiotics.

For the trials comparing different classes of antibiotics in different treatment arms, statistical heterogeneity was noted (p=0.09 for clinical cure) for the studies comparing folate inhibitors with other antibiotics. None was noted for studies comparing amoxycillin with other antibiotics.

Although some limitations to the validity are suggested above, these are unlikely to be substantial enough to invalidate the results of the review, therefore at this point it would be appropriate to examine the results. If you felt that the methods of the review were substantially flawed and could introduce major bias, you might choose to discard the results altogether.

* Of the Journals covered by MEDLINE, 52% are published in the USA. EMBASE is another large database of medical and scientific references. The range of journals covered differs to that of MEDLINE with only 33% of journals covered being published in North America.

** Jadad quality scale assigns a score to studies of between 0 and 5, where 5 is the maximum score representing the highest quality. The following questions are used:

(a) Was the study described as randomised? (1=Yes 0=No).
(b) Was the study described as being double blind? (1=Yes 0=No).
(c) Was there a description of withdrawals and dropouts? (1=Yes 0=No).
(d) Was the method of randomisation well described and appropriate? (1=Yes 0=No).
(e) Was the method of double blinding well described and appropriate? (1=Yes 0=No).
(f) Deduct one point if methods for randomisation or blinding were inappropriate.

What are the results?

What are the overall results?

Compared with placebo, antibiotics decreased the incidence of clinical failures by half, the risk ratio was 0.54 (95% confidence interval 0.37 to 0.79). For the point estimate the NNT to prevent clinical failures was calculated (see table below). 7 patients need to be treated with any antibiotic to ‘cure’ 1 patient (prevent treatment failure within 48 hours of the completion of treatment). As the reviewers point out the symptoms were noted to improve or disappear in 69% of patients without any antibiotic treatment. The risk of clinical failure did not differ significantly in the analysis of studies comparing amoxycillin with other antibiotics or in those studies comparing folate inhibitors with other antibiotic classes.
How precise were the results?

For the comparison between placebo and any antibiotic the 95% confidence intervals are between 0.37 and 0.79. This suggests that the true risk ratio may be as large as 0.37 or as small as 0.79 at the 95% level of confidence.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experimental event rate (Y)</th>
<th>Relative risk reduction $\text{RRR} = (X - Y)/X$</th>
<th>Absolute risk reduction $\text{ARR} = X - Y$</th>
<th>Number needed to treat (NNT = $1/\text{ARR}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical failure (antibiotics versus placebo)</td>
<td>Placebo 0.31 (31%)</td>
<td>Antibiotic 0.17 (17%)</td>
<td>$\text{RRR} = 1 - \text{RR}$ $\text{RRR} = 0.46$ 46%</td>
<td>0.14 (14%)</td>
<td>7</td>
</tr>
</tbody>
</table>

**Will the results help me in caring for my patient?**

| Can these results be applied to my patient care? | The mean ages of patients ranged from 25 to 44 years. Apart from this there is little demographic data about the primary studies included. An important consideration is the diagnostic criteria used to establish the diagnosis. Many of the studies included in the review used diagnostic criteria such as radiography and culture of sinus aspirations to establish the diagnosis. As pointed out by reviewers however, the one study that had inclusion criteria used by most family doctors (sinusitis-like symptoms) showed no benefit from antibiotics (after 10 days 85% of both the control and intervention groups had recovered). However a benefit from antibiotics in this setting is not excluded from the results of this individual study alone (note the confidence intervals for the study by Stalman et al 1997 in figure 1). Further studies would be needed before drawing firm conclusions about the generalisability of the findings to the general practice setting, where diagnosis is based on clinical criteria alone. At a minimum however, the higher rates of spontaneous improvement in this group of patients suggest that the NNT would be greater in the general practice setting. Of note recent guidelines do not recommend the use of routine radiography for the diagnosis of acute sinusitis. Clinical criteria are used to define suspected cases of acute bacterial sinusitis. |

| Is your patient so different from those in the study that its results cannot apply? | |

School of Population Health
Another factor that you may need to take into account is the local prevalence of antibiotic resistance.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the treatment feasible in your setting?</td>
<td>Oral antibiotic treatment is clearly feasible in the general practice or outpatient setting. The ability to perform procedures such as sinus aspiration to confirm diagnosis may not feasible.</td>
</tr>
<tr>
<td>Were all clinically important outcomes considered?</td>
<td>Clinical cure or failure rates seems a reasonable outcome measure, but as pointed out by the reviewers, the criteria used to establish this differed between studies and were not always well specified, making interpretation of this finding more difficult for practicing clinicians. Of note, amongst all the patients treated in the studies included in the review, there were no serious complications reported such as meningitis or brain abscess. Outcomes that would be important to patients but do not appear to have been covered in the primary studies might include quality of life, duration and intensity of symptoms such as pain and need for time off work.</td>
</tr>
<tr>
<td>Are the likely treatment benefits worth the potential harms and costs?</td>
<td>In discussing the need for treatment with your patient you might say, “the majority of patients with this conditions get better without antibiotic treatment. With antibiotics, however, treatment failures are less likely, but 7 patients need to be treated with antibiotics to prevent 1 treatment failure (48 hours after the treatment has been completed)”. Some patients will prefer antibiotics in the first instance while others may not. The benefits of treatment need to be weighed against the potential side effects of the medication. Side effects were not discussed in detail or clearly considered in the outcomes described. Costs were not considered but newer antibiotics are usually more expensive and there are concerns that their widespread use may increase rates of bacterial resistance. The results of this review suggest that there are no clear benefits from using newer antibiotics at this stage.</td>
</tr>
<tr>
<td><strong>When considering this question you should take into account your patient’s values and preferences.</strong></td>
<td><strong>In discussing the need for treatment with your patient you might say, “the majority of patients with this conditions get better without antibiotic treatment. With antibiotics, however, treatment failures are less likely, but 7 patients need to be treated with antibiotics to prevent 1 treatment failure (48 hours after the treatment has been completed)”. Some patients will prefer antibiotics in the first instance while others may not. The benefits of treatment need to be weighed against the potential side effects of the medication. Side effects were not discussed in detail or clearly considered in the outcomes described. Costs were not considered but newer antibiotics are usually more expensive and there are concerns that their widespread use may increase rates of bacterial resistance. The results of this review suggest that there are no clear benefits from using newer antibiotics at this stage.</strong></td>
</tr>
</tbody>
</table>
5.4 Self-complete exercises

Clinical scenario

You are a general practitioner. A longstanding patient (Mr CE), who has recently had a transient ischaemic attack (TIA), calls you to say his neurologist and neurosurgeon have recommended that he have a carotid endarterectomy but he would like to come and discuss it with you before agreeing to go ahead with the operation.

Mr CE is a 72 and an ex-smoker. He has had hypertension for 10 years, which is currently well controlled with amlodipine 10 mg daily. His only other regular medication is aspirin 150mg daily. He has no history of heart disease or peripheral vascular disease. In the past 6 months he has had 2 TIAs, the first occurred when he was not taking aspirin and the second while taking aspirin. On both occasions he presented with right facial and arm weakness associated with slurred speech, which resolved within 24 hours.

He was referred to a tertiary hospital for investigation and management. CT brain was normal. ECG and echocardiography were normal. Carotid angiography was performed after a carotid ultrasound suggested a stenosis of the left internal carotid artery. The angiogram confirmed the presence of 67% stenosis of the left carotid artery. There was no substantial abnormality noted in the right carotid arterial circulation.

After talking to Mr CE on the phone you arrange to see him the next day. You decide to examine the evidence regarding carotid endarterectomy for symptomatic carotid stenosis and perform a MEDLINE search, hoping to locate a relevant systematic review. You search MEDLINE using ‘PubMed Clinical Queries’, with the term ‘carotid endarterectomy’ in ‘Find Systematic Reviews’. There are ~230 citations in ‘All’ and 115 of these are within ‘Review’. You search the titles and abstracts and find the following systematic review: Cinà CS, Clase CM, Haynes BR. Refining the indications for carotid endarterectomy in patients with symptomatic carotid stenosis: A systematic review. Journal of Vascular Surgery 1999; 30:606-617.

Critically appraise the article and decide how the results should be applied to your patient. Use the guide below, which is based on the critical appraisal worksheet provided at the end of the section. Some additional questions have been added (in italics) to help guide you through the process.

Systematic review (of therapy) worksheet

Citation

Are the results of this systematic review valid?

<table>
<thead>
<tr>
<th>Did the systematic review address a focused clinical question?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Although the research question is not outlined in a single statement in the introduction the inclusions, exclusions and outcomes of interest are described. Can you describe the following components of the question?</td>
</tr>
<tr>
<td>Population – Intervention – Comparison – Outcome –</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Were the criteria used to select articles for inclusion appropriate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What were the inclusion criteria?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is it likely that important, relevant studies were missed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>If you think that important studies were not likely to be missed explain why.</td>
</tr>
<tr>
<td>Question</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Was the validity of the included studies appraised?</td>
</tr>
<tr>
<td><em>If validity was assessed, explain how.</em></td>
</tr>
<tr>
<td>Were assessments of studies reproducible?</td>
</tr>
<tr>
<td><em>How do you interpret the kappa statistics presented in this article?</em></td>
</tr>
<tr>
<td>Were the results similar from study to study? (homogeneous)</td>
</tr>
<tr>
<td><em>Did the authors assess the possibility of statistical heterogeneity?</em></td>
</tr>
</tbody>
</table>
**What are the results?**

**What are the overall results?**
Some of the main results are summarised in the table below. (Taken from table IV, Cinà et al).

Can you calculate the NNT using the data provided in the table?

**How precise were the results?**

<table>
<thead>
<tr>
<th>Stenosis</th>
<th>Relative risk reduction or increase (%)</th>
<th>Absolute risk reduction or increase (%)</th>
<th>Number needed to treat (NNT = 1/ARR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70% to 99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% to 69%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 49%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Number Needed to Harm

**Will the results help me in caring for my patient?**

<table>
<thead>
<tr>
<th>Can these results by applied to my patient care?</th>
<th>Is your patient so different from those in the study that its results cannot apply?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| What are the important factors that you need to consider for this patient? |
| If subgroup analyses were presented remember the guidelines outlined for assessing the reliability of these. |
|---------------------------------------------------------------------------|---------------------------------------------------------------------------------|</p>
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the treatment feasible in your setting?</td>
<td></td>
</tr>
<tr>
<td>Were all clinically important outcomes considered?</td>
<td></td>
</tr>
<tr>
<td>Are the likely treatment benefits worth the potential harms and costs?</td>
<td></td>
</tr>
<tr>
<td><em>When considering this question you should take into account your patient’s values and preferences.</em></td>
<td></td>
</tr>
</tbody>
</table>
5.5 Answers to self-complete exercises

Citation

Are the results of this systematic review valid?

<table>
<thead>
<tr>
<th>Did the systematic review address a focused clinical question?</th>
<th>Yes. Although the research question is not outlined in a single statement in the introduction, the inclusions, exclusions and outcomes of interest are described. The question has the following components:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population – patients with ipsilateral symptomatic carotid stenosis</td>
<td></td>
</tr>
<tr>
<td>Intervention – carotid endarterectomy</td>
<td></td>
</tr>
<tr>
<td>Comparison – compared with medical management (control group)</td>
<td></td>
</tr>
<tr>
<td>Outcome – Death or major disability from stroke.</td>
<td></td>
</tr>
</tbody>
</table>

| Were the criteria used to select articles for inclusion appropriate? | Yes. Inclusion criteria were based on study design (only randomised controlled studies were included), and the intervention, comparison and population of interest. It seems reasonable to exclude studies with asymptomatic carotid stenosis since such patients differ on prognostic grounds. It also seems reasonable to exclude studies that included unusual surgical techniques. |

| Is it likely that important, relevant studies were missed? | Unlikely. The search of medical databases was well described and comprehensive. There was no attempt to contact experts in the field or identify unpublished data. However one of the reviewers has been involved with the primary studies and therefore has some expertise in the field. It is difficult to examine for possible publication bias when the number of studies is small. The 3 studies identified for the review were all reasonably large and publication bias is more likely when there are number of smaller positive studies. |

| Was the validity of the included studies appraised? | Yes. The validity of included studies was appraised by 2 reviewers independently. The results were summarised |
in table III of the article. Overall the quality of studies appears high. As the reviewers point out, it is difficult to perform a double blind study with a surgical intervention. Although outcomes were assessed by independent reviewers, clinical data for such assessments was derived from unblinded clinical evaluations. This may lead to detection bias.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were assessments of studies reproducible?</td>
<td>Yes.</td>
</tr>
<tr>
<td>Studies were selected for inclusion by 2 reviewers. Inter-observer reliability was excellent (K = 0.95). Methodological quality was also assessed by 2 reviewers and agreement between reviewers was good (K= 0.75). Data extraction was performed by 2 reviewers independently using a standard form and the error rate was low.</td>
<td></td>
</tr>
<tr>
<td>Were the results similar from study to study? (homogeneous)</td>
<td>Yes.</td>
</tr>
<tr>
<td>There was no significant statistical or clinical heterogeneity between studies included in the review. As the authors point out, because only 2 studies are included in the actual meta-analysis the power to detect heterogeneity is low. The point estimates for each of the studies are similar and the confidence intervals overlap, making heterogeneity unlikely. (Refer to figure 1 Cinà et al)</td>
<td></td>
</tr>
</tbody>
</table>

**What are the results?**

**What are the overall results?**
The main results are summarised in the table below. (taken from table IV, Cinà et al).

**How precise were the results?**
For patients with a stenosis of 70% to 99% the results are precise with confidence intervals ranging from 10 to 31 for the NNT. This suggests that there is a 95% chance that the true NNT lies between 10 and 31. The confidence intervals are wider for the group of patients with a stenosis of 50 to 69%, the true NNT may lie between 11 and 125, at the 95% level of confidence.
### Will the results help me in caring for my patient?

<table>
<thead>
<tr>
<th>Stenosis</th>
<th>Relative risk reduction or increase (%)</th>
<th>Absolute risk reduction or increase (%)</th>
<th>Number needed to treat (NNT = 1/ARR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70% to 99%</td>
<td>48 (27 to 63)</td>
<td>6.7 (3.2 to 10)</td>
<td>NNT 15 (10 to 31)</td>
</tr>
<tr>
<td>50% to 69%</td>
<td>27 (5 to 44)</td>
<td>4.7 (0.8 to 8.7)</td>
<td>NNT 21 (11 to 125)</td>
</tr>
<tr>
<td>&lt; 49%</td>
<td>20 (0 to 44)</td>
<td>2.2 (0 to 4.4)</td>
<td>*NNH 45 (22 to ∞)</td>
</tr>
</tbody>
</table>

* Number Needed to Harm

**Will the results help me in caring for my patient?**

**Can these results by applied to my patient care?**

No.

Firstly, Mr CE appears to be similar to the types of patients included in the individual studies in this review in terms of demographic and clinical features (refer to table II in Cinà et al). Secondly, Mr CE has a stenosis of 67% and therefore fits into the group of patients with a carotid stenosis of 50% to 69%. (Note that these groups are based on stratified randomisation and not subgroup analysis and are therefore likely to be reliable.) In order to be able to extrapolate the findings of this review to your patient, you need to be sure that the methods used to assess the degree of carotid stenosis are standardized and comparable with those used by the primary studies in the review.

The results of this study suggest that the benefits of carotid endarterectomy are less, although still important, for patients with a stenosis of 50% to 69%. As pointed out by the reviewers of this article, for patients with this level of stenosis, consideration should be given to other variables in making decisions about surgery. For Mr CE, there are several factors that suggest he may derive an increased benefit from surgery compared with other patients in this category. These factors include male gender, older age, and the degree of stenosis. Within the categories of stenosis (50% to 69% and 70% to 99%) a gradient effect of benefit was observed with increasing degree of stenosis. This suggests that a patient with a stenosis of 67% may derive a greater benefit (all other things being equal) than a patient with a stenosis of 50%. Male patients and older patients also appeared to derive greater benefit. Gender appears to be an important determinant of outcome, this finding is consistent across the studies, the size of the effect is large and biologically plausible.
* Note that the subgroup analyses discussed in this article are a qualitative summary of the findings of the individual studies and not based on meta-analysis with separate subgroup analyses.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the treatment feasible in your setting?</td>
<td>Yes.</td>
</tr>
<tr>
<td>If Mr CE goes ahead with the surgery you are aware that he will be operated on in a particular tertiary institution where the combined morbidity and mortality of the procedure is known to be less than 6%. This is an important consideration. If the risk of post-operative death or stroke or other morbidity is greater than that reported in the individual studies then the benefits of surgery will be reduced.</td>
<td></td>
</tr>
<tr>
<td>Were all clinically important outcomes considered?</td>
<td>Yes.</td>
</tr>
<tr>
<td>Both death and functional outcomes were considered. Both these outcomes are clearly important to patients. Quality of life in addition to functional outcomes might be of interest but was not reported.</td>
<td></td>
</tr>
<tr>
<td>Are the likely treatment benefits worth the potential harms and costs?</td>
<td>The data suggests that for Mr CE the benefits are likely to outweigh the harms. Clearly his preferences and values need to be taken into account. The results could be explained to Mr CE. Note that for patients with a stenosis of 50% to 69% 21 patients need to be treated to prevent 1 death or disabling stroke, (over 5 years of follow up) although this number may be less for Mr CE given the factors discussed above. It is important to convey to Mr CE that although in the long term there is certainly a net benefit, in the immediate post operative period (30 days) there is an approximately 3% risk of death or major disabling stroke.</td>
</tr>
<tr>
<td>When considering this question you should take into account your patient’s values and preferences.</td>
<td></td>
</tr>
<tr>
<td>How would you rank the importance of immediate (small) risk of an adverse outcome versus the delayed (moderate) benefit in the longer term?</td>
<td></td>
</tr>
<tr>
<td>The value your patient places on these different outcomes clearly needs to be taken into account and the way that you convey information about the risk and benefits of the surgery may influence the informed decision making process. (See 8.1)</td>
<td></td>
</tr>
</tbody>
</table>

5.6 Recommended reading


5.7 Further reading

- Fletcher J. What is heterogeneity and is it important? *BMJ* 2007; 334:94-96


- ‘EPIQ’ (Effective Practice, Informatics & Quality Improvement) is produced by Professor Rod Jackson (University of Auckland). It has very useful materials including notes and tools. [http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm](http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm)

5.8 References


8. Greenhalgh T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997; 315:627-675.


32. The Cochrane Library.  
http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME


5.9 Critical appraisal worksheets (systematic reviews)

5.9.1 Example adapted from Centre for Evidence-Based Medicine, Mount Sinai Hospital
http://www.cebm.utoronto.ca/teach/materials/therapy.htm

Citation

---

**Are the results of this systematic review valid?**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the systematic review address a focused clinical question?</td>
<td></td>
</tr>
<tr>
<td>Were the criteria used to select articles for inclusion appropriate?</td>
<td></td>
</tr>
<tr>
<td>Is it likely that important, relevant studies were missed?</td>
<td></td>
</tr>
<tr>
<td>Was the validity of the included</td>
<td></td>
</tr>
</tbody>
</table>
### studies appraised?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Were assessments of studies reproducible?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Were the results similar from study to study? (homogeneous)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### What are the results?

#### What are the overall results?

- 
- 
- 

### How precise were the results?

- 
- 
-
Your calculations

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experimental event rate (Y)</th>
<th>Relative risk reduction ( RRR = \frac{(X - Y)}{X} )</th>
<th>Absolute risk reduction ( ARR = X - Y )</th>
<th>Number needed to treat (NNT = ( 1/ARR ))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Some reviews may use odds ratios for summary measures of effectiveness. The table below can be used to calculate the NNT from the odds ratio if the expected patient event rate is known.

**Will the results help me in caring for my patient?**

Can these results by applied to my patient care?

Is your patient so different from those in the study that its results cannot apply?

*If subgroup analyses were presented remember the guidelines outlined for assessing the reliability of these.*

Is the treatment feasible in your setting?
Were all clinically important outcomes considered?

Are the likely treatment benefits worth the potential harms and costs?

*Translating odds ratios to NNTs:*
The numbers in the body of the tables are the NNTs for the corresponding odds ratio at that particular patient's expected event rate (PEER).
When the odds ratio (OR) < 1
This table applies when a bad outcome is prevented by therapy.

<table>
<thead>
<tr>
<th>Patient’s expected event rate (PEER)</th>
<th>OR &lt; 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>0.05</td>
<td>2.09a</td>
</tr>
<tr>
<td>0.10</td>
<td>110</td>
</tr>
<tr>
<td>0.20</td>
<td>64</td>
</tr>
<tr>
<td>0.30</td>
<td>46</td>
</tr>
<tr>
<td>0.40</td>
<td>40</td>
</tr>
<tr>
<td>0.50</td>
<td>38</td>
</tr>
<tr>
<td>0.70</td>
<td>44</td>
</tr>
<tr>
<td>0.90</td>
<td>101c</td>
</tr>
</tbody>
</table>

- The relative risk reduction (RRR) here is 10%
- The RRR here is 49%
- For any OR, NNT is lowest when PEER = 0.50
- The RRR here is 9%

When the odds ratio (OR) > 1
This table applies both when a good outcome is increased by therapy and when a side effect is caused by therapy.

<table>
<thead>
<tr>
<th>Patient’s expected event rate (PEER)</th>
<th>OR &gt; 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>0.05</td>
<td>212</td>
</tr>
<tr>
<td>0.10</td>
<td>112</td>
</tr>
<tr>
<td>0.20</td>
<td>64</td>
</tr>
<tr>
<td>0.30</td>
<td>49</td>
</tr>
<tr>
<td>0.40</td>
<td>43</td>
</tr>
<tr>
<td>0.50</td>
<td>42</td>
</tr>
<tr>
<td>0.70</td>
<td>51</td>
</tr>
<tr>
<td>0.90</td>
<td>121</td>
</tr>
</tbody>
</table>

5.9.2 Example from Critical Appraisal Skills Programme (CASP)
See [http://www.phru.nhs.uk/casp/appraisa.htm](http://www.phru.nhs.uk/casp/appraisa.htm)
6. **DIAGNOSTIC TESTS**

6.1 **Learning objectives**

1. To understand, and be able to calculate, the parameters that define diagnostic test quality – sensitivity, specificity and likelihood ratio (which is derived from these) as well as the ‘predictive values’ of tests in populations with different prevalence of disease.
2. To understand the clinical uses of these parameters to rule-in or rule-out disease in individual patients and groups.
3. To understand the concepts of pre- and post-test probability of disease and how they are derived in clinical practice.
4. To understand and use a nomogram to define post-test probability of disease, given the pre-test probability and likelihood ratio.
5. To understand the key questions for a critical appraisal of an article on diagnosis.

6.2 **Core reading and key concepts**

**Introduction**

**Clinical Scenario**

A 12-month old baby boy is brought to your surgery by his worried mother. The baby has been feeding poorly, has fever, rhinorrhea and progressive cough for a week. What are the possible causes of these symptoms and signs, and how would you refine the probability of each of them?

As clinicians in this situation, we generate a **differential diagnosis** that includes the most important possible causes, and then we use various pieces of information (eg history, physical examination, clinical laboratory tests and radiographic studies) to help us **refine the probability of disease** for each significant possible diagnosis.

Diagnostic test may include any type of information that could be helpful in making a diagnosis, including history-taking and physical examination, but the term is usually applied to ward, laboratory or imaging tests.

The usual reasons given for requesting diagnostic tests are: to identify the disease responsible for a patient’s symptoms, to make better decisions about treatment, to give patients more accurate information about prognosis and to monitor treatment and disease progression.

The most important reason we request diagnostic tests is to **refine the probability of the presence or absence of disease**. The results of diagnostic tests **change the likelihood** of a particular diagnosis. We can improve our use of diagnostic tests by learning to apply the evidence-based techniques of refining probability.

A useful equation for the use of diagnostic tests is:

What we think before the test + Test Information = What we think after

Sensitivity and specificity, and likelihood ratio, which is derived from these, inform us about the chance, in patients with or without the disease, that the test is positive or negative. They define precisely how well a diagnostic test rules-in or rules-out the presence of disease, but have limitations when applied in isolation to patients with different probabilities of disease.
Using the pre-test probability of disease in combination with the likelihood ratio, clinicians can calculate the post-test probability of disease, and thus make informed decisions as to whether, and by how much, diagnostic tests improve diagnostic certainty for particular groups or individuals.

The importance of this approach is emphasised by the finding that misdiagnosis rates have not changed in major US teaching hospitals over the last four decades, despite an explosion of sophisticated diagnostic technologies. Unnecessary testing also exposes patients to unnecessary risks including false-positive diagnoses, and significantly increases healthcare costs.

Pre-test probability estimates

‘What we think before’ and ‘What we think after’ are estimates of the likelihood of disease, and can be expressed in probability terms:

What we think before test + Information about the test’s quality = What we think after

OR: Pre-test probability + information about test quality (eg Likelihood ratio) = Post-test probability

pre-test probability = probability that a patient has a disease before a test is done
post-test probability = the probability that a patient has a disease after a test is done

Estimates of pre-test probability can be generated from three sources:

- clinical data - derived from careful clinical assessment, which may include previous test results for the patient, and personal experience
- published data - giving the prevalence of disease in a defined clinical setting /population
- clinical prediction rules (clinical tools that quantify the individual contributions that various components of the history, physical examination, and basic investigations make towards diagnosis, prognosis or treatment of an individual.

Diagnostic test characteristics (test information)

Test information has traditionally been reported in terms of sensitivity, specificity and predictive value. Likelihood ratios, derived from sensitivity and specificity, are useful to refine the estimated probability of disease in our patients.

Sensitivity, Specificity and Predictive Values: the 2 x 2 Table

These test performance parameters are used to assess diagnostic tests that have only two outcomes (+/-), and require that the population studied undergoes a ‘gold standard’ test to ‘rule-in’ or ‘rule-out’ disease (eg evaluating the diagnostic capabilities of ultrasound for the diagnosis of gallstones in patients who all subsequently had a cholecystectomy).
<table>
<thead>
<tr>
<th>DISEASE POSITIVE</th>
<th>DISEASE NEGATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEST POSITIVE</strong></td>
<td><strong>TEST NEGATIVE</strong></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>True positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>False negative</td>
<td>True negative</td>
</tr>
<tr>
<td>a+c</td>
<td>b+d</td>
</tr>
<tr>
<td>Total with disease</td>
<td>Total without disease</td>
</tr>
</tbody>
</table>

**Sensitivity** = probability of a positive test when disease is present  
= \( \frac{a}{a+c} \)

**Specificity** = probability of a negative test when disease is absent  
= \( \frac{d}{b+d} \)

**Predictive Value** = probability of disease, given the results of the test. This is sometimes called the **post-test probability** of disease

- **positive predictive value** = probability of disease in patients with a positive (abnormal) test result  
  = post-test probability of disease given a positive test  
  = \( \frac{a}{a+b} \)

- **negative predictive value** = probability of no disease in patients with a negative (normal) test result  
  = post-test probability of disease given a negative test  
  = \( \frac{d}{c+d} \)

**Note:** Sensitivity, specificity and their derivative, likelihood ratio, are properties of tests, while predictive values are functions of both of these and the prevalence of disease in a group, or the chance (or probability) in an individual.

**Interpreting test results using sensitivity, specificity and predictive values**
Sensitivity and specificity describe the proportion of positive and negative test results in populations **when we already know who has the disease** or not. In isolation, they are of limited use in interpreting test results in the clinical setting, when we don’t know whether a patient does, or does not have, the disease\(^1\).

- **Tests with high sensitivity and/or high specificity** are especially helpful:
  - a highly sensitive test, when negative, rules out disease (SnNout)
  - a highly specific test, when positive, rules in disease (SpPin)
  - see self-complete exercise 1 (6.5)

The predictive value is determined by the sensitivity and specificity of the test, and also by the **prevalence of disease in the population being tested**. Predictive value, based on probability of disease in the group, is also of limited value in helping us refine probability in a particular patient, who may have a different pre-test probability of disease.

- When prevalence of disease is relatively high (over several percent), a highly specific test will perform well. But at lower prevalences, the positive predictive value drops to almost zero, and
the test becomes virtually useless for diagnosing disease, as a positive test result is most likely to be a false positive.
Example: HIV testing in persons at extra-ordinary low risk of HIV disease

Likelihood Ratios and Post-test probabilities
Diagnostic tests enable clinicians to revise the probability of disease, and to move from an initial assessment of the probability of disease (pre-test probability) to determine the final assessment of disease probability (post-test probability) -

\[
\text{Pre-test probability} + \text{Information about the test’s quality (Likelihood Ratio)} = \text{Post-test probability}
\]

Likelihood ratios are derived from sensitivity and specificity. The advantages of the likelihood ratio approach are:

- Likelihood ratios can deal with tests with more than two possible results (not just normal/abnormal).
- The magnitude of the likelihood ratio give intuitive meaning as to how strongly a given test result will raise (rule-in) or lower (rule-out) the likelihood of disease.
- Likelihood ratios are more easily applied to a series of diagnostic tests.
- Likelihood ratios can be easily used to determine post-test ‘odds’ and with a nomogram to determine post-test ‘probability’ (chance).

LR definitions

- defines how much a positive or negative test result modifies the probability of disease, and is expressed as a ratio:

\[
\text{LR} = \frac{\text{probability of the test result in a patient with the disease}}{\text{probability of the test result in a patient without the disease}}
\]

- positive and negative likelihood ratios are defined as:

\[
\text{LR}^+ = \frac{\text{probability of a positive test result in a patient with the disease}}{\text{probability of a positive test result in a patient without the disease}}
\]

\[
\text{LR}^- = \frac{\text{probability of a negative test result in an individual with the disease}}{\text{probability of a negative test result in an individual without the disease}}
\]

Effect of the value of a likelihood ratio on post-test probability of disease

- LR.s range from 0 to infinity
- A value of 1 means the test provides no additional information.
- Ratios above 1 increase the likelihood of disease
- Ratios below 1 decrease the likelihood of disease
- LR.s greater than 10 or less than 0.1 generate large and often conclusive changes from pre-test to post-test probability
- LR.s of 5 to 10 and 0.1 to 0.2 generate moderate shifts in pre-test to post-test probability
- LR.s of 2 to 5 and 0.5 to 0.2 generate small (but sometimes important) changes in probability
- LR.s of 1 to 2 and 0.5 to 1 alter probability to a small (and rarely important) degree
Examples of LRs:
* University of Toronto Centre for EBM
  http://www.cebm.utoronto.ca/glossary/lrs.htm#table
* Centre for Evidence-Based Medicine Oxford UK

Likelihood ratio, sensitivity and specificity
- LR+ and LR- are defined in terms of sensitivity and specificity:
  \[ \text{LR+} = \frac{\text{sensitivity}}{1-\text{specificity}} \text{ OR true+ve/false+ve} \]
  \[ \text{LR-} = \frac{1-\text{sensitivity}}{\text{specificity}} \text{ OR false-ve/true-ve} \]

See self-complete Exercise 3

- Interactive Calculators for diagnostic test 2 x 2 tables to estimate sensitivity, specificity, positive predictive value, negative predictive value and likelihood ratio are available from the following websites:
  * Centre for Evidence-Based Medicine Oxford UK:
    http://www.cebm.net/index.aspx?o=1160
  * University of Toronto Centre for EBM: http://www.cebm.utoronto.ca/practise/ca/statscal/

A palm version of this calculator is downloadable from:
http://www.cebm.utoronto.ca/palm/ebmcalc/

Using likelihood ratios with a nomogram
- Where the LR and pre-test probability are known, a nomogram can be used to quickly determine the post-test probability
• An easy-to-use interactive nomogram is available from the Centre for Evidence-Based Medicine Oxford UK: http://www.cebm.net/index.aspx?o=1161

Using likelihood ratios and odds
• The product of the LR and the pre-test odds determines post-test odds. Simple calculations are used to convert odds to probability:

\[
\text{Post-test odds} = \text{Pre-test odds} \times LR \\
\text{Odds} = \frac{\text{Probability}}{\text{1 - Probability}} \\
\text{Probability} = \frac{\text{Odds}}{\text{1 + Odds}}
\]

• Use of the nomogram saves having to convert odds to probability and probability to odds

Using likelihood ratios for multiple tests
• The likelihood ratio method can take into account test results at multiple different levels of severity (see Self-complete exercise 5)
• It is also easy to calculate post-test odds after a series of diagnostic tests:

\[
\text{Post-test Odds} = \text{Pre-test Odds} \times LR_1 \times LR_2 \times LR_3 \ldots \times LR_n
\]
6.3 Other useful concepts

**Accuracy**
Proportion of all test results, both positive and negative, that are correct.

**Reproducibility**
Refers to the repeatability, consistency, or stability of a measure or a test from one occasion to another. Tests may not always give the same result for a variety of reasons eg. test variability and observer interpretation.

**Intra- and inter-observer agreement**
Where tests involve observer interpretation, tests should be evaluated for a summary measure of observer variability.

This involves measuring agreement between repeat observations made by the same and between different observers.

The Kappa statistic is the usual measure of the percent agreement between raters that occurs beyond chance. (see also 5.2)

**Spectrum composition**
Note that the sensitivity and/or specificity of a test are not always constant, but may also depend on the characteristics of the population studied. Diagnostic test evaluation should include data on age distribution, sex distribution, summary of presenting clinical symptoms and/or disease stage, and eligibility criteria for study subjects, etc.

**Bias**
Bias may occur if the test result and disease status are not established independently of each other.

**Work-up bias**
- in some circumstances the result of the test may influence the subsequent diagnostic work-up needed
  - eg. negative test result - additional necessary diagnostic tests not requested

**Diagnostic-review bias**
- the test result may affect the use of data to establish diagnosis, which can lead to under and/or over diagnosis

**Test-review bias**
- test result interpretation occurs after diagnosis established
  - eg. after ‘gold standard’ test result known

**Incorporation bias**
- result of test is incorporated into evidence used to establish diagnosis of disease. The evidence used to establish the true diagnosis must be independent of the test result
6.4. Evaluating and applying the results of an article about diagnosis

Table 2: Evaluating and Applying the Results of an Article about Diagnosis

<table>
<thead>
<tr>
<th>Are the results of the study valid?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary guides:</strong></td>
</tr>
<tr>
<td>Was there an independent, blind comparison with a reference standard?</td>
</tr>
<tr>
<td>Did the patient sample include an appropriate spectrum of patient to whom the diagnostic test will be applied in clinical practice?</td>
</tr>
<tr>
<td><strong>Secondary guides:</strong></td>
</tr>
<tr>
<td>Did the results of the test being evaluated influence the decision to perform the reference standard?</td>
</tr>
<tr>
<td>Were the methods for performing the test described in sufficient detail to permit replication?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What were the results?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are likelihood ratios for the test results presented or data necessary for their calculation provided?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Will the results help me in caring for my patients?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will the reproducibility of the test result and its interpretation be satisfactory in my setting?</td>
</tr>
<tr>
<td>Are the results applicable to my patient?</td>
</tr>
<tr>
<td>Will patients be better off as a result of the test?</td>
</tr>
</tbody>
</table>

Worked example of an article about diagnosis

Introduction: the clinical problem

You are a general practitioner. A 68-year old male patient consults you with a history of increasing localised back pain. He is an ex-smoker and has a history of mild asthma. Three weeks previously, he had presented with right upper quadrant pain, and some shortness of breath. An ultrasound showed no right upper quadrant abnormality, and a review of his records showed a normal chest x-ray taken 12 months previously. The pain has continued and has localised to the lower back.

The patient’s weight is stable. Full blood count, routine biochemistry are normal, and ESR is 22mm/hour. An x-ray of the thoraco-lumbar spine shows degenerative changes only.

The patient’s pain continues over the next two months. Although controlled with medication, you are concerned about a non-mechanical cause of his pain. You are aware that plain x-rays are not ‘very good’ for detecting cancer, but have no idea exactly what value they have. You feel you may have been falsely reassured by the normal x-ray, and wonder about the significance of his clinical and laboratory findings.

You formulate the following question: In a middle aged man with back pain (the patient), which readily available and relatively cheap tests (the intervention) change the probability that cancer (the clinical outcome) is present, in comparison to clinical findings (the comparison)?

You conduct a MEDLINE Search. You use the MeSH browser for the terms ‘cancer’ and ‘back pain’ and ‘probability’ and find the MeSH terms ‘Neoplasms’ and ‘Back pain’ and ‘Probability’. Next you search with the appropriate subheading for each, ‘Back pain/etiology’ AND ‘Neoplasms/diagnosis' AND ‘Probability’, which leads to ~20 citations, the most relevant of which is:
Deyo RA, Diehl AK. Cancer as a cause of back pain: frequency, clinical presentation, and diagnostic strategies. J Gen Intern Med 1988; 3:230-238. The outline of this paper is summarised below, but you should peruse the whole paper as well:

**Question:** In a middle-aged man with back pain (the patient), which readily available and relatively cheap tests (the intervention) change the probability that cancer (the clinical outcome) is present, in comparison to clinical findings (the comparison intervention)?

**Design:** Prospective study to define sensitivity, specificity and likelihood ratios for clinical and diagnostic test findings in a large series of patients with back pain.

**Patients:** 1975 indigent patients, aged 15-86, attending a walk-in primary care clinic in Seattle, with a chief complaint of back pain.

**Main outcome measure:** Identification of cancer as the cause of back pain was performed by follow-up as performed through the regional cancer registry at least 6 months after the first visit.

**Main results:** 54% were seeking medical care for back pain for the first time. 76% had had pain for less than three months, and 84% had low back pain. The prevalence of underlying cancer was low, with thirteen patients (0.66%) found to have cancer as the cause of their back pain.

Patients with underlying malignancies often had few suggestive signs. Findings significantly associated with underlying cancer (p< 0.05) were: age greater than or equal to 50 years, previous history of cancer, duration of pain greater than 1 month, failure to improve with conservative therapy, elevated erythrocyte sedimentation rate (ESR), and anaemia.

**Conclusion:** The authors have developed an algorithm combining historical features and ESR results. With suggestive clinical features, an elevated ESR should prompt radiography. In the face of suggestive clinical findings and an elevated ESR, a negative spine x-ray should be interpreted with caution.

In this patient group, use of the algorithm would have limited x-ray utilisation to just 22% of subjects while recommending an x-ray for every cancer patient. It would further suggest which patients with negative x-ray findings require further work-up.

**Are the results of the study valid?**

**Was there an independent, blind comparison with a reference standard?**

The reference standard was registry identified cancer diagnosis. The tumour registry included every patient with a histologic diagnosis of cancer made in the local public area health system. While this method would fail to identify cancer patients who sought care outside this area health system, patients attending the walk-in clinic say they do not attend other area clinics, and are also highly likely at attend the local public hospital for complex illnesses.

It is possible however, that patients in this population might have left the area, and had cancer diagnosed elsewhere. The effect of such a bias would be to weaken the link between positive test results and diagnosis, but could potentially overestimate the link between a negative test result and no diagnosis of cancer.
Did the patient sample include an appropriate spectrum of patients to whom the diagnostic tests will be applied in clinical practice?

The study population was composed of poor, self-referred Mexican Americans, Hispanics or black Americans. This study population would not be exactly comparable to patients with back pain seen in most Australian general practices. You would expect it to have a similar spectrum of disease, but realise that this population could have higher prevalences of (work-related) injury and/or cancer as causes of back pain.

Tests were applied to approximately two thirds of all back pain patients presenting during the study period, but patients presenting at night and weekends were less represented than those presenting during daylight hours. Such patients may be ‘sicker’ and are less likely to be shift workers.

Not all patients received every diagnostic test, with ‘sicker’ patients more likely to have tests. Abnormal results among non-cancer patients may therefore be inflated.

Given these limitations and the lack of studies on other populations, you decide that the tests were applied to an acceptable spectrum of patients.

What were the results?
The diagnostic value for clinical and laboratory findings are presented in terms of sensitivity, 1-specificity, and LR (Deyo and Diehl, Tables 2 and 3).

A previous history of cancer had the highest LR of 14.7, and the presence of this finding would greatly increase the probability of cancer. Other findings, which increased the probability of cancer to a modest degree (LRs of 2-3), were unexplained weight loss, failure to improve after a previous visit in the last month, and age over 50.

For test results, an ESR>100 mm Hg had a LR of 55, and the presence of a lytic and blastic lesion on x-ray had a LR of 120. The presence of these findings would greatly increase the probability of cancer as the cause of back pain, whatever the pre-test probability of cancer.

Will the results help me in caring for my patients?

To apply the results to your patient, your review your notes, and decide that the pre-test probability of cancer was about 10% on the patient’s initial presentation, increasing to about 25%, following the patient's last visit. (You notice this approximates to the shift in probability expected when you use the nomogram and apply a LR of 3 for the clinical feature “failure to improve over a month”.) How does a negative spine x-ray result change the probability of cancer as the cause of back pain in your patient? You calculate the LR of a normal x-ray using data on sensitivity and specificity from Table 3 for “either compression fracture or lytic / blastic lesion”:

\[
LR_{-ve} = \frac{1 - \text{sensitivity}}{\text{specificity}} = \frac{0.3}{0.95}, \text{ giving a LR of 0.3.}
\]

Using the nomogram you see how the meaning of the test result depends on the pre-test probability of disease:

- pre-test probability of 10% : LR of 0.3 reduces post-test probability to <5%, which you are happy to accept
- pre-test probability of 25% : LR of 0.3 reduces post-test probability to about 10%, which is not a figure that totally reassures you that all is well

You decide that plain films have falsely reassured you in this patient. In future you might refer such a patient to a specialist or choose more sensitive investigations.
You may feel concerned about the precision of assigning a pre-test probability. The nomogram is useful for deciding at what probability level we would make or not make a particular management decision. For example in this patient, you might have requested a MRI examination if the probability of cancer was above 5%, and have been happy to wait and see if it was below this. With pre-test probabilities above 10%, the post-test probability of disease, given a negative x-ray, would be higher than our threshold for action. **So a precise estimate of pre-test probability is not necessary – just a reasonable estimate, the kind of estimate, which underpins our decisions on a daily basis.**

### 6.5 Self-complete exercises

**Exercise 1:**

**Calculating sensitivity and specificity for a dichotomous test result**

Davie et al assessed the test performance of key symptoms and signs for diagnosis of congestive heart failure (CHF).

Below are 2 x 2 tables relating to the symptom of ‘dyspnea on exertion’ (DOE) and the sign ‘gallop (S3) murmur’.

Calculate sensitivity and specificity for each test.

**Test = DOE**

<table>
<thead>
<tr>
<th></th>
<th>CHF</th>
<th>no CHF</th>
<th>Total positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOE</td>
<td>41</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>no DOE</td>
<td>0</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>218</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Total with disease</th>
<th>Total without disease</th>
<th>Grand total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOE</td>
<td>41</td>
<td>183</td>
<td>Total positive</td>
</tr>
<tr>
<td>no DOE</td>
<td>0</td>
<td>35</td>
<td>Total negative</td>
</tr>
</tbody>
</table>

**Test = S3 murmur**

<table>
<thead>
<tr>
<th></th>
<th>CHF</th>
<th>no CHF</th>
<th>Total positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3 murmur</td>
<td>10</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>no S3 murmur</td>
<td>31</td>
<td>215</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Total with disease</th>
<th>Total without disease</th>
<th>Grand total</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3 murmur</td>
<td>10</td>
<td>3</td>
<td>Total positive</td>
</tr>
<tr>
<td>no S3 murmur</td>
<td>31</td>
<td>215</td>
<td>Total negative</td>
</tr>
</tbody>
</table>

**Methodology for evaluating a 2 x 2 Table**

[http://gim.unmc.edu/dxtests/howto.htm](http://gim.unmc.edu/dxtests/howto.htm)
### Disease Present | Disease Absent
--- | ---
**Test Positive** |  
TP | FP | Total positive
**Test Negative** |  
FN | TN | Total negative

<table>
<thead>
<tr>
<th>Total with disease</th>
<th>Total without disease</th>
<th>Grand total</th>
</tr>
</thead>
</table>

Multiply the **Grand total** by the **Pre-test probability** to get the **Total with disease**.

Compute the **Total without disease** by subtraction.

Multiply the **Total with disease** by the **Sensitivity** to get the number of **True positives**.
Multiply the **Total without disease** by the **Specificity** to get the number of **True Negatives**.

Compute the number of **False positives** and **False negatives** by subtraction.

Compute the **Total positive tests** and **Total negative tests** by addition across the rows.

**Predictive value of a positive test** is **True positives** divided by **Total positive tests**.
**Predictive value of a negative test** is **True negatives** divided by **Total negative tests**.

---

**Exercise 2:**
**Calculating predictive values /post-test probability for a dichotomous test result**

Consider the use of the ANA (antinuclear antibody) test in the diagnosis of SLE (systemic lupus erythematosus). In a hypothetical rheumatology practice of 100,000 patients, the prevalence of SLE in patients on whom an ANA test was done was 2.88%. The sensitivity of the ANA for SLE is 98% and the specificity is 93%. A patient from this practice has a positive ANA test.
Use the methodology for evaluating a 2 x 2 table to determine:

<table>
<thead>
<tr>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>TP</td>
</tr>
<tr>
<td>Test Negative</td>
<td>FN</td>
</tr>
<tr>
<td>Total with disease</td>
<td>Total without disease</td>
</tr>
</tbody>
</table>

How many patients in the practice have SLE? ________________________

How many patients do not have SLE? ________________________

The TP, FN, FP, TN patient numbers for the practice. ________________________

How many patients test positive and negative for ANA? ________________________

What is the post-test probability of SLE, given the patient’s positive test result?

Exercise 3:
Calculating Likelihood Ratios (LRs) from sensitivity and specificity

Using the data from Exercise 2, calculate the likelihood ratio for a positive test result.

Exercise 4:
Calculating LRs & post-test probability for tests with more than 2 possible results

Likelihood ratios work well for tests with multiple qualitative results such as a ventilation perfusion (V/Q) scan which can be interpreted as normal, low probability, intermediate probability, and high probability of pulmonary embolism. For example, the PIOPED Study (JAMA 1990; 263:2753-2759) compared the V/Q scan with angiography and reported the following data:

<table>
<thead>
<tr>
<th>Scan Category</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
</tr>
</thead>
<tbody>
<tr>
<td>High probability</td>
<td>41</td>
<td>97</td>
</tr>
<tr>
<td>High or intermediate probability</td>
<td>82</td>
<td>52</td>
</tr>
<tr>
<td>High, intermediate or low probability</td>
<td>98</td>
<td>10</td>
</tr>
</tbody>
</table>
A patient with a 30% pre-test probability of pulmonary embolism has an intermediate probability V/Q scan.

**Calculate the likelihood ratio for a high or intermediate probability scan from the sensitivity and specificity data.**

**Use the nomogram to determine the post-test test probability of disease for this patient.**

This result, however, is not the best use of the available data because it lumps the high probability and intermediate probability scans together so that a sensitivity and specificity can be reported. The paper also lists the raw data by *individual* test category.

<table>
<thead>
<tr>
<th>Scan Category</th>
<th>PE Present</th>
<th>PE absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>High probability</td>
<td>102</td>
<td>14</td>
</tr>
<tr>
<td>Intermediate probability</td>
<td>105</td>
<td>217</td>
</tr>
<tr>
<td>Low probability</td>
<td>39</td>
<td>199</td>
</tr>
<tr>
<td>Normal or near normal</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>251</td>
<td>480</td>
</tr>
</tbody>
</table>

From these data (shown below in the two left columns), calculate the likelihood ratio for each test result.

Now using the nomogram, determine the post-test probability for the patient with a 30% pre-test probability and an intermediate probability scan.
Exercise 5:
Predicting pre-test probability and calculating predictive values /post-test probability

Case Scenarios

**Patient A** is a 55-year old mildly hypertensive man with a 4-week history of substernal pain that radiates to his neck, lower jaw and down the inner aspect of his left arm. It is precipitated by climbing stairs or waking uphill, and disappears after 3 to 5 minutes of rest. He had a mild episode while undressing for his examination. You think he has an S₄ gallop. You decide he has classic angina and consider requesting an exercise ECG to confirm the diagnosis.

**Patient B** is a 35-year old man, who is otherwise healthy and has no coronary risk factors. He has had ‘heartburn’ for years and now reports a 6-week history of non-exertional, squeezing pain deep to his lower sternum and epigastrium, usually radiating straight through to his back. It is most likely to occur when he lies down after a heavy meal. The remainder of his history and physical examination are negative. You think that his pain reflects oesophageal spasm and will pursue that diagnosis, but judge that an exercise ECG will rule out significant coronary disease and thus resolve the uncertainty for both of you.

**Patient C** is a 45-year old man with a negative past history and no coronary risk factors save a pack-a-day cigarette habit. He reports a 3-week history of pre-cordial and substernal pain, usually fleeting and of a stabbing quality. You find a costochondral junction that is slightly tender, but pressing it does not reproduce the pain. You conclude that he may have a typical angina and wonder if an exercise ECG would help.

A common problem in internal medicine is deciding how to manage the ambulatory patient with chest pain. The hypothesis list usually includes coronary artery disease and an exercise ECG is often considered to help with the diagnosis. A positive exercise test will suggest coronary artery disease and lead to an angiogram and negative exercise test will reassure the patient and direct our diagnostic effort elsewhere. Or will it?

The patients in a study of the relationship between graded exercise tests and angiographically documented coronary heart disease had a variety of chest pain syndromes and underwent both exercise electrocardiography and coronary angiography. One mm of ST-segment depression or more was selected as a cut-off for a positive exercise ECG. Angiography was taken as the ‘gold standard’: (that is, proof that the disease was present).

350 patients were studied, of which 137 had 70% coronary artery stenosis by angiography after having a positive exercise ECG. 11 did not have significant stenosis following positive test. Of those patients who had a negative exercise ECG (ST depression <1mm), 90 had at least 70% coronary artery stenosis on angiography and 112 had no significant coronary disease.
1. Display the above information in the 2 x 2 table now.

<table>
<thead>
<tr>
<th>Exercise ECG</th>
<th>70% coronary artery stenosis by angiogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>positive</td>
<td>a</td>
</tr>
<tr>
<td>negative</td>
<td>c</td>
</tr>
<tr>
<td>total</td>
<td>total</td>
</tr>
</tbody>
</table>

2. The sensitivity of a test indicates the probability of finding an abnormal result in a diseased patient. What is the sensitivity of an exercise ECG in detecting significant coronary artery disease based on the above study?

3. The specificity of a test indicates the probability of finding a normal result in a patient without the disease. What is the specificity of an exercise ECG in detecting significant coronary artery disease based on the above study?

4. Now consider the accompanying case scenarios. Estimate the pre-test probability of significant coronary artery narrowing in each of the three patients:

   Patient A: ____________________________________________

   Patient B: ____________________________________________

   Patient C: ____________________________________________

5. Which patient is going to benefit the most from having an exercise ECG?

One way of thinking about this is to compare the pre-test probability of the patient having the disease with the post-test probability. Let’s work through the exercise for each of the patients.
6. **Patient A:**
Assume the pre-test probability is 90%. If you have 1000 patients with the same presentation as Patient A, 900 will have coronary artery disease and 100 will not. Assuming that the exercise ECG has the sensitivity and specificity as established by the paper by Bartel et al, complete the following table:

<table>
<thead>
<tr>
<th>70% coronary artery stenosis by angiogram</th>
<th>present</th>
<th>absent</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>exercise ECG positive</td>
<td>a</td>
<td>b</td>
<td>total</td>
</tr>
<tr>
<td>negative c</td>
<td></td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>total 900</td>
<td>total</td>
<td>100</td>
<td>1000</td>
</tr>
</tbody>
</table>

a. The *positive predictive value* (or post-test probability) is the proportion of patients with a positive test result who actually have the disorder. What is the positive predictive value for the matrix above?

b. What is the chance that the patient has coronary artery disease if the test is *negative*?

c. What is the percentage change between the pre-test and post-test probabilities?

Comment on the benefit of doing an exercise ECG in Patient A.

7. **Patient B:**
Assume the pre-test probability is 5%. If you have 1000 patients with the same presentation as Patient B, 50 will have coronary artery disease and 950 will not. Assuming that the exercise ECG has sensitivity and specificity as established by paper by Bartel et al, complete the following table:
a. The positive predictive value (or post-test probability) is the proportion of patients with a positive test result who actually have the disorder. What is the positive predictive value for the matrix above?

b. What is the chance that the patient has coronary artery disease if the test is negative?

c. What is the percentage change between the pre-test and post-test probabilities?

Comment on the benefit of doing an exercise ECG in Patient B.

8. Patient C:
Assume the pre-test probability is 50%. If you have 1000 patients with the same presentation as Patient C, 500 will have coronary artery disease and 500 will not. Assuming that the exercise ECG has the sensitivity and specificity as established by the paper by Bartel et al\(^6\), complete the following table:
EBP Workbook 07

a. The positive predictive value (or post-test probability) is the proportion of patients with a positive test result who actually have the disorder. What is the positive predictive value for the matrix above?

b. What is the chance that the patient has coronary artery disease if the test is negative?

c. What is the percentage change between the pre-test and post-test probabilities?

Comment on the benefit of doing an exercise ECG in Patient C.

Exercise 6:
Critical appraisal of an article about diagnosis

Scenario
You are a resident looking after an emergency department patient with suspected pulmonary embolism. Helical CT, combined with other investigations, is commonly performed in your hospital for investigation of pulmonary embolism, but you are unsure of its accuracy and its role.

You have found the American College of Physicians (ACP) Journal Club site very useful for quickly consulting evidence on a subject, and do a basic search using terms ‘pulmonary embol*’ and ‘CT’. This yields commentaries on 4 articles, of which 2 look particularly relevant.

The first comments on a review article on the accuracy of helical CT. This states that the accuracy of helical CT for diagnosis of pulmonary embolism (PE) in symptomatic patients has not been adequately evaluated. Previous studies looking at this question have suffered from selection, workup and diagnostic-review biases, and have shown wide variations in sensitivity and specificity. Two recent reviews did not find one study that satisfied the basic methodological criteria necessary for the results to be regarded as valid (ACP Journal Club. 2000 July-Aug; 133:33: http://www.acpjc.org/Content/133/1/ISSUE/ACPJC-2000-133-1-033.htm)


Use the worksheet below to undertake a critical appraisal of this paper.
### Diagnostic worksheet

#### Citation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>

#### Are the results of this diagnostic study valid?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was there an independent, blind comparison with a reference ('gold') standard of diagnosis?</td>
<td></td>
</tr>
<tr>
<td>Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?</td>
<td></td>
</tr>
<tr>
<td>Was the reference standard applied regardless of the diagnostic test result?</td>
<td></td>
</tr>
<tr>
<td>Were the methods for performing the test described in sufficient detail to permit replication?</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**What were the results?**

<table>
<thead>
<tr>
<th>Are Likelihood Ratios for the test results presented or data necessary for their calculations (ie. sensitivity and specificity) provided?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Are the valid results of this diagnostic study important?

Your calculations

<table>
<thead>
<tr>
<th>Target Disorder</th>
<th>Present</th>
<th>Absent</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic test result</td>
<td>Present</td>
<td>Absent</td>
<td>Totals</td>
</tr>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>Test Negative</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Totals</td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

Can you apply this valid, important evidence about a diagnostic test in caring for your patient?

| Is the diagnostic test available, affordable, accurate, and precise in your setting? | |
| Can you generate a clinically sensible estimate of your patient's pre-test probability (from personal experience, prevalence statistics, practice databases, or primary studies)? Are the study patients similar to your own? Is it unlikely that the disease possibilities or probabilities have changed since the evidence was gathered? | |
Will the resulting post-test probabilities affect your management and help your patient? Could it move you across a test-treatment threshold? Would your patient be a willing partner in carrying it out?

| Would the consequences of the test help your patient? |

**Additional Notes**
6.6 Answers to self-complete exercises

Exercise 1: Calculating sensitivity and specificity for a dichotomous test result

Test = DOE
Sens = 41/(41+0) = 100%
Spec = 35/(183+35) = 17%

Note
If a patient does not complain of dyspnea on exertion, it is very unlikely that they have CHF (0 out of 41 patients with CHF did not have this symptom).
• A highly sensitive test, when negative, rules out disease (SnNout)
• The acronym "SnNout" is taken from the phrase: "Sensitive test when Negative rules Out disease"

Test = Gallop murmur
Sens = 10/41 = 24%
Spec = 215/218 = 99%

Note
Thus, if a patient has a gallop murmur, they probably have CHF (10 out of 13). A highly specific test, when positive, rules in disease (SpPin)
The acronym "SpPin" is taken from the phrase: "Specific test when Positive rules In disease"

Exercise 2: Calculating predictive values /post-test probability for a dichotomous test result

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>2822</td>
<td>6798</td>
</tr>
<tr>
<td>Test Negative</td>
<td>58</td>
<td>90322</td>
</tr>
<tr>
<td></td>
<td>2880</td>
<td>97120</td>
</tr>
</tbody>
</table>

Post-test probability of SLE, for the patient's positive test result
= 2822/9620 = 0.293
**Exercise 3:**
**Calculating likelihood ratios from sensitivity and specificity:**

\[ LR^+ = \frac{\text{probability of a positive test result in a patient with the disease}}{\text{probability of a positive test result in a patient without the disease}} \]

\[ LR^+ = \frac{98}{5} = 14 \]

\[ LR^- = \frac{1-98}{5} = 0.02 \]

**Exercise 4:**
**Calculating Likelihood Ratios and post-test probability for tests with more than two possible results**

Calculate the LR for a high or intermediate probability scan from the sensitivity and specificity data.

\[ LR^+ = \frac{82}{48} = 1.7 \]

Use the nomogram to determine the post-test test probability of disease for this patient.
Post-test probability = 0.4

From these data (shown below in the two left columns), calculate the LR for each test result.

For high probability: \( LR^+ = (102/251) / (14/480) = 13.9 \)

<table>
<thead>
<tr>
<th>Scan Category</th>
<th>PE Present</th>
<th>PE absent</th>
<th>Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>High probability</td>
<td>102</td>
<td>14</td>
<td>13.9</td>
</tr>
<tr>
<td>Intermediate probability</td>
<td>105</td>
<td>217</td>
<td>0.93</td>
</tr>
<tr>
<td>Low probability</td>
<td>39</td>
<td>199</td>
<td>0.37</td>
</tr>
<tr>
<td>Normal or near normal</td>
<td>5</td>
<td>50</td>
<td>0.19</td>
</tr>
<tr>
<td>Total</td>
<td>251</td>
<td>480</td>
<td></td>
</tr>
</tbody>
</table>

Now using the nomogram, determine the post-test probability for the patient with a 30% pre-test probability and an intermediate probability scan.
Post-test probability = 0.3

**Note**

This post-test probability is lower than that previously obtained because we are using all of the information in the data we have available. The LR approach allows us to work with individual test results without having to choose an arbitrary cut point by which to dichotomise the results into ‘positive’ and ‘negative’. Also notice again, the intuitive value of the likelihood ratio number. An intermediate probability scan has a LR very close to 1. This means that intermediate probability scans should not appreciably change your pre-test diagnostic suspicion.
Exercise 5:
Predicting pre-test probability and calculating predictive values / post-test probability

1. Display the above information in the 2 x 2 table now.

<table>
<thead>
<tr>
<th></th>
<th>70% coronary artery stenosis by angiogram</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>present</td>
</tr>
<tr>
<td>exercise ECG</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>negative</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>227</td>
</tr>
</tbody>
</table>

2. The sensitivity of an exercise ECG in detecting significant coronary artery disease
   \[ \frac{a}{a+c} = \frac{137}{137+90} = \frac{137}{227} = 60\% \]

3. The specificity of an exercise ECG in detecting significant coronary artery disease
   \[ \frac{d}{b+d} = \frac{112}{11+112} = 91\% \]

   The pre-test probability of significant coronary artery narrowing in each of the three patients:
   - Patient A: 90%
   - Patient B: 5%
   - Patient C: 50%

4. Which patient is going to benefit the most from having an exercise ECG?
   Patient C

5. Patient A:

<table>
<thead>
<tr>
<th></th>
<th>70% coronary artery stenosis by angiogram</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>present</td>
</tr>
<tr>
<td>exercise ECG</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>negative</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>900</td>
</tr>
</tbody>
</table>

   The positive predictive value = \[ \frac{a}{a+b} = \frac{540}{540+9} = 98\% \]

   What is the chance that the patient has coronary artery disease if the test is negative?
   \[ \frac{c}{b+c} = \frac{360}{100} = 80\% \]
What is the percentage change between the pre-test and post-test probabilities?
+ 8%

Comment on the benefit of doing an exercise ECG in Patient A.
No benefit - a negative test doesn't rule out CAD. Little extra information by doing the test.

6. Patient B:

<table>
<thead>
<tr>
<th></th>
<th>present</th>
<th>absent</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>exercise ECG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td>a 30</td>
<td>b 86</td>
<td>116</td>
</tr>
<tr>
<td>negative</td>
<td>c 20</td>
<td>d 864</td>
<td>884</td>
</tr>
<tr>
<td>total</td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

The positive predictive value = \( \frac{a}{a+b} = \frac{30}{30+86} = 26\% \)

What is the chance that the patient has coronary artery disease if the test is negative?
\( \frac{20}{884} = 2\% \)

What is the percentage change between the pre-test and post-test probabilities?
2%

Comment on the benefit of doing an exercise ECG in Patient A.
Little benefit in doing the test. High false positive rate.

7. Patient C:

<table>
<thead>
<tr>
<th></th>
<th>present</th>
<th>absent</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>exercise ECG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td>a 300</td>
<td>b 45</td>
<td>345</td>
</tr>
<tr>
<td>negative</td>
<td>c 200</td>
<td>d 455</td>
<td>655</td>
</tr>
<tr>
<td>total</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

The positive predictive value = \( \frac{a}{a+b} = \frac{300}{300+45} = 87\% \)

What is the chance that the patient has coronary artery disease if the test is negative?
\( \frac{200}{655} = 31\% \)

What is the percentage change between the pre-test and post-test probabilities?
+ 37%

Comment on the benefit of doing an exercise ECG in Patient A.
Largest shift - largest benefit.
### Exercise 6: Critical appraisal of an article about diagnosis

#### Diagnostic worksheet

<table>
<thead>
<tr>
<th>Citation</th>
</tr>
</thead>
</table>

#### Are the results of this diagnostic study valid?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was there an independent, blind comparison with a reference (‘gold’) standard of diagnosis?</td>
<td>All 299 patients presenting to an emergency department with suspected pulmonary embolus and a plasma D-dimer level &gt; 500 µg/L, had a blinded comparison of helical CT with a validated diagnostic algorithm. This algorithm used standard diagnostic techniques to make a definitive diagnosis of PE (the gold standard): Patients were considered to have PE if they had a positive pulmonary angiographic result, a high-probability lung scan result, or deep venous thrombosis detected by compression ultrasonography - at presentation or 3-month follow-up.</td>
</tr>
<tr>
<td>Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?</td>
<td>Yes. Patients comprised a consecutive series of emergency patients, with a broad spectrum of patient characteristics and clinical probability of PE (detailed in Table1). There were no significant differences between included and excluded patients for demographic and risk factors, nor for clinical presentation and clinical probability for PE. Prevalence of PE in groups with low, intermediate and high probability was appropriate. Proportion of inconclusive scans was appropriate.</td>
</tr>
<tr>
<td>Was the reference standard applied regardless of the diagnostic test result?</td>
<td>Yes. All 299 patients had helical CT scans. Scans were de-identified, were withheld from clinicians, and were read after 3 months by radiologists blinded to all clinical data and the results of the standard tests. These standard tests were interpreted without knowledge of CT scan results. Inter-observer agreement between 3 radiologists was calculated.</td>
</tr>
</tbody>
</table>
Were the methods for performing the test described in sufficient detail to permit replication?

Yes.

Helical CT was done on pulmonary arteries up to and including the segmental vessels from the level of the aortic arch to the lowest hemidiaphragm. Non-ionic contrast material was injected, and a subspecialty-trained chest radiologist scored each vessel for the presence or absence of clot.

What were the results?

Are LRs for the test results presented or data necessary for their calculations provided?

The prevalence of PE in the study population was 39%.

Inter-observer agreement among 3 radiologists was high (κ = 0.82 to 0.90). 12 CT scans were inconclusive. The sensitivity of CT was 70% (95% CI 62 to 78), and the specificity was 91% (CI 86 to 95). LRs for positive, negative, and inconclusive CT results are tabulated below.

Are the valid results of this diagnostic study important?

For patients with conclusive results:

<table>
<thead>
<tr>
<th>Target Disorder</th>
<th>Present</th>
<th>Absent</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>81</td>
<td>15</td>
<td>96</td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>a + b</td>
<td></td>
</tr>
<tr>
<td>Test Negative</td>
<td>35</td>
<td>156</td>
<td>191</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
<td>c + d</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

Can you apply this valid, important evidence about a diagnostic test in caring for your patient?

Is the diagnostic test available, affordable, accurate, and precise in your setting?

Helical CT is available in most hospitals where patients are likely to present with PE. The procedure is expensive for both public and private healthcare providers, and there will be cost savings through elimination of unnecessary testing.

For patients with conclusive results, sensitivity of helical CT was 70%, and specificity was 91%. The use of more stringent diagnostic criteria did not change the sens and spec. The likelihood of false positive results increased as the site of diagnosed embolus moved peripherally.

The sensitivity is too low to rule out PE without additional
tests. Where isolated thrombi are diagnosed at the segmental level and there is a low clinical probability of disease, diagnosis is uncertain. Patients with an inconclusive disease should undergo further tests.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can you generate a clinically sensible estimate of your patient's pre-test probability (from personal experience, prevalence statistics, practice databases, or primary studies)? Are the study patients similar to your own? Is it unlikely that the disease possibilities or probabilities have changed since the evidence was gathered?</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Will the resulting post-test probabilities affect your management and help your patient? Could it move you across a test-treatment threshold? Would your patient be a willing partner in carrying it out?</td>
<td>See additional notes</td>
</tr>
<tr>
<td>Would the consequences of the test help your patient?</td>
<td>In patients presenting to the emergency department with suspected pulmonary embolism, a positive helical computed tomographic result indicated a moderate increase in the probability of pulmonary embolism, but a negative result indicated only a small decrease in probability.</td>
</tr>
</tbody>
</table>

**Additional Notes**

**Source:**
ACP Journal Club. 2002 Jan-Feb;136:30
http://www.acpjc.org/Content/136/1/ISSUE/ACPJC-2002-136-1-030.htm or access through 'Clinicians Health Channel' or through your hospital library

The study by Perrier and colleagues is important because it extends our understanding of the ability of helical CT to diagnose or rule out PE. Through careful attention to minimization of selection, workup, and diagnostic-review biases, the authors largely succeeded in overcoming the methodologic flaws that have hampered previous attempts.

As with lung scanning, helical CT is less accurate than many believe. For example, a patient with a 50% pre-test probability of PE and a negative CT result would still have a post-test probability of 25%. A negative CT result is not equivalent to a normal result on a lung scan or an angiogram. Furthermore, when the clinical probability of PE is low, a positive CT result alone may not be sufficient to establish a diagnosis of PE, particularly
for isolated segmental abnormalities.

Although the observed likelihood ratios with CT are inferior to those previously shown for lung scanning, they were derived from different populations, and comparisons should thus be made with caution. This study cannot provide direct evidence for choosing one diagnostic modality over the other. However, the results provide no support in general for replacing lung scanning with helical CT to diagnose PE. Three advantages of helical CT deserve mention. First, this study did not consider the value of CT's ability to identify an alternative diagnosis in up to a third of patients with suspected PE. Second, as with lung scanning, when the combination of pre-test probability and helical CT results yield sufficiently high or low post-test probabilities, the test obviates the need for conventional angiography and its associated higher morbidity. Third, in practice settings in which clinicians can obtain a CT scan more quickly than a lung scan or an angiogram, the use of CT may prevent delays in therapy.

<table>
<thead>
<tr>
<th>CT test results</th>
<th>Likelihood ratios (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>8.3 (5.1 to 13.7)</td>
</tr>
<tr>
<td>Negative</td>
<td>0.34 (0.26 to 0.45)</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0.31 (0.08 to 1.21)</td>
</tr>
</tbody>
</table>

6.7 Recommended reading

- Jaeschke R, Guyatt G, Sackett DL, for the Evidence Based Working Group. Users' IA Guides to the Medical Literature, III: how to use an article about a diagnostic test, B: what are the results and will they help me in caring for my patients? JAMA. 1995; 274: 1630-1632
- ‘EPIQ’ (Effective Practice, Informatics & Quality Improvement) is produced by Professor Rod Jackson (University of Auckland). It has very useful materials including notes and tools. [http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm](http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm)

6.8 Further reading


6.9 References
### 6.10 Critical appraisal worksheets

#### 6.10.1 Example adapted from Centre for Evidence-Based Medicine, Mount Sinai Hospital

http://www.cebm.utoronto.ca/teach/materials/therapy.htm

**Diagnostic worksheet**

<table>
<thead>
<tr>
<th>Citation</th>
</tr>
</thead>
</table>

**Are the results of this diagnostic study valid?**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was there an independent, blind comparison with a reference ('gold') standard of diagnosis?</td>
<td></td>
</tr>
<tr>
<td>Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?</td>
<td></td>
</tr>
<tr>
<td>Was the reference standard applied regardless of the diagnostic test result?</td>
<td></td>
</tr>
<tr>
<td>Were the methods for performing the test described in sufficient detail to permit replication?</td>
<td></td>
</tr>
</tbody>
</table>
What were the results?

Are likelihood ratios for the test results presented or data necessary for their calculations provided?

Are the valid results of this diagnostic study important?

YOUR CALCULATIONS

<table>
<thead>
<tr>
<th>Target Disorder</th>
<th>Present</th>
<th>Absent</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic test result</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Test Negative</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Can you apply this valid, important evidence about a diagnostic test in caring for your patient?

Is the diagnostic test available, affordable, accurate, and precise in your setting?

Can you generate a clinically sensible estimate of your patient's pre-test probability (from personal experience, prevalence statistics, practice databases, or primary studies)? Are the study patients similar to your own? Is it unlikely that the disease possibilities or probabilities have changed since the evidence was gathered?
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will the resulting post-test probabilities affect your management and help your patient?</td>
<td></td>
</tr>
<tr>
<td>Could it move you across a test-treatment threshold? Would your patient be a willing partner in carrying it out?</td>
<td></td>
</tr>
<tr>
<td>Would the consequences of the test help your patient?</td>
<td></td>
</tr>
</tbody>
</table>

**Additional Notes**

**6.10.2 Example from Critical Appraisal Skills Programme**

See [http://www.phru.nhs.uk/casp/apprais.htm](http://www.phru.nhs.uk/casp/apprais.htm)
6.11 Nomogram

7. STUDIES OF PROGNOSIS AND RISK

7.1 Learning objectives

- To understand what types of study designs are appropriate for evaluating clinical questions about prognosis and risk.
- To understand the potential sources of bias in studies of prognosis and risk.
- To be able to critically appraise studies of disease prognosis and risk and decide how to apply them to the results to the care of individual patients.

7.2 Core reading and key concepts

Introduction
You are a general practitioner. A colleague calls you at home one night to ask you to see a relative of hers the following day. She has just seen her aunt and made a clinical diagnosis of herpes zoster but feels it would be inappropriate for her to manage the problem herself. The patient is a 56-year old woman who has previously been well. She has had 4 days of pain on the left chest wall in a T 6 dermatome distribution. In the last 48 hours she has noticed an erythematous maculopapular rash (in the same area), which has now become vesicular. She feels otherwise well and is on no regular medication.

You agree to see her the following morning. You are aware that antiviral treatment is recommended for the treatment of herpes zoster in patients seen within 72 hours of the onset of vesicles. Studies suggest that some antivirals may reduce the duration of postherpetic neuralgia. In your experience however severe and persistent pain following herpes zoster is not common in patients younger than 60 and antiviral therapy is relatively costly. You decide to see if you can find out more information on the likelihood of developing postherpetic neuralgia in patients who present with herpes zoster in General Practice.

You conduct a search of MEDLINE using the search filter (see 3.2.9) for ‘prognosis’ in ‘PubMed Clinical Queries’, entering the terms ‘postherpetic neuralgia’ AND ‘herpes zoster’. If a ‘specific’ search is selected in ‘PubMed Clinical Queries’, there are ~ 30 citations. You search the titles and abstracts of these articles and find a study that sounds relevant to your question: Helgason S. Petursson G. Gudmundsson S. Sigurdsson JA. Prevalence of postherpetic neuralgia after a first episode of herpes zoster: prospective study with long-term follow-up. BMJ. 2000; 321:794-6.

This study is a prospective follow up study that evaluates the prevalence of post-herpetic neuralgia in patients with a clinical diagnosis of herpes zoster presenting to general practitioners. In order to be able to decide if this study will be helpful in answering your question, you need to decide if the study is both valid and applicable to your patients. In the following sections we will outline some of the potential sources of bias in such follow up studies and discuss the use of critical appraisal guides.
General principles related to the interpretation of prognosis studies

Prognosis is the prediction of the future course of disease following its onset. You may also sometimes hear the term ‘natural history’ being used. This refers to the evolution of disease without medical intervention. Prognostic information is useful to clinicians for two main reasons: patients often seek information about the likely course of their disease, and information about prognosis can be useful for making treatment decisions.

Risk factors for disease and prognostic factors

The factors that predispose individuals to developing a disease are called ‘risk factors’. These may not be the same as the ‘prognostic factors’, which influence the outcome of a disease once established. For example, in a systematic review of psychosocial factors in the aetiology and prognosis of coronary heart disease (prospective cohort studies) Hemingway et al found that depression and anxiety were both risk factors for the development of coronary artery disease and predicted prognosis in patients who developed coronary artery disease. However the evidence suggested that a Type A behaviour pattern was a possible risk factor for the development of coronary artery disease, but did not predict prognosis in those who developed heart disease.

Study types

Several different study types can provide prognostic information. For example randomised controlled trials of health care interventions may provide prognostic information on certain outcomes in the control and intervention groups. In particular where the control group receive placebo or no active intervention, the outcomes in this group may provide information on the natural history of the disease. However randomised controlled trials often have some limitations with respect to examining natural history. Firstly follow up is often limited to a relatively short time frame. Secondly patients that take part in such studies are usually highly selected and may not be representative of the types of patients of interest to you in your practice. Finally placebo effects and the other factors related to participation in the trial may independently influence the natural history in a way that we would not expect in day-to-day practice.

Prognostic information about disease is often obtained from observational studies. Case control studies and cohort studies can be used to examine prognostic factors in relation to particular diseases. Cohort studies can take several forms, for example they may be retrospective or prospective. When examining risk factors for the development of disease, studies often assemble a group of individuals exposed to a potential risk factor and compare the rates of disease during follow up with another group of individuals not exposed to the risk factor. For studies examining prognosis, however often a cohort of patients with established disease will be assembled and followed through time to detect the rates of particular outcome events. Potential prognostic factors can be explored (for example age, sex or severity of disease at baseline). Study design is displayed in Figure 7.1, note that bias may be introduced at each of the steps outlined.
Assembly of patients for prognosis studies
The way that patients are assembled for inclusion in cohort studies can influence the results. Prognostic studies are most useful if the patients have been followed up from a similar and well-described point in their illness. The term inception cohort has been used to describe a group of people who are assembled near the onset (‘inception’) of their disease. However it is not always necessary to follow individuals from an early stage, the important point is that they are followed from a uniform point in their disease. If this is not well described then it is difficult to interpret the results. A study by Hubbard et al illustrates the importance of defining at what time point in their disease patients have been included in prognostic studies. In this cohort study examining survival in patients with cryptogenic fibrosing alveolitis (idiopathic pulmonary fibrosis), the researchers found that the survival of patients already attending clinics (prevalent cases) was better than that in individuals who were newly diagnosed (incident cases). The median survival of incident cases was 2.9 years, compared with 9 years for prevalent cases. It is likely that the prevalent group excludes patients with aggressive disease that have died early in the course of their illness (survival bias).

Because descriptions of the course of disease must be based on samples of patients, they may be susceptible to sampling bias. A common theme for prognostic studies across different diseases is that the prognosis of individuals who attend tertiary referral centres is often worse than that found in community based studies. For example the 5-year survival after a diagnosis of dilated cardiomyopathy was found to be 36% in patients referred to a specialist centre compared with 80% in a community based cohort. There are several reasons that prognosis may differ between community based studies and those based on individuals attending tertiary referral centres. Firstly patients with more severe disease are more likely to be referred for expert opinion and management. Secondly experts may preferentially investigate, manage and follow up individuals that are thought to be more severe or more challenging. Finally access to diagnostic and treatment services may be linked with sociodemographic and geographic factors and in some circumstances these factors may themselves be markers of disease prognosis.
For these reasons the sample of individuals included in the study should be well described. Details about how patients are selected for the study and from what sources are important. For clinicians interpreting prognostic studies the samples should be representative of the types of patients seen in their practice if the results are going to be helpful to them. If the sample is from a tertiary centre the referral process should be described. The description should include the source of patients, inclusion or exclusion criteria and what objective criteria were used to confirm the disease. Note that when examining the findings of prognostic studies you need enough information to be able to decide if the sample of patients included in the study is likely to differ systematically from those you see in practice or wish to apply the results to. A study based on patients from a tertiary centre may be of limited value to a general practitioner managing patients outside of this setting, but the results would be helpful to specialists working in similar tertiary settings.

Follow up

The results of prognostic studies are likely to be more useful if losses to follow up are minimal and if the duration of follow up is sufficient. The duration of follow up will depend on the disease being examined but for some disease it may need to be quite long to detect all-important outcomes. For example to detect cases of cirrhosis following infection with hepatitis C a follow up period of 15 years or more may be required\(^7\). If follow up is incomplete this may threaten the validity of the results. Bias may occur if those that are lost to follow up are systematically different from those in who follow up is complete, and especially if these differences are related to prognosis. For example patients who are more severely affected or disabled by their illness may be more difficult to contact or may fail to attend for follow up review.

Where individuals are followed up by surveys, non-response bias may influence the results. Non-response bias cannot always be excluded by showing that baseline characteristics are equally distributed among responders and non-responders. For example in a cohort study of cement factory workers, non-responders to a follow up survey were subsequently found to have twice the rate of hospital admission for respiratory diseases, after controlling for factors such as smoking and age. However, responders and non-responders did not differ on the basis of smoking habits, respiratory symptoms and lung function\(^7\).

The potential impact of losses to follow up on the study results can be examined by recalculating the results taking into account the losses and assuming a best and a worst-case scenario (sometimes called a sensitivity analysis). For example, in a study of the natural history of acute back pain in patients attending primary care, the risk of having chronic low back pain at 3 months was found to be approximately 2%\(^8\). This study initially enrolled 103 patients but complete follow up data was only available on 92 patients (11 drop outs). 2 patients developed chronic low back pain. If we assume that all those patients lost to follow up developed chronic pain (worst case scenario) then the actual prevalence of chronic pain would be \(11+2/103 = 23\%\). If we assume that all those lost to follow up did not develop chronic pain (best case scenario) then the prevalence would be \(2/103 = 2\%\). Thus the true rate of chronic pain could range between 2 and 23%. The clinical implications of these disparate figures are clearly quite different and therefore we need to be cautious about interpreting these results.

The effect of losses to follow up are usually greater where the risk of the prognostic outcome is low as demonstrated in the back pain example\(^8\). The clinical implications may not be as substantial when the prognostic outcome is more common. For example, in a long term follow up study of children with a history of wheezing or asthma before the age of 7, 27% (238 subjects) of those surveyed at age 33 reported wheezes during the previous 12 months\(^9\). The original sample consisted of 1046 children but follow up data was only available on 880 (166 losses to follow up). We can recalculate the proportion with wheeze at age 33, using the approach above. Firstly, we assume that all those lost to follow up do experience wheeze at age 33, in this case the prevalence would be \(238+166/1046 = 39\%\). If we assume that all those lost to follow up did not have
persistent wheeze at age 33, then the prevalence would be 238/1046 = 23%. Across this range (23% to 39%) the clinical implications are similar, that is, a reasonable proportion of patients with childhood wheeze or asthma have persistent symptoms in adulthood.

**Outcome measurement**

Studies of prognosis may examine different sorts of outcomes. Some studies will report diagnostic test results, for example, as ‘surrogate markers’. It is important however that such studies also consider outcomes that are important to both patients and clinicians. Such outcomes might include duration or severity of symptoms, death, disease recurrence/relapse, functional status (for example return to work) or quality of life.

As with other types of studies, outcome assessment should be conducted in a standardised fashion. In some cases blinding of observers may be necessary to avoid measurement bias. Blinding of observers may be particularly important when subjective outcomes are being assessed, such as pain, and prognostic factors are being examined. In this case blinding to prognostic factors may be necessary. Outcomes should be clearly defined, and explicit criteria used.

**Adjustment for important prognostic factors**

Often prognostic studies will compare outcome rates amongst patient groups with different characteristics. For example age is often a prognostic factor in disease and outcomes in older populations are sometimes compared with outcomes in younger patients. For many diseases, multiple factors may influence the disease outcomes and the associations amongst these different variables may be complex\(^1\). Where there are several factors that appear to be of prognostic significance, it is important to consider what the particular relationships between these variables might be, in particular:

- potential prognostic variables may be related to each other as well as the outcome of interest, in this case one of the variables may merely be a ‘marker’ for the other prognostic factor, but not predictive of prognosis independently (confounding)
- two or more variables might interact together so that their combined effect is larger than the sum of their individual effects

Multivariate analysis is often used to simultaneously evaluate the effect of multiple variables on outcomes. With this method mathematical models are used to relate independent variables to outcome. Different methods (such as multiple linear or logistic regression) are in common use and some assumptions are necessary. The details are beyond the scope of this course and it would be unreasonable to expect that the average reader of medical studies should be able to determine if appropriate statistical methods have been used for such analyses. It is important, however, that readers consider whether the researchers have taken into account all the potentially important prognostic factors before drawing conclusions about the relationship between certain variables and outcomes. Multivariate analysis can be used to simultaneously adjust for the effects of several variables so that the independent effects of one can be determined. For example in a study of prognosis following first myocardial infarction, Marrugat et al found that women have a higher 28-day case fatality and 6 month morbidity and mortality than men\(^10\). They also found however that the women in the study were on average older and more often had certain co-morbidities than the males in the study. The relationship between gender and outcome however remained significant even after adjustment for age and co-morbidity.

Sometimes multivariate statistics are used to develop predictive models. In this case a number of factors including demographic and clinical variables might be combined to make predictions about a particular prognosis. Models based on a single patient sample may however fail to distinguish between important prognostic variables and unimportant idiosyncrasies related to that particular
Sample. Such models should be validated by assessing the predictive power of the model on a second independent sample of patients.

**Examining the results of prognostic studies**

Prognostic outcomes are often expressed as a rate. For example the 5-year survival rate is the proportion of patients surviving 5 years from some point in the course of their disease. To be able to interpret a rate we need to know at what point in their disease patients entered the study, how long they were followed up for, how outcome events were defined and who the people at risk were.

Summarising results as a rate at a single time point does not convey any information about how the chance of survival changes over time. Rates convey simple information but conducting a survival analysis and displaying survival curves provides more information about the survival probabilities with time. Survival curves may be used to examine ‘time to event’ data other than survival, including, but not limited to; disease recurrence, functional recovery following illness, or time to conception in patients with poor fertility. Figure 7.2 shows two hypothetical survival curves for 2 different diseases, the probability of surviving is displayed on the vertical axis and the horizontal axis is the period of time from the beginning of the observation. In both diseases the 5-year survival is approximately 10%. In disease A the majority of patients die within the first 6 months and thereafter the risk of dying improves. In disease B the probability of survival in the first 6 to 12 months is greater than disease A.

A detailed description of survival analysis is beyond the scope of this course but the following is a brief overview that may help with the interpretation of survival curves. Using the Kaplan-Meier method, survival probabilities are calculated for a series of discrete time intervals. The time intervals used will depend on the particular study but often short intervals are used such as days. For each time interval the probability that those who have survived to the beginning will survive to the end is calculated. This is termed conditional probability, that is, the probability of being a survivor at the end of the interval on condition that the subject was a survivor at the beginning of the interval. For each time period the probability of surviving is based on the number of individuals at risk for that period. Individuals who have already died or have been censored are not at risk and are not used to estimate survival for that time period. Patients may be censored if they have dropped out, died of another cause or have not yet been followed up to that point in time. Survival curves are used to display the survival probabilities over time. Survival curves may have a ‘step’ appearance because sudden changes in the estimated probability occur when an event has been observed. For many of the time intervals there will be no outcome events recorded and the probability of survival will be unchanged over that period until the next event occurs. Standard errors and confidence intervals can be calculated for survival probabilities at different time periods.

*Some points about interpreting survival curves*

The number of observations decreases over time (as the proportion of individuals at risk diminishes) and therefore the estimates of survival probability at the tail of the survival curve may be less precise (with wider confidence intervals) than the estimates from the left hand side of the curve. Similarly some survival curves give the impression of a plateau effect at the tail end, however the smaller slope of the curve at this end may be due to the fact that survival rates are based on a diminishing number of individuals.
7.3 Worked example

Recall the example discussed in the introduction to this section. The following article seems to address the clinical question:


Abstract

**Objective:** To estimate the frequency, duration, and clinical importance of postherpetic neuralgia after a single episode of herpes zoster.

**Design:** Prospective cohort study with long-term follow up.

**Setting:** Primary healthcare in Iceland.

**Participants:** 421 patients with a single episode of herpes zoster.
Main outcome measures: Age and sex distribution of patients with herpes zoster, point prevalence of postherpetic neuralgia, and severity of pain at 1, 3, 6, and 12 months and up to 7.6 years after the outbreak of zoster.

Results: Among patients younger than 60 years, the risk of postherpetic neuralgia three months after the start of the zoster rash was 1.8% (95% confidence interval 0.59% to 4.18%) and pain was mild in all cases. In patients 60 years and older, the risk of postherpetic neuralgia increased but the pain was usually mild or moderate. After three months severe pain was recorded in two patients older than 60 years (1.7%, 2.14% to 6.15%). After 12 months no patient reported severe pain and 14 patients (3.3%) had mild or moderate pain. Seven of these became pain free within two to seven years, and five reported mild pain and one moderate pain after 7.6 years of follow up. Sex was not a predictor of postherpetic neuralgia. Possible immunomodulating comorbidity (such as malignancy, systemic steroid use, diabetes) was present in 17 patients.

Conclusions: The probability of longstanding pain of clinical importance after herpes zoster is low in an unselected population of primary care patients essentially untreated with antiviral drugs.

The full text of this article can be downloaded from the BMJ website http://www.bmj.com. Click on Search/archive. Under Author enter ‘Helgason, S’ and click Search.

In the following section, this article has been appraised using the worksheet provided at the end of this section.
# Prognosis Worksheet

**Citation**


## Are the results of this prognosis study valid?

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Was a defined, representative sample of patients assembled at a similar point in the course of their disease?</strong></td>
<td>The researchers have assembled an inception cohort for this study. (Patients with a first episode of herpes zoster). Although not explicitly stated in the article, presumably the patients were referred to the study after their first presentation to their GP and once the clinical diagnosis was established. (Ideally factors such as the duration of symptoms prior to enrolment in the study should have been described.) The clinical criteria used by GPs to diagnose herpes zoster were not described. The sample included in the study would seem to be representative of patients attending general practice in Iceland. 62 selected GPs were used including both rural and urban practices. In addition, the completeness of enrolment was assessed by examining the records of patients attending hospitals covering the same catchment area. Only GPs with a computerised record system were included which may not be representative of all GPs however the total sample is drawn from a large catchment area in Iceland representing more than a third of the population of Iceland. Significant sampling bias is unlikely.</td>
</tr>
<tr>
<td><strong>Was patient follow-up sufficiently complete?</strong></td>
<td>Follow up rates at 1 month, 3 months and 12 months were 81%, 93% and 99%. The poorer follow up 1 and 3 months could affect the prevalence figures for pain at these time points. Losses to follow up by age categories are not described and therefore it is difficult to recalculate a best and worst case scenario. However recall that where the prevalence rates are low (categories of moderate or severe pain), small losses to follow up could substantially affect the results. It is however reassuring that the follow up at 12 months was almost complete and we can be reassured that the prevalence of protracted post herpetic neuralgia is low.</td>
</tr>
<tr>
<td><strong>Were objective and unbiased outcome criteria applied?</strong></td>
<td>All patients were contacted and surveyed by a single researcher using a systematic set of questions. A simple scale was used to assess pain (none, mild, moderate or severe). They do not state whether this is a validated tool, but no standardised questionnaires (eg such as</td>
</tr>
</tbody>
</table>
quality of life) were used. Ideally a standardised questionnaire, which has been validated, would be useful. For subjective outcomes such as pain, the way that questions are phrased, for example, may influence the results. In this study we are not able to examine such factors, because a standard questionnaire reported elsewhere in the literature is not available. There is no mention of whether the investigator was blinded to prognostic factors such as age when undertaking the follow up.

| If subgroups with different prognoses are identified, was adjustment for important prognostic factors undertaken? | The only possible prognostic factors considered in the analysis were sex and age. Gender was not found to be a significant predictor of outcome. Age was a significant predictor of pain at each of the follow up points. The odds ratio per 10 years age difference was 1.87 (95% CI: 1.56 to 2.23) at 1 month and 2.45 (1.5 to 4.01) after 6 months. |

Are the valid results of this prognosis study important?

| How likely are the outcomes over time? | In patients aged 50 to 59, the prevalence of mild pain at 1 month was 16% and the prevalence of moderate pain at 1 month was 7%. At 3 and 12 months the prevalence of mild pain was 4% and moderate pain 0%. |
| How precise are the estimates of likelihood? | The numbers of patients experiencing pain are relatively small and the estimates of prevalence are therefore somewhat imprecise (wide confidence intervals). For mild pain at 1 month the confidence intervals are 6.7% to 30%. For mild pain at 12 months the confidence intervals are narrower, 0.5% to 12.8%. |

Can you apply this valid, important evidence about prognosis in caring for your patients?

| Were the study patients and their management similar to those in my practice? | Patients included in the study appear to be similar to your own, 54 patients in the age range 50 to 59 were included. Because the clinical criteria used to diagnose herpes zoster are not well defined it is difficult to assess whether your patient is similar on clinical grounds, however your patient’s presentation is fairly typical for herpes zoster and is likely to be similar to those included in the study. Of note cultural differences may be important in extrapolating the results of this study to your patients. Pain thresholds have been noted to vary across cultures. This study was limited to patients in Iceland. |
| Was follow up sufficiently long? | 12 months follow up is sufficiently long to capture cases of post herpetic neuralgia. In addition a sub-sample were followed for a longer period (up to 7.6 years). |
Can I use the results in the management of patients in my practice?

Yes, you are happy that, based on the results of this study, the likelihood of your patient developing severe or protracted post herpetic neuralgia is very low. You are aware however, that when this complication does occur it can be very debilitating. You can use the information from this study to discuss the risks of post herpetic neuralgia and the treatment implications.

In the table below calculations for the number needed to treat to prevent pain at 6 months, based on a study examining the efficacy of Famciclovir are presented\textsuperscript{12,13}. In this study however, the prevalence of pain at 6 months in the placebo group (23.8\%) was greater than that reported from the population based study by Helgason et al. The number needed to treat has been recalculated using prevalence data from Helgason et al and assuming the relative risk reduction remains the same. Using data from the Famciclovir study suggests that 11 patients need to treated to prevent 1 patient having persistent pain at 6 months. However when the prevalence rates from a population-based study are used we find that 68 patients need to be treated to prevent 1 patient from having persistent pain at 6 months.

You can use this information to explain the potential benefits of treatment to your patient. You also explain the cost of treatment and the possible side effects (which are relatively minor with Famciclovir).

Other antivirals (eg acyclovir) appear to have some efficacy in the treatment of herpes zoster, but the benefit is less clearly defined. Other studies also suggest that tricyclic antidepressants administered early in the course of herpes zoster infection may help to reduce the risk of post herpetic neuralgia. For those interested, the systematic review by Alper et al is very comprehensive\textsuperscript{13}. 

Note
Pain at 6 months following presentation with herpes zoster, comparison of the NNT from a treatment study with famciclovir with the NNT calculated based on a cohort of patients followed up from general practice.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Control event rate (X)</th>
<th>Experiment event rate (Y)</th>
<th>Relative risk reduction RRR = (X –Y)/X</th>
<th>Absolute risk reduction ARR = X – Y</th>
<th>Number needed to treat (NNT = 1/ARR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculations based on Famciclovir study (12) (13)</td>
<td>23.8%</td>
<td>15%</td>
<td>37%</td>
<td>0.088 (8.8%)</td>
<td>11</td>
</tr>
<tr>
<td>Number needed to treat based on patients aged 50 to 59, using prevalence data from Helgason et al. (14)</td>
<td>4%</td>
<td>2.5%</td>
<td>37%</td>
<td>0.015 (1.5%)</td>
<td>68</td>
</tr>
</tbody>
</table>

7.4 Self-complete exercise

A patient of yours (Mr AS) who has just been to see his cardiologist calls you. Mr. AS is a 68-year-old man who has recently been diagnosed with aortic stenosis. Because he is asymptomatic his cardiologist has recommended that surgery be delayed until symptoms develop, however Mr AS is worried about the risk of dying if he does not have surgery in the near future and would like to discuss this with you further.

Aortic stenosis was diagnosed after you noticed a loud ejection systolic murmur during an examination several months ago. Subsequently you referred Mr AS for echocardiography and doppler examination, which confirmed the presence of severe aortic stenosis with a peak aortic-jet velocity of 4.8m per second. The aortic valve was moderately calcified. Trivial aortic regurgitation was noted but no other valvular lesions were noted. Left ventricular size and function was normal. Currently Mr AS is well and has no symptoms related to the aortic stenosis. He is an ex smoker, with a history of hypercholesterolemia currently controlled with diet. He has no other relevant past history.

You arrange to see Mr AS the following day and decide to read up on the prognosis of patients with asymptomatic severe aortic stenosis. You conduct a MEDLINE search in ‘PubMed’, by typing ‘aortic stenosis’ into the MeSH browser search screen (see 3.2), and find that the MeSH term is ‘aortic valve stenosis’. You search ‘PubMed’ using the subheading ‘mortality’ and retrieve ~900 citations, too many to peruse. You inspect the MeSH tree in the MeSH browser and elect not to ‘explode’ the MeSH term to include ‘supravalvular’ and ‘subvalvular’ stenoses, and also to restrict the search to ‘majored’ citations, by ticking the box ‘Restrict Search to Major Topic headings only’ (see 3.2). This leads to ~ 60 citations, which you peruse and find the following relevant publication:
Abstract

**Background:** Whether to perform valve replacement in patients with asymptomatic but severe aortic stenosis is controversial. Therefore, we studied the natural history of this condition to identify predictors of outcome.

**Methods:** During 1994, we identified 128 consecutive patients with asymptomatic, severe aortic stenosis (59 women and 69 men; mean [±SD] age, 60±18 years; aortic-jet velocity, 5.0±0.6 m per second). The patients were prospectively followed until 1998.

**Results:** Follow-up information was available for 126 patients (98 percent) for a mean of 22±18 months. Event-free survival, with the end point defined as death (8 patients) or valve replacement necessitated by the development of symptoms (59 patients), was 67±5 percent at one year, 56±5 percent at two years, and 33±5 percent at four years. Five of the six deaths from cardiac disease were preceded by symptoms. According to multivariate analysis, only the extent of aortic-valve calcification was an independent predictor of outcome, whereas age, sex, and the presence or absence of coronary artery disease, hypertension, diabetes, and hypercholesterolemia were not. Event-free survival for patients with no or mild valvular calcification was 92±5 percent at one year, 84±8 percent at two years, and 75±9 percent at four years, as compared with 60±6 percent, 47±6 percent, and 20±5 percent, respectively, for those with moderate or severe calcification. The rate of progression of stenosis, as reflected by the aortic-jet velocity, was significantly higher in patients who had cardiac events (0.45±0.38 m per second per year) than those who did not have cardiac events (0.14±0.18 m per second per year, P<0.001), and the rate of progression of stenosis provided useful prognostic information. Of the patients with moderately or severely calcified aortic valves whose aortic-jet velocity increased by 0.3 m per second or more within one year, 79 percent underwent surgery or died within two years of the observed increase.

**Conclusions:** In asymptomatic patients with aortic stenosis, it appears to be relatively safe to delay surgery until symptoms develop. However, outcomes vary widely. The presence of moderate or severe valvular calcification, together with a rapid increase in aortic-jet velocity, identifies patients with a very poor prognosis. These patients should be considered for early valve replacement rather than have surgery delayed until symptoms develop. (N Engl J Med 2000; 343:611-617.)

A full text copy of this article can be obtained by accessing full text journals online though the University Library, or through the Clinicians Health Channel or your hospital library.
### 7.5 Answers to self-complete exercises

#### Prognosis Worksheet

##### Citation


#### Are the results of this prognosis study valid?

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was a defined, representative sample of patients assembled at a similar point in the course of their disease?</td>
<td>Patients appear to have been assembled at a similar point in the course of their disease, that is they all have asymptomatic aortic stenosis at entry to the study. Only patients who were referred to an echocardiography laboratory were studied. The referral patterns were not described but presumably this is a tertiary referral centre or major hospital (Vienna General Hospital). The study will clearly not be representative of all patients with severe asymptomatic aortic stenosis, since many patients in the community will remain undiagnosed or may be diagnosed but not referred for echocardiography at such centres. However the patient sample probably is representative of patients referred to or managed in the tertiary setting. As the authors point out, one of the main limitations to the study is the fact that some of the patients underwent valve replacement shortly after diagnosis despite being asymptomatic. They could not be used to examine the natural history of severe asymptomatic aortic stenosis. If this group of patients differ to those who did not undergo surgery then the possibility of bias exists. Importantly the group of patients undergoing surgery early may represent more severe cases, and the risk of sudden death, for example, may be underestimated. These patients do appear to differ on the basis of age and aortic jet velocity at baseline and this is clearly a potential source of bias. However the study represents current practice, that is that some patients with asymptomatic severe aortic stenosis will be referred for aortic valve replacement at the discretion of their treating physician. Other studies have noted similar circumstances.</td>
</tr>
<tr>
<td>Was patient follow-up sufficiently complete?</td>
<td>Out of 128 patients only 2 were lost to follow up. These losses to follow up are unlikely to substantially affect the results of the study.</td>
</tr>
<tr>
<td>Were objective and unbiased outcome criteria applied?</td>
<td>The outcomes were death or aortic-valve replacement necessitated by the development of symptoms. Cause</td>
</tr>
</tbody>
</table>
of death was also examined. There is no mention that outcome assessors were blind to prognostic factors being considered, however the outcomes are fairly objective. Cause of death can be open to interpretation and different methods were used to determine the cause of death in different patients. However the number of deaths is relatively small, most were due to cardiac causes and it is unlikely that the risk of cardiac death has been underestimated.

The outcome measure of aortic valve replacement necessitated by symptoms is somewhat problematic in relation to the examination of prognostic variables. Clinicians may take into account prognostic variables identified in this study such as aortic valve calcification, or changes in aortic jet velocity over time when making decisions to refer patients for valve replacement. Although these patients were classified as having symptoms, this is open to interpretation by the clinician, for example a clinician may be more likely to refer a patient with relatively mild exertional dyspnoea for valve replacement if they are concerned about other markers of disease severity such as an increase in aortic jet velocity over time.

If subgroups with different prognoses are identified, was adjustment for important prognostic factors undertaken?

In examining prognostic factors the researchers examined multiple variables, including clinical variables (age, sex, and presence/absence of coronary artery disease, hypertension, diabetes, and hypercholesterolemia) and echocardiographic variables (degree of aortic-valve calcification and aortic-jet velocity). Change in aortic jet velocity over time was also examined.

The extent of aortic-valve calcification was a strong predictor of subsequent events. Although age, coronary artery disease and diabetes were predictors of outcome in the univariate analyses, aortic-valve calcification was the only prognostic factor that remained significant after adjustment for other possible prognostic factors in the multivariate analysis.

The researchers appear to have considered most of the important possible prognostic variables in their analysis.
### Are the valid results of this prognosis study important?

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely are the outcomes over time?</td>
<td>During the follow up period (mean 22 months) there were 8 deaths and 59 valve replacements due to the development of symptoms. 6 of the 8 deaths were due to cardiac causes and these were all presumed to be related to aortic stenosis. However apart from 1 sudden death they were all preceded by the development of symptoms (aortic valve replacement was not performed in these patients for various reasons). Overall survival was 87% at four years. Event free survival was 33 percent at 4 years. For patients with moderate or severe aortic calcification event free survival was 20% at 4 years. All deaths were noted amongst patients with moderate or severe valvular calcification.</td>
</tr>
<tr>
<td>How precise are the estimates of likelihood?</td>
<td>Standard errors and standard deviations were provided rather than confidence intervals.</td>
</tr>
</tbody>
</table>

### Can you apply this valid, important evidence about prognosis in caring for your patient?

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were the study patients and their management similar to those in my practice?</td>
<td>Your patient is similar to those included in the study on the basis of severity of AS and demographic factors. Your patient has also recently been assessed and managed in a tertiary referral setting.</td>
</tr>
<tr>
<td>Was follow up sufficiently long?</td>
<td>Yes, a large number of outcome events have occurred during the follow up period.</td>
</tr>
<tr>
<td>Can I use the results in the management of patients in my practice?</td>
<td>Yes, although you are concerned about some of the limitations of this study, you feel this study represents the best evidence on which to base advice for your patient. In summary the article suggests that amongst those patients who are first diagnosed with asymptomatic aortic stenosis and are not referred for valve replacement shortly after diagnosis, the long-term prognosis in terms of the risk of sudden death is good. A large proportion of patients end up requiring aortic valve replacement because of the development of symptoms however. You reassure your patient that the risk of sudden death is less than 1 percent per year. The immediate perioperative risk associated with isolated aortic valve replacement is around 3.5%(^{15}) (with higher rates for older populations or in the presence of co-morbidities) and there are also long-term complications associated with prosthetic valves that need to be considered(^{16}). You advise your patient to have regular review with his cardiologist as planned and tell him to report the</td>
</tr>
</tbody>
</table>
following symptoms to you immediately if they develop; shortness of breath on exertion, chest pain or discomfort, particularly on exertion and loss of consciousness or light-headedness on exertion.

7.6 Suggested reading


7.7 Further reading


- ‘EPIQ’ (Effective Practice, Informatics & Quality Improvement) is produced by Professor Rod Jackson (University of Auckland). It has very useful materials including notes and tools. http://www.health.auckland.ac.nz/comhealth/epiq/epiq.htm


7.8 References


7.9 Critical appraisal worksheets

Example adapted from Centre for Evidence-Based Medicine, Mount Sinai Hospital
http://www.cebm.utoronto.ca/teach/materials/therapy.htm

Prognosis worksheet

<table>
<thead>
<tr>
<th>Citation</th>
</tr>
</thead>
</table>

Are the results of this prognosis study valid?

<table>
<thead>
<tr>
<th>Was a defined, representative sample of patients assembled at a similar point in the course of their disease?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Was patient follow-up sufficiently complete?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Were objective and unbiased outcome criteria applied?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>If subgroups with different prognoses are identified, was adjustment for important prognostic factors undertaken?</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Are the valid results of this prognosis study important?</td>
</tr>
<tr>
<td>How likely are the outcomes over time?</td>
</tr>
<tr>
<td>How precise are the estimates of likelihood?</td>
</tr>
<tr>
<td>Can you apply this valid, important evidence about prognosis in caring for your patient?</td>
</tr>
<tr>
<td>Were the study patients and their management similar to those in my practice?</td>
</tr>
<tr>
<td>Was follow up sufficiently long?</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Can I use the results in the management of patients in my practice?</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Additional Notes**
8. **Helping Patients (‘Consumers’) to Make Informed Decisions Which Integrate Their Values With Clinical Evidence**

### 8.1 Learning objectives

- **To further understand how evidence from studies (eg therapy, diagnosis, prognosis, cause and risk) applies to individual patients.**
- **To recognise the complexity of individuals’ perceptions of the risks and benefits of interventions, and how the presentation of data from clinical studies affects its interpretation.**
- **To appreciate different models for clinician-patient interactions and decision-making by patients.**
- **To assist patients seeking answers to questions, about prognosis and cause, and about the risks and benefits of treatments and tests.**

### 8.2 Core reading, key concepts, and worked examples

#### Introduction

Patients have two major concerns about applying the results of clinical evidence derived from observational and interventional studies to themselves:

- With the exceptions of ‘n–of–1’ trials, where the patients themselves are specifically included in control and treatment phases, studies almost always involve other people, and are in different settings. Therefore the results may not apply to them.

- Results are expressed as ‘averages’ with ‘ranges’ for the groups studied. Patients see themselves as individuals, possibly different from the ‘average’, and would like to have more specific information about results for themselves.

Perceptions about health issues and choices about lifestyle and interventions are determined by complex experiential and sociological phenomena. Patients also know from their experiences as ‘consumers’ of products and services unrelated to health, that advertising, packaging and marketing can influence both the perceived value of products and their uptake, and recognise that the same applies to health information.

Clinicians are often requested by patients to assist them to make decisions affecting their health from the various options available. The extent of the request varies from virtually no assistance at all, to complete dependence on the clinician’s advice. A frequent request is to satisfy patients’ desires for additional information.
8.2.1 Applicability

You will recall from 'Ask the answerable question' (see also 2.2) that all clinical evidence requires specification of the characteristics of persons being studied as well as details of outcomes being assessed. With interventional studies the precise nature of the intervention is also required.

Applicability refers to the translation of results from the group studied to other persons, often in different settings. One question is whether an individual's characteristics are so different from the group studied that the results no longer apply. Even if your patient's characteristics closely resembles those of the group studied, another question is the relevance of the study in the patient's own setting.

Although there are sound principles on which to address these questions, the answers reflect judgements and opinions, which are based on knowledge, wisdom and experience. Factors, which affect the applicability of clinical studies to a particular patient, can be classified into 4 groups, although there is considerable overlap.

Patient factors eg age; gender; genes (including race); comorbid conditions; ingested substances (including medications); individual probability of adverse and beneficial disease outcomes; individual risk of treatments and investigations; compliance.

Disease factors eg clinical presentation; duration of disease; severity of disease; presence and type of symptoms.

Environmental factors eg social, geographic, ecological and economic settings; definitions of diseases and of outcomes.

Interventional factors eg the nature and consistency of investigations and treatments. (see also 4.2 “external validity and generalisability”).

Example 1: Patients with the characteristics of your patient have not yet been studied.

Your patient is a 60-year old heavy smoking diabetic with hypertension and hyperlipidemia. Using the New Zealand cardiovascular risk calculator (http://www.nps.org.au/docs/pdfs/cardiovascularrisk.pdf) you estimate that he has a 25% risk of developing coronary heart disease over the next 10 years. He is keen to quit smoking but past attempts have failed.

Which medications might help him to quit smoking?

A search of ‘Clinical Evidence’ (available from the ‘Clinicians Health Channel' through teaching hospital computers) indicates that Bupropion is “likely to be beneficial”, but that there are no studies in diabetic patients. Rather than deciding that the data is inapplicable, you and the patient ask if there is any reason that these data might not apply. You decide that there is no reason and proceed to recommend Bupropion.

Although the effectiveness of this medication in diabetics is unknown, it is most unlikely that people with diabetes would respond differently than non-diabetics. In circumstances like this it is worth considering the question: “Is there any reason that the results of this study should not apply to this patient?”

Example 2: Spectrum bias in diagnostic tests (see also 6.3)
The sensitivity, specificity and likelihood ratios (see also 6.2) of tests are not always constant across the clinical spectrum, but may vary according to a range of factors including age; gender; concurrent medications; the presence or absence of symptoms; duration of illness; severity of illness.

Your patient is a 19-year old Arts student presenting with sore throat, fever and generalized lymphadenopathy for 5 days. As the probability of HIV infection is minimal, you wonder about infectious mononucleosis and request Ebstein-Barr virus serology. The test result is negative then, but positive when repeated 2 weeks later.

Why is the antibody test negative then, but positive 2 weeks later?

The test result is negative at that time, yet positive 2 weeks later, because the production of antibody production takes time. Thus ‘duration of illness’ is one of the factors that influence the sensitivity, specificity and likelihood ratio of a diagnostic test.

Example 3: Subgroup analysis.

In some observational studies and clinical trials, it may be possible to refine applicability by studying smaller ‘subgroups’, with clinical characteristics more closely resembling your patient.

There are two major pitfalls of subgroup analyses. The reduction in size of the cohort may result in false-negative conclusions, as confidence intervals (see also 4.2) become wider. Furthermore, if the subgroups are not defined prospectively on the basis of scientific rationale (eg clinical knowledge, pathophysiology etc), retrospective ‘data dredging’ may result in irrational false-positive conclusions. An inappropriate subgroup analysis showed fewer deaths for stroke patients when treatment was chosen according to the throw of coloured dice for some, but not all colours1! A structured approach to subgroup analysis helps distinguish between artifactual and real differences in outcomes of clinical trials between subgroups2.

Example 4: Individualising risk and prognostic data.

Analysis of subgroups, within both cohort studies and the placebo arm of intervention studies, serves to increase the precision of absolute and relative assessments for the risk of developing diseases and for outcomes of diseases (see also 7.2). Case control studies enable more precise estimations of relative but not of absolute risk (see also 4.2).

Making individual assessments of the risk of developing diseases and their outcomes is not only important for information on prognosis, but helps patients to choose interventions which are most likely to improve their outcomes. There are many examples of interventions in which relative risk reduction (see also 4.2) remains fairly constant over a broad range of risk. This means that the Number Needed to Treat (NNT) for patients with higher risk is smaller, and hence the benefits greater, for those at higher risk. In many circumstances, the risk of complications of treatment, in contrast, remains fairly constant across a broad range of risk for adverse disease outcomes. Thus individualization of both risk of adverse disease outcomes and the risk of interventions helps patients to make better decisions about potentially useful interventions.

Your patient is a 78-year old man with atrial fibrillation for 10 years. There is no suggestion of other cardiac or of general disease following thorough clinical evaluation. His blood pressure is 130/85 mmHg. You know that several randomised controlled trials have shown that anticoagulation with warfarin prevents stroke in patients with atrial fibrillation. Warfarin, although more effective than aspirin, carries greater risks.
Would you recommend aspirin or warfarin for him?

A literature search draws your attention to a publication which shows that across a range of risks, relative risk reduction of stroke is fairly constant, so absolute risk reduction is proportional to risk. You interpret this to mean that the Number Needed to Treat (NNT) to prevent one stroke is inversely proportional to risk. On the other hand, the risk of bleeding from warfarin is independent of the risk of stroke from atrial fibrillation. Two clinical prediction rules considering the stroke risk for aspirin-treated patient in trials comparing aspirin and warfarin to prevent stroke show that this otherwise healthy man has a relatively low risk of stroke. Thus you recommend aspirin rather than the more risky warfarin for him.

8.2.2 Perceptions and presentations

8.2.2.1 Perceptions of risks and benefits
In health-related, as in all other matters, each person perceives information about prognosis, the cause and risks of disease, and the pros and cons of life-style changes and the benefits and risks of diagnostic and therapeutic interventions quite differently. Determinants of an individual’s decisions, selected from the breadth of options available, relate to many influences. These include past experiences; family, cultural, religious and ethical values; education; numeracy; literacy; access and opportunities; media coverage; and the input of significant other people. Decisions may be seen as ‘right’ or ‘wrong’ when viewed from clinician, patient and other persons’ perspectives.

Depending on the model chosen for the clinician-patient interaction, clinicians sometimes impose their own values to modify patient choice. Most clinicians accept the responsibility to provide relevant information in a form, which the patient can fully comprehend, so that the patient can, to the extent they wish, make their own decisions.

Studies show that other determinants of the perceptions of the risks and benefits of interventions include:

- voluntary versus mandatory choices
- treatment choice versus behavioural change
- clinician estimate versus patient estimate of risk and benefit
- immediate versus longer-term outcomes
- novel rather than established phenomena and interventions
- selective denial of significant data

8.2.2.2 Presentation (‘framing’) of data
The manner in which data is presented on the risk, and prognosis of disease and on the benefits and risk of interventions, is a strong determinant of its interpretation, perception and uptake by clinicians and patients.

Example: Absolute and/or relative data?

- bigger numbers make more impact than smaller numbers
- ‘relative’ (ie. ‘proportion’ or % or ‘chance’ or ‘probability’) data are more difficult for clinicians and patients to understand than ‘absolute’ (ie. ‘natural frequency’) data.

Imagine that you are a 70-year old woman deciding whether to take combination oestrogen/progesterone ‘HRT’ (Hormone Replacement Therapy) to prevent osteoporotic fractures.
Your clinician tells you of a Randomised Controlled Trial of ‘HRT’ compared to placebo in women of your age, and offers you the choice of two treatments: the first will reduce your risk of hip or vertebral fracture by ⅓ (30%), while the second would give you a 1/400 (0.25%) chance of avoiding a fracture with ‘HRT’ over the next 5 years. Both statements are derived from the same data and both are correct. The reason for the apparent difference in the magnitude of the figures is that while the risk-reduction is relatively large, the absolute risk is fairly small, and so the absolute benefits of treatment are perceived as small.

Which manner of data presentation to the patient is likely to appeal to the manufacturers of ‘HRT’, i.e. most likely to increase sales?
Should the first or the second presentation of data (or both) be offered to the patient?

For most people a 30% risk reduction will be more persuasive than an absolute benefit of 0.25%. Thus manufacturers often present relative rather than absolute data. In order to facilitate patient understanding, both methods should be used. The expression of absolute benefit and risk as Number Needed to be Treated (NNT) for 1 person to be benefited or harmed (NNH) provides an easily understood option for clinicians, patients and others (see also 4.2).

Other examples: The following differences in the way data is presented have also been also shown to alter the perceptions of risk:
- mortality data in contrast to survival data
- qualitative in contrast to quantitative data
- survival curves in contrast to numerical data
- pictorial representations in contrast to text
- personalised data in contrast to group data
- ‘plain English’ text in contrast to technical jargon
- the order in which data is presented on benefits and risks
- ‘decision aids’ in contrast to verbal explanations
- the source of the data

8.2.3 Models for clinician-patient interactions

The next step in the practice of EBM, which follows asking questions, searching and appraising the literature, applying results to individuals, and presenting it in appropriate ways to patients, is to determine the particular model of clinician-patient interaction that the patient is seeking. Possibilities range from no clinician input at all, even for the provision of appraised evidence, to complete dependence on the clinician for the interpretation of evidence and the choice of interventions.

There are many determinants of such models. Failure to appreciate the differences in models and the range of possible determinants is a significant source of miscommunication and discontent in clinician-patient relationships. Broad determinants, which may be different at different times, even for the same patient-clinician pairs include:
- the particular clinician
- the particular patient
- the setting in which patient and clinician interact
- the acuity (real or perceived) of the problem, disease or illness

As for the perceptions of the particular risks of diseases and benefits of interventions, the selection of specific clinician-patient interaction models also reflects cultural and sociological influences. The choice of model cannot always be assumed, however, and patients and clinicians can usefully interact to establish that model which is best suited to the current situation.
A variety of models have been suggested. The first is called ‘paternalistic’, where the clinician makes decisions, the second, ‘informative’, where the clinician provides all relevant data, or access to such data. The third is called ‘interpretative’, the clinician providing selected data after making efforts to determine patients’ values. The fourth, ‘deliberative’, is where selected data are provided after the clinician has assisted the patient to define their own values. Frameworks for more ‘integrated decision making’ and for the teaching and learning of ‘shared decision-making’ are evolving.

8.2.4 Answering patients’ questions concerning prognosis and cause and the risks and benefits of treatments and tests

In spite of clinician adherence to the issues of applicability, patients’ perceptions, and concern for the manner of information presentation, additional questions often arise. These are often resolved by further discussion but the patient may want more information. The patient-clinician interaction model may also change again at this stage, and in either direction, towards greater patient autonomy or dependence.

For all patients, especially those choosing greater autonomy, many readily-accessible resources are now available. These can enhance clinician-patient concept-sharing and dialogue. Although the clinician cannot and should not guarantee the validity and applicability of information, even from ‘trusted’ resources, the data obtained often stimulates further discussion and better resolution of those issues which bother patients.

Consumers, unlike some clinicians, have very few opportunities to access decision support systems, although some ‘Synopses’, and many ‘Syntheses’ and primary ‘Studies’ are now available to consumers through resources like the Cochrane Library and ‘PubMed’. Most of this information is designed for clinicians, but increasing numbers of resources, specifically designed for consumers, are now becoming available. These more scientific and evidence-based resources supplement traditional consumer ‘background’ clinical data, derived from the experiences of others, especially family and friends. Text, visual and electronic material from numerous sources, ‘framed’ in various ways, still complements the verbal accounts of others, to shape information and attitudes. Evidence-based consumer-specific information, which can enhance ‘background’, and at times ‘foreground’ information, includes pamphlets and electronic material prepared by discipline-specific professional societies (eg Gastroenterological Society of Australia) and foundations (eg National Heart Foundation, Stroke Foundation). Several medical journals (eg Annals of Internal Medicine, JAMA, ‘UpToDate’), and the Cochrane Collaboration Consumer Network also publish and/or collate consumer-specific material.

Useful web addresses for both ‘background’ and ‘foreground’ data for patients seeking these are listed below. Some of these sites also contain worthwhile ‘links’ to other resources as well:

This is a Victorian DHS facility. Search in the window provided, but also check the ‘links’ for more Australian and International sites.

“Entrez-PubMed”
Search in ‘Google’ (http://www.google.com) for ‘Entrez-PubMed’. Then click on “consumer health related resources” in the lower-left third of the ‘Entrez-PubMed’ home page, and select a topic to search. This USA site gives access to much data for consumers, but not all of it may be valid.
http://www.healthinsite.gov.au
This is an Australian Commonwealth Government site.

http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/ForPatients.html
This provides specific information for patients about Cochrane products.

Cochrane Library
Free access from Australia is through the Australian National Institute of Clinical Studies (NICS) web page: http://www.nicsl.com.au. The NICS home page also includes instructions on how to search the Cochrane Library, which includes systematic reviews and randomised controlled trials.

8.3 Recommended reading

- Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor DC, Wilson MC, Richardson WS. Users’ guides to the medical literature XXV. Evidence-based medicine: Principles for applying the users’ guides to patient care. JAMA 2000; 284:1290-1296
- Komesaroff PA. Ethical perspectives on the communication of risk. Australian Prescriber 2003;26:44-45

8.4 Further reading

- Dans AL, Dans LF, Guyatt GH, Richardson R. Users’ guides to the medical literature XIV. How to decide on the applicability of clinical trial results to your patient. JAMA 1998; 279:545-549

• Lloyd AJ. The extent of patients’ understanding of the risks of treatments. Quality in Health care 2001; 10:i14-i18

• McAlister FA, Straus SE, Guyatt GH, Haynes RB. Users’ guides to the medical literature XX. Integrating research evidence with the care of the individual patient. JAMA 2000; 283:2829-2836


8.5 References


9. **CONTRIBUTORS AND ACKNOWLEDGEMENTS**

The first edition of this workbook was prepared in 2002 by A/Professor Don Campbell (Director), Dr Renee Manser (Research Fellow) and Dr Marg de Campo (Project Director), Clinical Epidemiology and Health Service Evaluation Unit, Royal Melbourne Hospital / Melbourne Health.

These authors were assisted through formal and informal discussion with the members of the Unit, by academics from the Faculty of Medicine Dentistry and Health Sciences, University of Melbourne, especially those from the School of Population Health and the Faculty Education Unit, and by clinicians from the teaching hospitals. The authors acknowledge discussions, input and illustrations provided by Professor Rod Jackson, Professor of Epidemiology, School of Population Health and University of Auckland.

The contributions of Professor Terry Nolan, A/Professor Geoff McColl, Drs Caroline Brand, Roger Brown, Shlomo Cohney, Jenny Conn, Peter Greenberg and Martin Wright, Mr Noel Atkinson and Mr Bruce Mann are noted. Many medical students also contributed to the development of the workbook. Ms Wendy Lemaire, Ms Lori McLeod and Ms Janice Powell provided invaluable support.

Editing and additions for 2003 and 2004 were made by Dr Peter Greenberg and Professor Jeremy Anderson, supported by Ms Rita Ward, Ms Janice Powell and Ms Katie Dumas. In 2005 and 2006, editing and additions to the first three chapters were made by Dr Peter Greenberg, assisted by Ms Rita Ward and Ms Katie Dumas, with input from Ms Elizabeth Deans and Dr Simon Chatfield. In 2007, editing and additions, especially to the first 3 chapters were made by Dr Peter Greenberg, with assistance from Ms Meg McKay and Ms Rita Ward.
10. **GLOSSARY OF TERMS**

For more detailed glossaries see: [http://www.cebm.utoronto.ca/glossary/](http://www.cebm.utoronto.ca/glossary/) and [http://www.cochrane.org/resources/glossary.htm](http://www.cochrane.org/resources/glossary.htm)

Terms used in therapeutics

**Allocation concealed**: deemed to have taken adequate measures to conceal allocation to study group assignments from those responsible for assessing patients for entry in the trial (e.g., central randomisation; sequentially numbered, opaque, sealed envelopes; sealed envelopes from a closed bag; numbered or coded bottles or containers; drugs prepared by the pharmacy; or other descriptions that contain elements convincing of concealment).

**Allocation not concealed**: deemed to have not taken adequate measures to conceal allocation to study group assignments from those responsible for assessing patients for entry in the trial (e.g., no concealment procedure was undertaken, sealed envelopes that were not opaque, or other descriptions that contain elements not convincing of concealment).

**Unclear allocation concealment**: the authors of the article did not report or provide us with a description of an allocation concealment approach that allowed for classification as concealed or not concealed.

**Blinded**: any or all of the clinician, patients, participants, outcome assessors, or statisticians were unaware of who received which study intervention. Those that are blinded are indicated in parentheses. If ‘initially’ is indicated (e.g., blinded [patients and outcome assessor initially]), the code was broken during the trial, for instance, because of adverse effects.

**Blinded (unclear)**: the authors did not report or provide with an indication of whom, if any one, was unaware of who received which study intervention.

**Unblinded**: all participants in the trial (clinicians, patients, participants, outcome assessors, and statisticians) were aware of who received which study intervention.

**RRR (relative risk reduction)**: the proportional reduction in the rates of bad events between experimental (experimental event rate [EER]) and control (control event rate [CER]) patient in the trial, calculated as \( \frac{\text{EER} - \text{CER}}{\text{CER}} \), and accompanied by a 95% confidence interval (CI).

**ARR (absolute risk reduction)**: the absolute arithmetic difference in event rates, \( \text{EER} - \text{CER} \).

**NNT (number needed to treat)**: the number of patients who need to be treated to prevent one additional bad outcome; calculated as \( 1/\text{AAR} \), rounded up to the next highest whole number, and accompanied by its 95% CI.

**RBI (relative benefit increase)**: the increase in the rates of good event, comparing experimental and control patients in a trial, also calculated as \( \frac{\text{EER} - \text{CER}}{\text{CER}} \).

**ABI (absolute benefit increase)**: the absolute arithmetic difference in event rates, \( \text{EER} - \text{CER} \).

**NNT**: calculated as \( 1/\text{ABI} \); denotes the number of patients who must receive the experimental treatment to create one additional improved outcome in comparison with the control treatment.
WHEN THE EXPERIMENTAL TREATMENT INCREASES THE PROBABILITY OF AN UNFAVOURABLE OUTCOME

**RRI (relative risk increase):** the increase in rates of bad events, comparing experimental patients to control patients in a trial, and calculated as for RBI. RRI is also used in assessing the effect of risk factors for disease.

**ARI (absolute risk increase):** the absolute difference in rates of bad events, when the experimental treatment harms more patients that the control treatment; calculated as for ABI.

**NNH (number needed to harm):** the number of patients who, if they received the experimental treatment, would lead to one additional person being harmed compared with patients who receive the control treatment; calculated as 1/ARI.

**Confidence interval (CI):** the CI quantifies the degree of precision or certainty in measurement, usually reported as 95% CI, which is the range of values within which we can be 95% sure that the true value for the whole population lies.

**Terms used in diagnosis**

**Sensitivity:** the proportion of patients with the target disorder who have a positive test result (a/[a + c]) (figure).

**Specificity:** the proportion of patients without the target disorder who have a negative test result (d/[b + d]) (figure).

**Pre-test probability (prevalence):** the proportion of patients who have the target disorder, as determined before the test is carried out ([(a + c)/([a + b + c + d])] (figure).

**Pre-test odds:** the odds that the patient has the target disorder before the test is carried out (pre-test probability /[1 – pre-test probability]).

**Likelihood ratio (LR):** the ratio of the probability of a test result among patients with the target disorder to the probability of that same test result among patients who are free of the target disorder. The LR for a positive test is calculated as sensitivity /(1 – specificity). The LR for negative test is calculated as (1 – sensitivity)/specificity.

**Post-test odds:** the odds that the patient has the target disorder after the test is carried out (pre-test odds x LR)

**Post-test probability:** the proportion of patients with that particular test result who have the target disorder (post-test odds /[1+ post-test odds]).