

Principles of evidence based medicine

A K Akobeng

Arch Dis Child 2005;90:837–840. doi: 10.1136/adc.2005.071761

Health care professionals are increasingly required to base clinical decisions on the best available evidence. Evidence based medicine (EBM) is a systematic approach to clinical problem solving which allows the integration of the best available research evidence with clinical expertise and patient values. This paper explains the concept of EBM and introduces the five step EBM model: formulation of answerable clinical questions; searching for evidence; critical appraisal; applicability of evidence; evaluation of performance. Subsequent articles will focus on the principles and critical appraisal of randomised controlled trials, systematic reviews, and meta-analyses, and provide a practical demonstration of the five step EBM model using a real life clinical scenario.

preferences, and should also incorporate expertise in performing clinical history and physical examination. Figure 1 illustrates a typical flow chart of EBM, depicting how knowledge and experience may be integrated with patients' preferences and available evidence in the making of clinical decisions.

WHY EVIDENCE BASED MEDICINE?

The most important reason for practising EBM is to improve quality of care through the identification and promotion of practices that work, and the elimination of those that are ineffective or harmful.⁴ EBM promotes critical thinking. It demands that the effectiveness of clinical interventions, the accuracy and precision of diagnostic tests, and the power of prognostic markers should be scrutinised and their usefulness proven. It requires clinicians to be open minded and look for and try new methods that are scientifically proven to be effective and to discard methods shown to be ineffective or harmful.

It is important that health care professionals develop key EBM skills including the ability to find, critically appraise, and incorporate sound scientific evidence into their own practice.

THE FIVE STEP EBM MODEL

The practice of EBM involves five essential steps^{3–5}: first, converting information needs into answerable questions; second, finding the best evidence with which to answer the questions; third, critically appraising the evidence for its validity and usefulness; fourth, applying the results of the appraisal into clinical practice; and fifth, evaluating performance.

Step 1: Formulating answerable clinical questions

One of the difficult steps in practising EBM may be the translation of a clinical problem into an answerable question.⁶ When we come across a patient with a particular problem, various questions may arise for which we would like answers. These questions are frequently unstructured and complex, and may not be clear in our minds. The practice of EBM should begin with a well formulated clinical question. This means that we should develop the skill to convert our information needs into answerable questions. Good clinical questions should be clear, directly focused on the problem at hand, and answerable by searching the medical literature.⁷

A useful framework for making clinical questions more focused and relevant has been suggested by Sackett *et al.*¹ They proposed that

WHAT IS EVIDENCE BASED MEDICINE?

The concept of evidence based medicine (EBM), defined as the "integration of best research evidence with clinical expertise and patient values",¹ has been gaining popularity in the past decade. The practice of EBM involves a process of lifelong self directed learning in which caring for patients creates the need for important information about clinical and other health care issues. EBM recognises that the research literature is constantly changing.² What the evidence points to as the best method of practice today may change next month or next year. The task of staying current, although never easy, is made much simpler by incorporating the tools of EBM such as the ability to track down and critically appraise evidence, and incorporate it into everyday clinical practice.

The work of people in the field of paediatrics and child health centres on the problems of children and their families and carers. Questions about diagnosis, prognosis, and treatment often arise and sometimes the answers to these questions need to be sought. EBM allows the integration of good quality published evidence with clinical expertise and the opinions and values of the patients and their families or carers. Deciding on how to treat patients should not be based solely on the available evidence. Other factors such as personal experience, judgement, skills, and more importantly patient values and preferences must be considered.

The practice of EBM should therefore aim to deliver optimal patient care through the integration of current best evidence and patient

Correspondence to:
Dr A K Akobeng,
Department of Paediatric
Gastroenterology, Central
Manchester and
Manchester Children's
University Hospitals, Booth
Hall Children's Hospital,
Charlestown Road,
Blackley, Manchester, M9
7AA, UK; tony.akobeng@
cmmc.nhs.uk

Accepted 22 April 2005

Abbreviations: EBM, evidence based medicine; CASP, critical appraisal skills programme

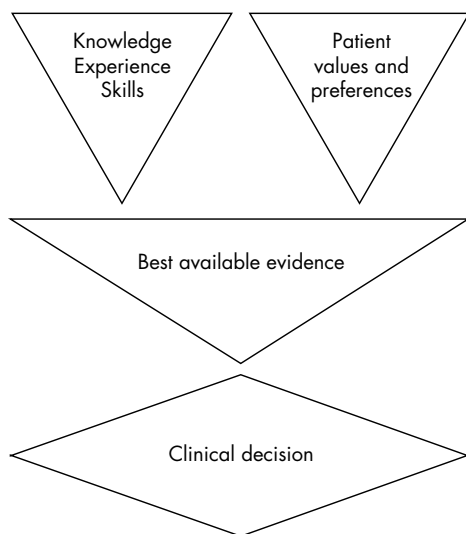


Figure 1 Flow chart of evidence based medicine.³

a good clinical question should have four (or sometimes three) essential components:

- the patient or problem in question;
- the intervention, test, or exposure of interest;
- comparison interventions (if relevant);
- the outcome, or outcomes, of interest.

Thus an answerable clinical question should be structured in the **PICO** (**P**atient or **P**roblem, **I**ntervention, **C**omparison, **O**utcome/s) or **PIO** (**P**atient or **P**roblem, **I**ntervention, **O**utcome/s) format.

To illustrate the concept of PICO/PIO, imagine that you have a four month old baby admitted to your ward with viral bronchiolitis. The child's symptoms get progressively worse and you wonder whether giving corticosteroids might help the child improve and reduce the length of stay in hospital. You decide to use "clinical score" as a measure of improvement. The key components of your clinical question would be:

Patient or problem: 4 month old baby with viral bronchiolitis.

Intervention: corticosteroids.

Comparison: no corticosteroids.

Outcomes: clinical score, length of hospital stay.

A four part clinical question may be formulated as follows:
In a 4 month old baby with viral bronchiolitis, does the administration of corticosteroids compared with not giving corticosteroids improve clinical score and reduce length of hospital stay?

Step 2: Finding the evidence

Once you have formulated your clinical question, the next step is to seek relevant evidence that will help you answer the question. There are several sources of information that may be of help. Traditional sources of information such as textbooks and journals are often too disorganised or out of date.⁸ You may resort to asking colleagues or "experts" but the quality of information obtained from this source is variable. Secondary sources of reliable summarised evidence which may help provide quick evidence based answers to specific clinical questions include *Archimedes* (<http://adc.bmjournals.com/cgi/collection/archimedes>), *Clinical Evidence* (<http://www.clinicalevidence.com/ceweb/conditions/index.jsp>), and *BestBets* (<http://www.bestbets.org/index.html>).

Other important sources of evidence include the online electronic bibliographic databases, which allow thousands of articles to be searched in a relatively short period of time in

an increasing number of journals. The ability to search these databases effectively is an important aspect of EBM. Effective searches aim to maximise the potential of retrieving relevant articles within the shortest possible time. Studies have shown that, even in countries where hospitals have facilities for internet access allowing health care personnel access to a number of electronic databases, many people are not familiar with the process of carrying out efficient searches and often conduct searches which result in too few or too many articles.^{9 10} It is therefore important for health care professionals to undergo basic training in search skills, either through their local library services or through the attendance at formal courses.

BASIC SEARCH PRINCIPLES

Convert the clinical problem into an answerable question

The key to successful searching is to convert your clinical problem into a clear answerable question, which should ideally be framed in the PICO/PIO format as discussed above.

Generate appropriate keywords

A word list can be generated, based on keywords from the clinical question. For example, from the clinical question above, the following keywords could be used for the search: viral bronchiolitis (*patient or problem*); corticosteroids and synonyms: glucocorticoids, steroids, prednisolone, dexamethasone (*intervention*); clinical score, hospital stay (*outcomes*)

Choose a bibliographic database

Numerous online databases are available. These include the Cochrane Library databases, MEDLINE, EMBASE, and CINAHL. In day to day clinical practice, I will suggest that becoming familiar with one or two databases will suffice in most cases. I recommend the Cochrane Library databases and MEDLINE. The Cochrane Library databases—which include the Cochrane database of systematic reviews, the Database of abstracts of reviews of effectiveness, and the Cochrane controlled trials register—is maintained by the Cochrane collaboration, an international initiative which began in the early 1990s and was designed to prepare, maintain, and disseminate systematic reviews of health care interventions.³ The Cochrane Library is updated quarterly and is available through the internet or CD-rom. There is usually a charge for

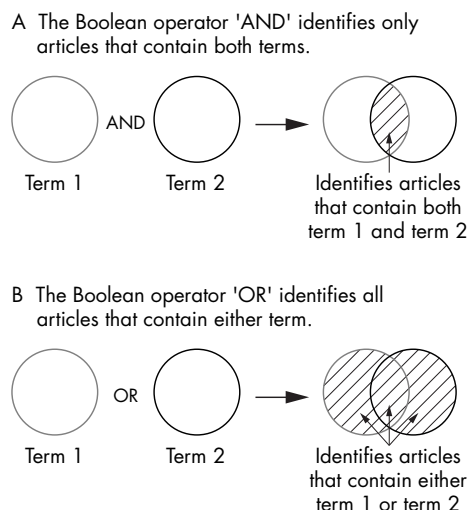


Figure 2 Venn diagram illustrating the use of Boolean operators AND and OR.

using the library, although NHS staff in the United Kingdom have free access to the service through the National Electronic Library for Health.

MEDLINE is probably the most widely used database for searching the biomedical literature.³ It is maintained by the National Library of Medicine, USA. A version of MEDLINE (PUBMED) is freely available on the internet, is updated regularly, and is relatively user friendly.

When looking for articles on effectiveness of interventions or treatments, the first point of call should probably be the Cochrane database of systematic reviews or the other secondary sources mentioned above such as Archimedes, Clinical Evidence, and BestBets. The Cochrane controlled trials register provides an index of published randomised controlled trials. Randomised controlled trials and systematic reviews may also be searched for using MEDLINE. SUMsearch (<http://sumsearch.uthscsa.edu>) is a useful search engine that allows direct searches of external databases with a focus on clinical topics.

Conduct the search

Once the key words and databases have been identified, the next thing is to run the search. At the basic level, an efficient method is to combine individual words or terms using the Boolean operators "AND" and "OR".¹¹ If you are combining two terms, AND allows only articles containing both terms to be retrieved, while OR allows articles containing either term to be retrieved. A simple Venn diagram consisting of two overlapping circles may be used to illustrate this principle. In fig 2A, the two terms have been combined using AND, and in fig 2B, they are combined using OR.

When too many articles come up after the initial search (which is often the case), PUBMED has a feature that allows you to limit the results of your search. You can limit your search by publication type (for example, randomised controlled trials or review articles); by date of publication; by language, by study population, and so on. PubMed also has a feature called "Clinical queries" which provides an easy to use approach to evidence based searching within the Medline database. "Clinical queries" is a preprogrammed research methodology filter that helps busy practitioners access the best available evidence by providing a quick access to reliable clinical studies related to therapy, diagnosis, aetiology, or prognosis.

Example of a basic search strategy

To try to find evidence to answer the clinical question I formulated earlier, we can use the keywords generated to search the Cochrane database of systematic reviews and PUBMED, using the following search strategy:

- (1): Viral bronchiolitis
- (2): Corticosteroids OR steroids OR glucocorticoids OR prednisolone OR dexamethasone
- (3): Clinical score OR hospital stay
- (4): (1) AND (2) AND (3).

When this search strategy was used to search the Cochrane database of systematic reviews on 10 December 2004, four articles were retrieved, but only one of these was relevant.¹²

Other strategies that may be used to improve the sensitivity and specificity of literature searches have been described by Sackett *et al.*¹

Step 3: Appraising the evidence

After you have obtained relevant articles on a subject, the next step is to appraise the evidence for its validity and clinical usefulness. Although there is a wealth of research articles available, the quality of these is variable. Putting

unreliable evidence into practice could lead to harm being caused or limited resources being wasted.

Research evidence may be appraised with regard to three main areas: validity, importance, and applicability to the patient or patients of interest. Critical appraisal provides a structured but simple method for assessing research evidence in all three areas.¹³ Developing critical appraisal skills involves learning how to ask a few key questions about the validity of the evidence and its relevance to a particular patient or group of patients. Such skills may be learnt within small tutorials, workshops, interactive lectures, and at the bedside.¹³

Several tools for appraising research articles are available. I like the tools developed by the Critical Appraisal Skills Programme (CASP), Oxford, UK. These include tools for appraising randomised controlled trials, systematic reviews, case-control studies, and cohort studies. The CASP tools are simple, easy to use, and freely available on the internet.¹⁴

A detailed discussion of the critical appraisal of randomised controlled trials and systematic reviews will be provided in the next two articles of the series.

Step 4: Applying the evidence

When we decide after critical appraisal that a piece of evidence is valid and important, we then have to decide whether that evidence can be applied to our individual patient or population. In deciding this we have to take into account the patient's own personal values and circumstances. The evidence regarding both efficacy and risks should be fully discussed with the patient or parents, or both, in order to allow them to make an informed decision. This approach allows a "therapeutic alliance" to be formed with the patient and the parents and is consistent with the fundamental principle of EBM: the integration of good evidence with clinical expertise and patient values.¹⁵ The decision to apply evidence should also take account of costs and the availability of that particular treatment in your hospital or practice. A practical illustration of issues to consider before applying research evidence will be provided in the fourth article of the series.

Step 5. Evaluating performance

As we incorporate EBM into routine clinical practice, we need to evaluate our approach at frequent intervals and to decide whether we need to improve on any of the four steps discussed above. As Strauss and Sackett have suggested, we need to ask whether we are formulating answerable questions, finding good evidence quickly, effectively appraising the evidence, and integrating clinical expertise and patient's values with the evidence in a way that leads to a rational, acceptable management strategy.¹⁵ Formal auditing of performance may be needed to show whether the EBM approach is improving patient care.

CONCLUSIONS

EBM aims to improve quality of care through the integration of best research evidence with clinical expertise and patient's and parents' preferences. In this article, I have explained the five essential steps for practising EBM, which are: formulating answerable clinical questions; searching for evidence; making a critical appraisal; assessing the applicability of the evidence; and evaluating performance. The principles and critical appraisal of randomised controlled trials, systematic reviews, and meta-analyses, and a practical demonstration of the five step EBM model will be explored further in later articles in this series.

Competing interests: none declared

REFERENCES

- 1 Sackett DL, Strauss SE, Richardson WS, *et al.* *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.
- 2 Steves R, Hootman JM. Evidence-based medicine: what is it and how does it apply to athletic training? *J Athl Train* 2004;**39**:83–7.
- 3 Brownson RC, Baker EA, Leet TL, *et al.* *Evidence based public Health*. New York: Oxford University Press, 2003.
- 4 Gray GE, Pinson LA. Evidence-based medicine and psychiatric practice. *Psychiatr Q* 2003;**74**:387–99.
- 5 Sackett DL. Evidence-based medicine. *Semin Perinatol* 1997;**21**:3–5.
- 6 Levi M. Formulating clinical questions. In: McGovern DPB, Valori RM, Summerskill WSM, Levi M, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001.
- 7 Carneiro AV. The correct formulation of clinical questions for the practice of evidence based medicine. *Acta Med Port* 1998;**11**:745–8.
- 8 Rosenberg WM, Sackett DL. On the need for evidence-based medicine. *Therapie* 1996;**51**:212–7.
- 9 Jordaan M, Jones R. Adoption of internet technology by UK postgraduate centres: a questionnaire survey. *Health Libr Rev* 1999;**16**:166–73.
- 10 Rosenberg WM, Deeks J, Lusher A, *et al.* Improving searching skills and evidence retrieval. *J R Coll Physicians Lond* 1998;**32**:557–63.
- 11 Craig JV, Smyth RL. *The evidence-based manual for nurses*. London: Churchill Livingstone, 2002.
- 12 Patel H, Platt R, Lozano JM, *et al.* Glucocorticoids for acute viral bronchiolitis in infants and young children. *The Cochrane database of systematic reviews*, 2004, issue 3.
- 13 Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* 1995;**310**:1122–6.
- 14 **Critical Appraisal Skills Programme**. Appraisal Tools. Oxford, UK. <http://www.phru.nhs.uk/casp/appraisa.htm> (accessed 10 Dec 2004).
- 15 Straus SE, Sackett DL. Using research findings in clinical practice. *BMJ* 1998;**317**:339–42.

EVIDENCE BASED CHILD HEALTH 2

Understanding randomised controlled trials

A K Akobeng

Arch Dis Child 2005;**90**:840–844. doi: 10.1136/adc.2004.058222

The hierarchy of evidence in assessing the effectiveness of interventions or treatments is explained, and the gold standard for evaluating the effectiveness of interventions, the randomised controlled trial, is discussed. Issues that need to be considered during the critical appraisal of randomised controlled trials, such as assessing the validity of trial methodology and the magnitude and precision of the treatment effect, and deciding on the applicability of research results, are discussed. Important terminologies such as randomisation, allocation concealment, blinding, intention to treat, p values, and confidence intervals are explained.

Figure 1 illustrates such a hierarchy. The ranking has an evolutionary order, moving from simple observational methods at the bottom, through to increasingly rigorous methodologies. The pyramid shape is used to illustrate the increasing risk of bias inherent in study designs as one goes down the pyramid.³ The randomised controlled trial (RCT) is considered to provide the most reliable evidence on the effectiveness of interventions because the processes used during the conduct of an RCT minimise the risk of confounding factors influencing the results. Because of this, the findings generated by RCTs are likely to be closer to the true effect than the findings generated by other research methods.⁴

The hierarchy implies that when we are looking for evidence on the effectiveness of interventions or treatments, properly conducted systematic reviews of RCTs with or without meta-analysis or properly conducted RCTs will provide the most powerful form of evidence.³ For example, if you want to know whether there is good evidence that children with meningitis should be given corticosteroids or not, the best articles to look for would be systematic reviews or RCTs.

WHAT IS A RANDOMISED CONTROLLED TRIAL?

An RCT is a type of study in which participants are randomly assigned to one of two or more clinical interventions. The RCT is the most scientifically rigorous method of hypothesis testing available,⁵ and is regarded as the gold standard trial for evaluating the effectiveness of interventions.⁶ The basic structure of an RCT is shown in fig 2.

Abbreviations: Abbreviations: CONSORT, consolidated standards of reporting trials; EBM, evidence based medicine; PCDAI, paediatric Crohn's disease activity index; RCT, randomised controlled trial

In the first article of the series,¹ I described evidence based medicine (EBM) as a systematic approach to clinical problem solving, which allows the integration of the best available research evidence with clinical expertise and patient values. In this article, I will explain the hierarchy of evidence in assessing the effectiveness of interventions or treatments, and discuss the randomised controlled trial, the gold standard for evaluating the effectiveness of interventions.

HIERARCHY OF EVIDENCE

It is well recognised that some research designs are more powerful than others in their ability to answer research questions on the effectiveness of interventions. This notion has given rise to the concept of "hierarchy of evidence". The hierarchy provides a framework for ranking evidence that evaluates health care interventions and indicates which studies should be given most weight in an evaluation where the same question has been examined using different types of study.²

Correspondence to:
Dr A K Akobeng,
Department of Paediatric
Gastroenterology, Central
Manchester and
Manchester Children's
University Hospitals, Booth
Hall Children's Hospital,
Charlestown Road,
Blackley, Manchester, M9
7AA, UK; tony.akobeng@mmc.nhs.uk

Accepted 22 April 2005

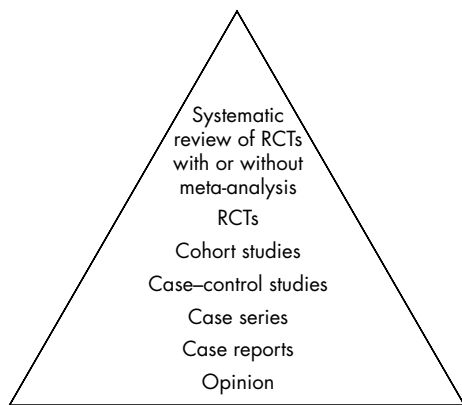


Figure 1 Hierarchy of evidence for questions about the effectiveness of an intervention or treatment.

A sample of the population of interest is randomly allocated to one or another intervention and the two groups are followed up for

a specified period of time. Apart from the interventions being compared, the two groups are treated and observed in an identical manner. At the end of the study, the groups are analysed in terms of outcomes defined at the outset. The results from, say, the treatment A group are compared with results from the treatment B group. As the groups are treated identically apart from the intervention received, any differences in outcomes are attributed to the trial therapy.⁶

WHY A RANDOMISED CONTROLLED TRIAL?

The main purpose of random assignment is to prevent selection bias by distributing the characteristics of patients that may influence the outcome randomly between the groups, so that any difference in outcome can be explained only by the treatment.⁷ Thus random allocation makes it more likely that there will be balancing of baseline systematic differences between intervention groups with regard to known and unknown factors—such as age, sex, disease activity, and duration of disease—that may affect the outcome.

APPRAISING A RANDOMISED CONTROLLED TRIAL

When you are reading an RCT article, the answers to a few questions will help you decide whether you can trust the results of the study and whether you can apply the results to your patient or population. Issues to consider when reading an RCT may be condensed into three important areas⁸:

- the validity of the trial methodology;
- the magnitude and precision of the treatment effect;
- the applicability of the results to your patient or population.

A list of 10 questions that may be used for critical appraisal of an RCT in all three areas is given in box 1.⁹

ASSESSING THE VALIDITY OF TRIAL METHODOLOGY

Focused research question

It is important that research questions be clearly defined at the outset. The question should be focused on the problem of interest, and should be framed in such a way that even somebody who is not a specialist in the field would understand why the study was undertaken.

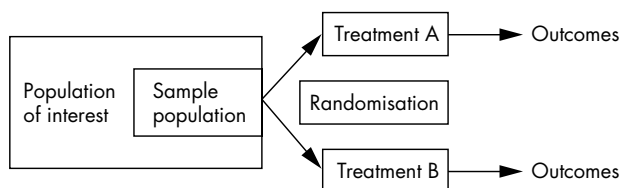


Figure 2 The basic structure of a randomised controlled trial.

Randomisation

Randomisation refers to the process of assigning study participants to experimental or control groups at random such that each participant has an equal probability of being assigned to any given group.¹⁰ The main purpose of randomisation is to eliminate selection bias and balance known and unknown confounding factors in order to create a control group that is as similar as possible to the treatment group.

Methods for randomly assigning participants to groups, which limits bias, include the use of a table of random numbers and a computer program that generates random numbers. Methods of assignment that are prone to bias include alternating assignment or assignment by date of birth or hospital admission number.¹⁰

In very large clinical trials, simple randomisation may lead to a balance between groups in the number of patients allocated to each of the groups, and in patient characteristics. However, in “smaller” studies this may not be the case. Block randomisation and stratification are strategies that may be used to help ensure balance between groups in size and patient characteristics.¹¹

Block randomisation

Block randomisation may be used to ensure a balance in the number of patients allocated to each of the groups in the trial. Participants are considered in blocks of, say, four at a time. Using a block size of four for two treatment arms (A and B) will lead to six possible arrangements of two As and two Bs (blocks):

AABB BBAA ABAB BABA ABBA BAAB

A random number sequence is used to select a particular block, which determines the allocation order for the first four

Box 1: Questions to consider when assessing an RCT⁹

- Did the study ask a clearly focused question?
- Was the study an RCT and was it appropriately so?
- Were participants appropriately allocated to intervention and control groups?
- Were participants, staff, and study personnel blind to participants' study groups?
- Were all the participants who entered the trial accounted for at its conclusion?
- Were participants in all groups followed up and data collected in the same way?
- Did the study have enough participants to minimise the play of chance?
- How are the results presented and what are the main results?
- How precise are the results?
- Were all important outcomes considered and can the results be applied to your local population?

subjects. In the same vein, treatment group is allocated to the next four patients in the order specified by the next randomly selected block.

Stratification

While randomisation may help remove selection bias, it does not always guarantee that the groups will be similar with regard to important patient characteristics.¹² In many studies, important prognostic factors are known before the study. One way of trying to ensure that the groups are as identical as possible is to generate separate block randomisation lists for different combinations of prognostic factors. This method is called stratification or stratified block sampling. For example, in a trial of enteral nutrition in the induction of remission in active Crohn's disease, potential stratification factors might be disease activity (paediatric Crohn's disease activity index (PCDAI) ≤ 25 v > 25) and disease location (small bowel involvement v no small bowel involvement). A set of blocks could be generated for those patients who have PCDAI ≤ 25 and have small bowel disease; those who have PCDAI ≤ 25 and have no small bowel disease; those who have PCDAI > 25 and have small bowel disease; and those who have PCDAI > 25 and have no small bowel disease.

Allocation concealment

Allocation concealment is a technique that is used to help prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. The technique prevents researchers from consciously or unconsciously influencing which participants are assigned to a given intervention group. For instance, if the randomisation sequence shows that patient number 9 will receive treatment A, allocation concealment will remove the ability of researchers or other health care professionals from manoeuvring to place another patient in position 9.

In a recent observational study, Schulz *et al* showed that in trials in which allocation was not concealed, estimates of treatment effect were exaggerated by about 41% compared with those that reported adequate allocation concealment.¹³

A common way for concealing allocation is to seal each individual assignment in an opaque envelope.¹⁰ However, this method may have disadvantages, and "distance" randomisation is generally preferred.¹⁴ Distance randomisation means that assignment sequence should be completely removed from those who make the assignments. The investigator, on recruiting a patient, telephones a central randomisation service which issues the treatment allocation.

Although an RCT should, in theory, eliminate selection bias, there are instances where bias can occur.¹⁵ You should not assume that a trial methodology is valid merely because it is stated to be an RCT. Any selection bias in an RCT invalidates the study design and makes the results no more reliable than an observational study. As Torgesson and Roberts have suggested, the results of a supposed RCT which has had its randomisation compromised by, say, poor allocation concealment may be more damaging than an explicitly unrandomised study, as bias in the latter is acknowledged and the statistical analysis and subsequent interpretation might have taken this into account.¹⁴

Blinding

There is always a risk in clinical trials that perceptions about the advantages of one treatment over another might influence outcomes, leading to biased results. This is particularly important when subjective outcome measures are being used. Patients who are aware that they are receiving what they believe to be an expensive new treatment may report being better than they really are. The judgement of a doctor who expects a particular treatment to be more

effective than another may be clouded in favour of what he perceives to be the more effective treatment. When people analysing data know which treatment group was which, there can be the tendency to "overanalyse" the data for any minor differences that would support one treatment.

Knowledge of treatment received could also influence management of patients during the trial, and this can be a source of bias. For example, there could be the temptation for a doctor to give more care and attention during the study to patients receiving what he perceives to be the less effective treatment in order to compensate for perceived disadvantages.

To control for these biases, "blinding" may be undertaken. The term blinding (sometimes called masking) refers to the practice of preventing study participants, health care professionals, and those collecting and analysing data from knowing who is in the experimental group and who is in the control group, in order to avoid them being influenced by such knowledge.¹⁶ It is important for authors of papers describing RCTs to state clearly whether participants, researchers, or data evaluators were or were not aware of assigned treatment.

In a study where participants do not know the details of the treatment but the researchers do, the term "single blind" is used. When both participants and data collectors (health care professionals, investigators) are kept ignorant of the assigned treatment, the term "double blind" is used. When, rarely, study participants, data collectors, and data evaluators such as statisticians are all blinded, the study is referred to as "triple blind".⁵

Recent studies have shown that blinding of patients and health care professionals prevents bias. Trials that were not double blinded yielded larger estimates of treatment effects than trials in which authors reported double blinding (odds ratios exaggerated, on average, by 17%).¹⁷

It should be noted that, although blinding helps prevent bias, its effect in doing so is weaker than that of allocation concealment.¹⁷ Moreover, unlike allocation concealment, blinding is not always appropriate or possible. For example, in a randomised controlled trial where one is comparing enteral nutrition with corticosteroids in the treatment of children with active Crohn's disease, it may be impossible to blind participants and health care professionals to assigned intervention, although it may still be possible to blind those analysing the data, such as statisticians.

Intention to treat analysis

As stated earlier, the validity of an RCT depends greatly on the randomisation process. Randomisation ensures that known and unknown baseline confounding factors would balance out in the treatment and control groups. However, after randomisation, it is almost inevitable that some participants would not complete the study for whatever reason. Participants may deviate from the intended protocol because of misdiagnosis, non-compliance, or withdrawal. When such patients are excluded from the analysis, we can no longer be sure that important baseline prognostic factors in the two groups are similar. Thus the main rationale for random allocation is defeated, leading to potential bias.

To reduce this bias, results should be analysed on an "intention to treat" basis.

Intention to treat analysis is a strategy in the conduct and analysis of randomised controlled trials that ensures that all patients allocated to either the treatment or control groups are analysed together as representing that treatment arm whether or not they received the prescribed treatment or completed the study.⁵ Intention to treat introduces clinical reality into research by recognising that for several reasons, not all participants randomised will receive the intended treatment or complete the follow up.¹⁸

According to the revised CONSORT statement for reporting RCTs, authors of papers should state clearly which participants are included in their analyses.¹⁹ The sample size per group, or the denominator when proportions are being reported, should be provided for all summary information. The main results should be analysed on the basis of intention to treat. Where necessary, additional analyses restricted only to participants who fulfilled the intended protocol (per protocol analyses) may also be reported.

Power and sample size calculation

The statistical power of an RCT is the ability of the study to detect a difference between the groups when such a difference exists. The power of a study is determined by several factors, including the frequency of the outcome being studied, the magnitude of the effect, the study design, and the sample size.⁵ For an RCT to have a reasonable chance of answering the research question it addresses, the sample size must be large enough—that is, there must be enough participants in each group.

When the sample size of a study is too small, it may be impossible to detect any true differences in outcome between the groups. Such a study might be a waste of resources and potentially unethical. Frequently, however, small sized studies are published that claim no difference in outcome between groups without reporting the power of the studies. Researchers should ensure at the planning stage that there are enough participants to ensure that the study has a high probability of detecting as statistically significant the smallest effect that would be regarded as clinically important.²⁰

MAGNITUDE AND SIGNIFICANCE OF TREATMENT EFFECT

Once you have decided that the methodology of a study is valid within reason, the next step is to decide whether the results are reliable. Two things usually come into mind in making this decision—how big is the treatment effect, and how likely is it that the result obtained is due to chance alone?

Magnitude of treatment effect

Magnitude refers to the size of the measure of effect. Treatment effect in RCTs may be reported in various ways including absolute risk, relative risk, odds ratio, and number needed to treat. These measures of treatment effect and their advantages and disadvantages have recently been reviewed.²¹ A large treatment effect may be more important than a small one.

Statistical significance

Statistical significance refers to the likelihood that the results obtained in a study were not due to chance alone. Probability (p) values and confidence intervals may be used to assess statistical significance.

p Value

A p value can be thought of as the probability that the observed difference between two treatment groups might have occurred by chance. The choice of a significance level is artificial but by convention, many researchers use a p value of 0.05 as the cut off for significance. What this means is that if the p value is less than 0.05, the observed difference between the groups is so unlikely to have occurred by chance that we reject the null hypothesis (that there is no difference) and accept the alternative hypothesis that there is a real difference between the treatment groups. When the p value is below the chosen cut off, say 0.05, the result is generally referred to as being statistically significant. If the p value is greater than 0.05, then we say that the observed difference might have occurred by chance and we fail to reject the null

hypothesis. In such a situation, we are unable to demonstrate a difference between the groups and the result is usually referred to as not statistically significant.

Confidence intervals

The results of any study are estimates of what might happen if the treatment were to be given to the entire population of interest. When I test a new asthma drug on a randomly selected sample of children with asthma in the United Kingdom, the treatment effect I will get will be an estimate of the “true” treatment effect for the whole population of children with asthma in the country. The 95% confidence interval (CI) of the estimate will be the range within which we are 95% certain that the true population treatment effect will lie. It is most common to report 95% CI, but other intervals, such as 90% and 99% CI, may also be calculated for an estimate.

If the CI for a mean difference *includes* 0, then we have been unable to demonstrate a difference between the groups being compared (“not statistically significant”), but if the CI for a mean difference does *not* include 0, then a statistically significant difference between the groups has been shown. In the same vein, if the CI for relative risk or odds ratio for an estimate includes 1, then we have been unable to demonstrate a statistically significant difference between the groups being compared, and if it does not include 1, then there is a statistically significant difference.

Confidence intervals versus p values

CIs convey more useful information than p values. CI may be used to assess statistical significance, provide a range of plausible values for a population parameter, and gives an idea about how precise the measured treatment effect is (see below). Authors of articles could report both p values and CIs.²² However, if only one is to be reported, then it should be the CI, as the p value is less important and can be deduced from the CI; p values tell us little extra when CIs are known.^{22–23}

Clinical significance

A statistically significant finding by itself can have very little to do with clinical practice and has no direct relation to clinical significance. Clinical significance reflects the value of the results to patients and may be defined as a difference in effect size between groups that could be considered to be important in clinical decision making, regardless of whether the difference is statistically significant or not. Magnitude and statistical significance are numerical calculations, but judgements about the clinical significance or clinical importance of the measured effect are relative to the topic of interest.² Judgements about clinical significance should take into consideration how the benefits and any adverse events of an intervention are valued by the patient.

PRECISION OF TREATMENT EFFECT

CI is important because it gives an idea about how precise an estimate is. The width of the interval indicates the precision of the estimate. The wider the interval, the less the precision. A very wide interval may indicate that more data should be collected before anything definite can be said about the estimate.

APPLYING RESULTS TO YOUR OWN PATIENTS

An important concept of EBM is that clinicians should make decisions about whether the valid results of a study are applicable to their patients. The fact that good evidence is available on a particular asthma treatment does not necessarily mean that all patients with asthma can or should be given that treatment. Some of the issues one needs to consider before deciding whether to incorporate a particular

piece of research evidence into clinical practice are briefly discussed below.

Are the participants in the study similar enough to my patients?

If a particular drug has been found to be effective in adults with meningitis in the USA, you need to decide whether there is any biological, geographical, or cultural reason why that particular drug will not be effective in children with meningitis in the United Kingdom.

Do the potential side effects of the drug outweigh the benefits?

If a particular treatment is found to be effective in an RCT, you need to consider whether the reported or known side effects of the drug may outweigh its potential benefits to your patient. You may also need to consider whether an individual patient has any potential co-morbid condition which may alter the balance of benefits and risks. In such a situation, you may, after consultation with the patient or carers, decide not to offer the treatment.

Does the treatment conflict with the patient's values and expectations?

Full information about the treatment should be given to the patient or carers, and their views on the treatment should be taken into account. A judgement should be made about how the patient and carers value the potential benefits of the treatment as against potential harms.

Is the treatment available and is my hospital prepared to fund it?

There will be no point in prescribing a treatment which cannot either be obtained in your area of work or which your hospital or practice is not in a position to fund, for whatever reason, including cost.

CONCLUSIONS

An RCT is the most rigorous scientific method for evaluating the effectiveness of health care interventions. However, bias could arise when there are flaws in the design and management of a trial. It is important for people reading medical reports to develop the skills for critically appraising RCTs, including the ability to assess the validity of trial

methodology, the magnitude and precision of the treatment effect, and the applicability of results.

Competing interests: none declared

REFERENCES

- 1 Akobeng AK. Evidence-based child health. 1. Principles of evidence-based medicine. *Arch Dis Child* 2005;**90**:837–40.
- 2 Rychetnik L, Hawe P, Waters E, et al. A glossary for evidence based public health. *J Epidemiol Community Health* 2004;**58**:538–45.
- 3 Craig JV, Smyth RL. *The evidence-based manual for nurses*. London: Churchill Livingstone, 2002.
- 4 Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003;**12**:77–84.
- 5 Last JM. *A dictionary of epidemiology*. New York: Oxford University Press, 2001.
- 6 McGovern DPB. Randomized controlled trials. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:26–9.
- 7 Roberts C, Torgerson D. Randomisation methods in controlled trials. *BMJ* 1998;**317**:1301–10.
- 8 Sackett DL, Strauss SE, Richardson WS, et al. *evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.
- 9 Critical Appraisal Skills Programme. *Appraisal tools*. Oxford, UK, <http://www.phru.nhs.uk/casp/rcts.htm> (accessed 8 December 2004).
- 10 Lang TA, Secic M. *How to report statistics in medicine*. Philadelphia: American College of Physicians, 1997.
- 11 Altman DG, Bland JM. How to randomise. *BMJ* 1999;**319**:703–4.
- 12 Chia KS. Randomisation: magical cure for bias. *Ann Acad Med Singapore* 2000;**29**:563–4.
- 13 Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.
- 14 Torgerson DJ, Roberts C. Randomisation methods: concealment. *BMJ* 1999;**319**:375–6.
- 15 Torgerson DJ, Torgerson CJ. Avoiding bias in randomised controlled trials in educational research. *Br J Educ Stud* 2003;**51**:36–45.
- 16 Day SJ, Altman DG. Blinding in clinical trials and other studies. *BMJ* 2000;**321**:504.
- 17 Schulz KF. Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evid Based Nurs* 2000;**5**:36–7.
- 18 Summerskill WSM. Intention to treat. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:105–7.
- 19 Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomised controlled trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.
- 20 Devane D, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. *J Adv Nurs* 2004;**47**:297–302.
- 21 Akobeng AK. Understanding measures of treatment effect in clinical trials. *Arch Dis Child*, 2005;**90**, 54–6.
- 22 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall/CRC, 1991:152–78.
- 23 Coggon D. *Statistics in clinical practice*. London: BMJ Publishing Group, 1995.

EVIDENCE BASED CHILD HEALTH 3

Understanding systematic reviews and meta-analysis

A K Akobeng

Arch Dis Child 2005;90:845–848. doi: 10.1136/adc.2004.058280

This review covers the basic principles of systematic reviews and meta-analyses. The problems associated with traditional narrative reviews are discussed, as is the role of systematic reviews in limiting bias associated with the assembly, critical appraisal, and synthesis of studies addressing specific clinical questions. Important issues that need to be considered when appraising a systematic review or meta-analysis are outlined, and some of the terms used in the reporting of systematic reviews and meta-analyses—such as odds ratio, relative risk, confidence interval, and the forest plot—are introduced.

Health care professionals are increasingly required to base their practice on the best available evidence. In the first article of the series, I described basic strategies that could be used to search the medical literature.¹ After a literature search on a specific clinical question, many articles may be retrieved. The quality of the studies may be variable, and the individual studies might have produced conflicting results. It is therefore important that health care decisions are not based solely on one or two studies without account being taken of the whole range of research information available on that topic.

Health care professionals have always used review articles as a source of summarised evidence on a particular topic. Review articles in the medical literature have traditionally been in the form of “narrative reviews” where experts in a particular field provide what is supposed to be a “summary of evidence” in that field. Narrative reviews, although still very common in the medical field, have been criticised because of the high risk of bias, and “systematic reviews” are preferred.² Systematic reviews apply scientific strategies in ways that limit bias to the assembly, a critical appraisal, and synthesis of relevant studies that address a specific clinical question.²

THE PROBLEM WITH TRADITIONAL REVIEWS

The validity of a review article depends on its methodological quality. While traditional review articles or narrative reviews can be useful when conducted properly, there is evidence that they are usually of poor quality. Authors of narrative reviews often use informal, subjective methods to collect and interpret studies and tend to be selective in citing reports that reinforce their preconceived ideas or promote their own views on a topic.^{3,4} They are also rarely explicit about

how they selected, assessed, and analysed the primary studies, thereby not allowing readers to assess potential bias in the review process. Narrative reviews are therefore often biased, and the recommendations made may be inappropriate.⁵

WHAT IS A SYSTEMATIC REVIEW?

In contrast to a narrative review, a systematic review is a form of research that provides a summary of medical reports on a specific clinical question, using explicit methods to search, critically appraise, and synthesise the world literature systematically.⁶ It is particularly useful in bringing together a number of separately conducted studies, sometimes with conflicting findings, and synthesising their results.

By providing in a clear explicit fashion a summary of all the studies addressing a specific clinical question,⁴ systematic reviews allow us to take account of the whole range of relevant findings from research on a particular topic, and not just the results of one or two studies. Other advantages of systematic reviews have been discussed by Mulrow.⁷ They can be used to establish whether scientific findings are consistent and generalisable across populations, settings, and treatment variations, or whether findings vary significantly by particular subgroups. Moreover, the explicit methods used in systematic reviews limit bias and, hopefully, will improve reliability and accuracy of conclusions. For these reasons, systematic reviews of randomised controlled trials (RCTs) are considered to be evidence of the highest level in the hierarchy of research designs evaluating effectiveness of interventions.⁸

METHODOLOGY OF A SYSTEMATIC REVIEW

The need for rigour in the preparation of a systematic review means that there should be a formal process for its conduct. Figure 1 summarises the process for conducting a systematic review of RCTs.⁹ This includes a comprehensive, exhaustive search for primary studies on a focused clinical question, selection of studies using clear and reproducible eligibility criteria, critical appraisal of primary studies for quality, and synthesis of results according to a predetermined and explicit method.^{3,9}

WHAT IS A META-ANALYSIS?

Following a systematic review, data from individual studies may be pooled quantitatively and reanalysed using established statistical methods.¹⁰ This technique is called meta-analysis. The

Abbreviations: RCT, randomised controlled trial

Correspondence to:
Dr A K Akobeng,
Department of Paediatric
Gastroenterology, Central
Manchester and
Manchester Children's
University Hospitals, Booth
Hall Children's Hospital,
Charlestown Road,
Blackley, Manchester, M9
7AA, UK; tony.akobeng@
mmc.nhs.uk

Accepted 22 April 2005

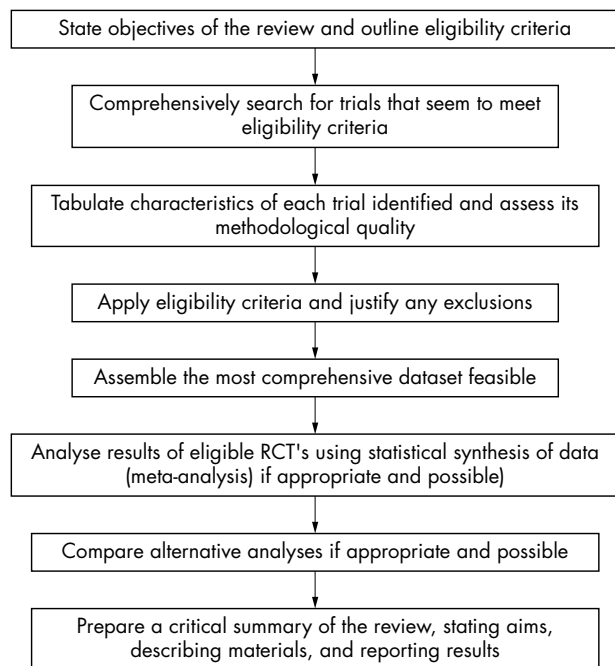


Figure 1 Methodology for a systematic review of randomised controlled trials.⁹

rationale for a meta-analysis is that, by combining the samples of the individual studies, the overall sample size is increased, thereby improving the statistical power of the analysis as well as the precision of the estimates of treatment effects.¹¹

Meta-analysis is a two stage process.¹² The first stage involves the calculation of a measure of treatment effect with its 95% confidence intervals (CI) for each individual study. The summary statistics that are usually used to measure treatment effect include odds ratios (OR), relative risks (RR), and risk differences.

In the second stage of meta-analysis, an overall treatment effect is calculated as a weighted average of the individual summary statistics. Readers should note that, in meta-analysis, data from the individual studies are not simply combined as if they were from a single study. Greater weights are given to the results from studies that provide more information, because they are likely to be closer to the "true effect" we are trying to estimate. The weights are often the inverse of the variance (the square of the standard error) of the treatment effect, which relates closely to sample size.¹² The typical graph for displaying the results of a meta-analysis is called a "forest plot".¹³

The forest plot

The plot shows, at a glance, information from the individual studies that went into the meta-analysis, and an estimate of the overall results. It also allows a visual assessment of the amount of variation between the results of the studies (heterogeneity). Figure 2 shows a typical forest plot. This figure is adapted from a recent systematic review and meta-analysis which examined the efficacy of probiotics compared with placebo in the prevention and treatment of diarrhoea associated with the use of antibiotics.¹⁴

Description of the forest plot

In the forest plot shown in fig 2, the results of nine studies have been pooled. The names on the left of the plot are the first authors of the primary studies included. The black squares represent the odds ratios of the individual studies,

and the horizontal lines their 95% confidence intervals. The area of the black squares reflects the weight each trial contributes in the meta-analysis. The 95% confidence intervals would contain the true underlying effect in 95% of the occasions if the study was repeated again and again. The solid vertical line corresponds to no effect of treatment (OR = 1.0). If the CI includes 1, then the difference in the effect of experimental and control treatment is not significant at conventional levels ($p > 0.05$).¹⁵ The overall treatment effect (calculated as a weighted average of the individual ORs) from the meta-analysis and its CI is at the bottom and represented as a diamond. The centre of the diamond represents the combined treatment effect (0.37), and the horizontal tips represent the 95% CI (0.26 to 0.52). If the diamond shape is on the **Left** of the line of no effect, then **Less** (fewer episodes) of the outcome of interest is seen in the treatment group. If the diamond shape is on the **Right** of the line, then **more** episodes of the outcome of interest are seen in the treatment group. In fig 2, the diamond shape is found on the left of the line of no effect, meaning that less diarrhoea (fewer episodes) was seen in the probiotic group than in the placebo group. If the diamond touches the line of no effect (where the OR is 1) then there is no statistically significant difference between the groups being compared. In fig 2, the diamond shape does not touch the line of no effect (that is, the confidence interval for the odds ratio does not include 1) and this means that the difference found between the two groups was statistically significant.

APPRAISING A SYSTEMATIC REVIEW WITH OR WITHOUT META-ANALYSIS

Although systematic reviews occupy the highest position in the hierarchy of evidence for articles on effectiveness of interventions,⁸ it should not be assumed that a study is valid merely because it is stated to be a systematic review. Just as in RCTs, the main issues to consider when appraising a systematic review can be condensed into three important areas⁸:

- The validity of the trial methodology.
- The magnitude and precision of the treatment effect.
- The applicability of the results to your patient or population.

Box 1 shows a list of 10 questions that may be used to appraise a systematic review in all three areas.¹⁶

ASSESSING THE VALIDITY OF TRIAL METHODOLOGY

Focused research question

Like all research reports, the authors should clearly state the research question at the outset. The research question should include the relevant population or patient groups being studied, the intervention of interest, any comparators (where relevant), and the outcomes of interest. Keywords from the research question and their synonyms are usually used to identify studies for inclusion in the review.

Types of studies included in the review

The validity of a systematic review or meta-analysis depends heavily on the validity of the studies included. The authors should explicitly state the type of studies they have included in their review, and readers of such reports should decide whether the included studies have the appropriate study design to answer the clinical question. In a recent systematic review which determined the effects of glutamine supplementation on morbidity and weight gain in preterm babies the investigators based their review only on RCTs.¹⁷

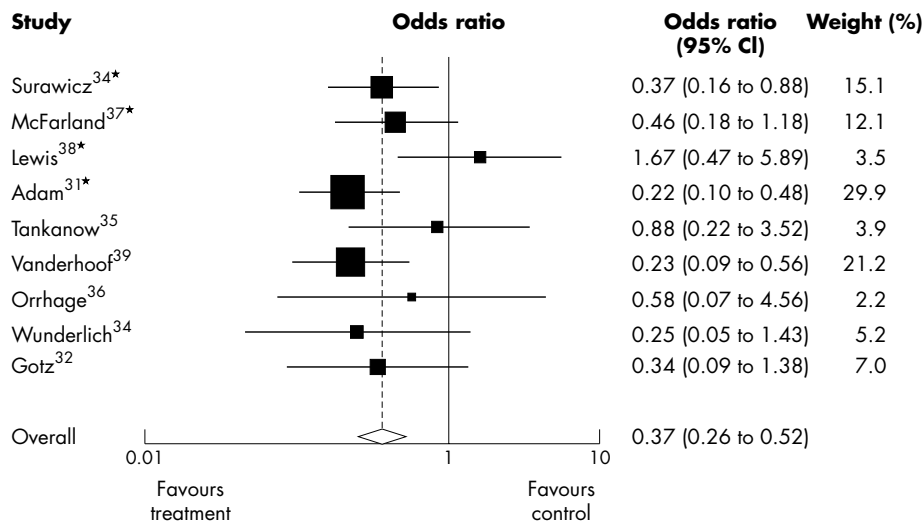


Figure 2 Effect of probiotics on the risk of antibiotic associated diarrhoea.¹⁴

Search strategy used to identify relevant articles

There is evidence that single electronic database searches lack sensitivity and relevant articles may be missed if only one database is searched. Dickersin *et al* showed that only 30–80% of all known published RCTs were identifiable using MEDLINE.¹⁸ Even if relevant records are in a database, it can be difficult to retrieve them easily. A comprehensive search is therefore important, not only for ensuring that as many studies as possible are identified but also to minimise selection bias for those that are found. Relying exclusively on one database may retrieve a set of studies that are unrepresentative of all studies that would have been identified through a comprehensive search of multiple sources. Therefore, in order to retrieve all relevant studies on a topic, several different sources should be searched to identify relevant studies (published and unpublished), and the search strategy should not be limited to the English language. The aim of an extensive search is to avoid the problem of publication bias which occurs when trials with statistically significant results are more likely to be published and cited, and are preferentially published in English language journals and those indexed in Medline.

In the systematic review referred to above, which examined the effects of glutamine supplementation on

morbidity and weight gain in preterm babies, the authors searched the Cochrane controlled trials register, Medline, and Embase,¹⁷ and they also hand searched selected journals, cross referencing where necessary from other publications.

Quality assessment of included trials

The reviewers should state a predetermined method for assessing the eligibility and quality of the studies included. At least two reviewers should independently assess the quality of the included studies to minimise the risk of selection bias. There is evidence that using at least two reviewers has an important effect on reducing the possibility that relevant reports will be discarded.¹⁹

Pooling results and heterogeneity

If the results of the individual studies were pooled in a meta-analysis, it is important to determine whether it was reasonable to do so. A clinical judgement should be made about whether it was reasonable for the studies to be combined based on whether the individual trials differed considerably in populations studied, interventions and comparisons used, or outcomes measured.

The statistical validity of combining the results of the various trials should be assessed by looking for homogeneity of the outcomes from the various trials. In other words, there should be some consistency in the results of the included trials. One way of doing this is to inspect the graphical display of results of the individual studies (forest plot, see above) looking for similarities in the direction of the results. When the results differ greatly in their direction—that is, if there is significant heterogeneity—then it may not be wise for the results to be pooled. Some articles may also report a statistical test for heterogeneity, but it should be noted that the statistical power of many meta-analyses is usually too low to allow the detection of heterogeneity based on statistical tests. If a study finds significant heterogeneity among reports, the authors should attempt to offer explanations for potential sources of the heterogeneity.

Magnitude of the treatment effect

Common measures used to report the results of meta-analyses include the odds ratio, relative risk, and mean differences. If the outcome is binary (for example, disease *v* no disease, remission *v* no remission), odds ratios or relative risks are used. If the outcome is continuous (for example, blood pressure measurement), mean differences may be used.

Box 1: Questions to consider when appraising a systematic review¹⁶

- Did the review address a clearly focused question?
- Did the review include the right type of study?
- Did the reviewers try to identify all relevant studies?
- Did the reviewers assess the quality of all the studies included?
- If the results of the study have been combined, was it reasonable to do so?
- How are the results presented and what are the main results?
- How precise are the results?
- Can the results be applied to your local population?
- Were all important outcomes considered?
- Should practice or policy change as a result of the evidence contained in this review?

ODDS RATIOS AND RELATIVE RISKS

Odds and odds ratio

The *odds* for a group is defined as the number of patients in the group who achieve the stated end point divided by the number of patients who do not. For example, the odds of acne resolution during treatment with an antibiotic in a group of 10 patients may be 6 to 4 (6 with resolution of acne divided by 4 without = 1.5); in a control group the odds may be 3 to 7 (0.43). The *odds ratio*, as the name implies, is a ratio of two odds. It is simply defined as the ratio of the odds of the treatment group to the odds of the control group. In our example, the odds ratio of treatment to control group would be 3.5 (1.5 divided by 0.43).

Risk and relative risk

Risk, as opposed to odds, is calculated as the number of patients in the group who achieve the stated end point divided by the *total* number of patients in the group. Risk ratio or relative risk is a ratio of two "risks". In the example above the risks would be 6 in 10 in the treatment group (6 divided by 10 = 0.6) and 3 in 10 in the control group (0.3), giving a risk ratio, or relative risk of 2 (0.6 divided by 0.3).

Interpretation of odds ratios and relative risk

An odds ratio or relative risk greater than 1 indicates increased likelihood of the stated outcome being achieved in the treatment group. If the odds ratio or relative risk is less than 1, there is a decreased likelihood in the treatment group. A ratio of 1 indicates no difference—that is, the outcome is just as likely to occur in the treatment group as it is in the control group.¹¹ As in all estimates of treatment effect, odds ratios or relative risks reported in meta-analysis should be accompanied by confidence intervals.

Readers should understand that the odds ratio will be close to the relative risk if the end point occurs relatively infrequently, say in less than 20%.¹⁵ If the outcome is more common, then the odds ratio will considerably overestimate the relative risk. The advantages and disadvantages of odds ratios *v* relative risks in the reporting of the results of meta-analysis have been reviewed elsewhere.¹²

Precision of the treatment effect: confidence intervals

As stated earlier, confidence intervals should accompany estimates of treatment effects. I discussed the concept of confidence intervals in the second article of the series.⁸ Ninety five per cent confidence intervals are commonly reported, but other intervals such as 90% or 99% are also sometimes used. The 95% CI of an estimate (for example, of odds ratios or relative risks) will be the range within which we are 95% certain that the true population treatment effect will lie. The width of a confidence interval indicates the precision of the estimate. The wider the interval, the less the precision. A very long interval makes us less sure about the accuracy of a study in predicting the true size of the effect. If the confidence interval for relative risk or odds ratio for an estimate includes 1, then we have been unable to demonstrate a statistically significant difference between the groups being compared; if it does not include 1, then we say that there is a statistically significant difference.

APPLICABILITY OF RESULTS TO PATIENTS

Health care professionals should always make judgements about whether the results of a particular study are applicable to their own patient or group of patients. Some of the issues that one need to consider before deciding whether to incorporate a particular piece of research evidence into clinical practice were discussed in the second article of the series.⁸ These include similarity of study population to your population, benefit *v* harm, patients preferences, availability, and costs.

CONCLUSIONS

Systematic reviews apply scientific strategies to provide in an explicit fashion a summary of all studies addressing a specific question, thereby allowing an account to be taken of the whole range of relevant findings on a particular topic. Meta-analysis, which may accompany a systematic review, can increase power and precision of estimates of treatment effects. People working in the field of paediatrics and child health should understand the fundamental principles of systematic reviews and meta-analyses, including the ability to apply critical appraisal not only to the methodologies of review articles, but also to the applicability of the results to their own patients.

Competing interests: none declared

REFERENCES

- 1 Akobeng AK. Evidence based child health 1. Principles of evidence based medicine. *Arch Dis Child* 2005;**90**:837–40.
- 2 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;**126**:376–80.
- 3 Pai M, McCulloch M, Gorman JD, et al. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *Natl Med J India* 2004;**17**:86–95.
- 4 McGovern DPB. Systematic reviews. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:17–9.
- 5 McAlister FA, Clark HD, van Walraven C, et al. The medical review article revisited: has the science improved? *Ann Intern Med* 1999;**131**:947–51.
- 6 Sackett DL, Strauss SE, Richardson WS, et al. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill Livingstone, 2000.
- 7 Mulrow CD. Systematic reviews: rationale for systematic reviews. *BMJ* 1994;**309**:597–9.
- 8 Akobeng AK. Evidence based child health 2. Understanding randomised controlled trials. *Arch Dis Child* 2005;**90**:840–4.
- 9 Greenhalgh T. How to read a paper: papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997;**315**:672–5.
- 10 Muir Gray JA. *Evidence based healthcare. How to make health policy and management decisions*. London: Churchill Livingstone, 2001:125–6.
- 11 Lang TA, Secic M. *How to report statistics in medicine*. Philadelphia: American College of Physicians, 1997.
- 12 Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. *Systematic reviews in healthcare: meta-analysis in context*. London: BMJ Publishing Group, 2001:285–312.
- 13 Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001;**322**:1479–80.
- 14 D'Souza AL, Rajkumar C, Cooke J, et al. Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis. *BMJ* 2002;**324**:1361.
- 15 Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;**315**:1533–7.
- 16 Critical Appraisal Skills Programme. Appraisal Tools. Oxford, UK. <http://www.phru.nhs.uk/casp/appraisa.htm> (accessed 10 Dec 2004).
- 17 Tubman TRJ, Thompson SW. Glutamine supplementation for prevention of morbidity in preterm infants. The Cochrane Database of Systematic Reviews 2001, Issue 4.
- 18 Dickersin K, Scherer R, Lefebvre C. Systematic reviews: identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.
- 19 Clarke M, Oxman AD, eds. Selecting studies. Cochrane reviewers' handbook 4.2.0 [updated March 2003]. In: *The Cochrane library*, issue 2. Oxford: Update Software, 2003.

EVIDENCE BASED CHILD HEALTH 4

Evidence in practice

A K Akobeng

Arch Dis Child 2005;90:849–852. doi: 10.1136/adc.2004.058248

A clinical scenario is used to illustrate how the principles outlined in the previous articles in the series could be applied to help improve patient care. A practical demonstration of the art of formulating answerable clinical questions, finding evidence, critically appraising evidence, and putting evidence into practice is provided. The importance of integrating evidence with patient's preferences, and taking account of issues such as availability of interventions, costs, and so on is discussed. Finally, some of the issues involved in the development of evidence based policies within clinical teams are outlined.

had previously received some training in evidence based medicine (EBM) and knew that clinical questions are best formatted in the **PICO** or **PIO** format to include the **P**atient's problem or diagnosis; the **I**ntervention of interest, as well as any **C**omparison intervention (if relevant), and the **O**utcome of interest.¹

His focused clinical question was therefore based on the following key elements arising out of his current information need:

- **Patient:** a 16 year old girl with active Crohn's disease unresponsive to conventional treatment
- **Intervention:** infliximab
- **Comparison:** no infliximab
- **Outcome:** induction of remission

These key elements were used to formulate the following clinical question:

In a 16 year old girl with active Crohn's disease unresponsive to conventional therapy, is infliximab effective in inducing remission?

SEARCH STRATEGY

As Dr B's question was about a treatment (an intervention), he was most interested in obtaining relevant systematic reviews of randomised controlled trials with or without meta-analysis, and randomised controlled trials (RCTs). He searched the following databases to which he had easy access at his hospital: the Cochrane database of systematic reviews, and Medline via PubMed.

Keywords needed for the search were drawn from the above focused clinical question. These were: Crohn's disease or its synonym Crohn disease; infliximab or its synonyms remicade and monoclonal antibodies cA2; and remission.

The following search strategy was used to search the Cochrane database of systematic reviews from the Cochrane Library (issue 4, 2004), and PUBMED (1966 to present) on the 28 December 2004: (1) Crohn's disease *or* Crohn disease; (2) infliximab *or* remicade *or* monoclonal antibodies cA2; (3) remission; (4): (1) *and* (2) *and* (3).

To make sure he was not missing any relevant systematic reviews, he also searched the "clinical queries" option for systematic reviews in PUBMED using the search strategy "infliximab *and* Crohn's disease".

Abbreviations: CDAI, Crohn's disease activity index; CDEI, Crohn's disease endoscopic index; EBM, evidence based medicine; RCT, randomised controlled trial; RR, relative risk; AR, absolute risk; ARR, absolute risk reduction; NNT, number needed to treat

CLINICAL SCENARIO

Laura, a 16 year old girl, was diagnosed with Crohn's disease 5 years earlier. Her disease was fairly well controlled in the first few years with intermittent use of corticosteroids, enteral nutrition, and mesalazine. In the past 6 months, however, she has been having persistent symptoms, and despite several courses of enteral nutrition and steroids together with immunosuppressive therapy (azathioprine), she remained quite unwell with bloody diarrhoea, abdominal pain, and marked weight loss. She was admitted to hospital with worsening symptoms and required treatment with intravenous steroids, with little benefit. Gastrointestinal endoscopies and barium imaging at this time confirmed severe extensive inflammation of the colon and small bowel. Laura's disease was clearly unresponsive to conventional treatment. Surgery was thought not to be an option because of the extensive nature of her disease. At one of their ward rounds, Dr B, specialist registrar in paediatrics told the consultant paediatrician that he had recently attended a meeting where he heard that a drug called infliximab could be useful in patients with Crohn's disease unresponsive to conventional treatment. The consultant asked Dr B to look for further evidence on the effectiveness of infliximab in the induction of remission in Crohn's disease. Dr B initially looked through his main paediatric gastroenterology textbook (published 7 years previously) for more information on this drug, but found nothing. He decided to investigate further for relevant evidence.

FOCUSED CLINICAL QUESTION

Dr B first decided to convert his information needs into an answerable clinical question. He

Correspondence to:
Dr A K Akobeng,
Department of Paediatric
Gastroenterology, Central
Manchester and
Manchester Children's
University Hospitals, Booth
Hall Children's Hospital,
Charlestown Road,
Blackley, Manchester, M9
7AA, UK; tony.akobeng@
mmc.nhs.uk

Accepted 22 April 2005

Results of the search

Cochrane library search

One relevant article was found in the Cochrane Database of Systematic Reviews.²

PubMed search

The main PubMed search initially yielded 125 articles. These were too many, so he limited the search further to try and retrieve the most relevant articles. Initially limiting the search to "meta-analysis", no article was obtained. He then limited the search to "randomised controlled trials", and 12 articles were retrieved. A quick glance at the titles or abstracts of these articles showed that only two of them were potentially relevant.^{3,4} The rest were either narrative reviews, case series, or trials on the use of infliximab for other indications. However, he realised that the two articles identified had been incorporated into the systematic reviews identified from the Cochrane Library search and he decided that there was no need to collect those articles separately. The "clinical queries" search identified the systematic reviews already retrieved from the Cochrane library.

Summary of the systematic review

After reviewing a number of articles, the authors of the systematic review included one RCT that had investigated the efficacy of infliximab in the induction of remission in Crohn's disease.³ In that study, 108 patients (age range 26 to 46 years) with moderate to severe Crohn's disease resistant to conventional treatment were recruited from 18 centres in North America and Europe. They were randomised to receive a single two hour intravenous infusion of infliximab or placebo. Outcome measures were "clinical remission" defined as a Crohn's disease activity index (CDAI) less than 150 at four weeks after the infusion, and "clinical response" defined as a reduction of 70 points or more in the CDAI score at four weeks after the infusion. In a subgroup analysis of the same study, changes in Crohn's disease endoscopic index score (CDEIS) from baseline after four weeks in a 30 patient subset of the original cohort, who also underwent colonoscopy before the infusion and four weeks after infusion, were reported.⁵

After reviewing the available evidence, the authors of the systematic review concluded that a single intravenous infusion of infliximab may be effective for inducing remission in Crohn's disease. Based on their review, they recommended a dose of 5 mg/kg.

CRITICAL APPRAISAL

Dr B assessed the quality of the systematic review using the critical appraisal checklist published by the Critical Appraisal Skills Programme (CASP), Oxford, UK.⁶ A summary of Dr B's appraisal is shown below:

Summary of the appraisal

Did the review address a clearly focused question?

The reviewers aimed to evaluate the efficacy of infliximab in the induction of remission of Crohn's disease and to determine adverse events associated with infliximab treatment in Crohn's disease.

Did the reviewers try to identify all relevant studies?

They searched MEDLINE, EMBASE, the Cochrane central register of controlled trials from the Cochrane Library, and the Cochrane Inflammatory Bowel Disease Review Group specialised trials register. They also hand searched articles cited in each publication obtained, and contacted leaders in the field and manufacturers of infliximab to try and obtain unpublished articles. The search strategy was not limited by language.

Did the reviewers assess the quality of all studies included?

The reviewers included one RCT on the effectiveness of infliximab in the induction of remission in Crohn's disease. The methodological quality of the trial was assessed independently by two reviewers using the criteria described in the *Cochrane Reviewers' Handbook*⁷ and the Jadad scale.⁸ Allocation of participants to intervention or placebo was random and allocation concealment was adjudged adequate. Investigators, all other study personnel, and patients were blinded to treatment assignments. Data were analysed according to the intention to treat principle. Although no subgroup analysis based on colonoscopic findings was prespecified by the authors of the primary study, they also subsequently reported results for a 30 patient subset of the original cohort, who also underwent colonoscopy before and four weeks after infusion.⁵

If the result of the study has been combined, was it reasonable to do so?

As only one RCT was found on the use of infliximab in CD, no meta-analysis was undertaken.

What are the main results and how precise are the main results?

As stated earlier, outcome measures of the primary study were "clinical remission" defined as a Crohn's disease activity index (CDAI) less than 150 at four weeks after the infusion, and "clinical response" defined as a reduction of 70 points or more in the CDAI score at four weeks. The authors of the systematic review reported their main results in relative risks (RR) and their 95% confidence intervals (CI). Four weeks after the infusion, 27 of 83 patients in the infliximab group v 1 of 25 patients in the placebo group were in remission (RR = 8.1 (95% CI, 1.2 to 56.9)). Fifty four of 83 patients in the infliximab group v 4 of 25 in the placebo achieved clinical response (RR = 4.1 (1.6 to 10.1)).

Dr B decided to also express the results on clinical remission (which he considered to be the most important outcome) in terms of absolute risk (AR) measures.⁹ Based on data in the article, he calculated the AR of achieving clinical remission for patients receiving infliximab as $27/83 = 0.325$ or 32.5%, and the AR of achieving remission for patients receiving placebo as $1/25 = 0.04$ or 4%. Thus the absolute risk reduction (ARR) or risk difference with regard to the achievement of clinical remission is $(0.325 - 0.040)$ or $0.285 = 28.5\%$. This means that in the study, receiving infliximab increased a patient's chance of going into remission by 28.5%. Dr B also felt that it would be necessary to express the results in a more user friendly way by calculating the number needed to treat (NNT) which, in this context, is defined as the number of patients who need to receive infliximab in order to achieve remission in one of them. The concept of NNTs has been discussed elsewhere.⁹ Dr B calculated the NNT as the reciprocal of the ARR or risk difference, and this was $1/0.285$ or 3.5. This means that in the population studied, about four people with Crohn's disease unresponsive to conventional treatment needed to receive infliximab in order to get one of them into remission.

Dr B was, however, aware that because the study was undertaken in a different population, he might not be able to extrapolate the calculated ARR or NNT directly to his own population, as these measures are dependent on baseline risk.⁹ The calculated ARR or NNT apply only to populations whose baseline risk (the risk of going into remission without infliximab) is similar to the study population. He would be able calculate an ARR for his population using the relative risk figures if he knew the baseline risk of his population,⁹ but no such local data on this were available.

In the subgroup analysis of patients who also underwent endoscopy, mean CDEIS in the infliximab group decreased significantly from (mean (SD)) 13.0 (7.1) to 5.3 (4.4) ($p < 0.001$). No significant change in mean CDEIS was observed in the placebo group (changed from 8.4 (6.3) to 7.5 (5.4)). Changes in CDEIS showed a significant correlation with changes in the CDAI ($r = 0.56$, $p = 0.002$).

Were other important outcomes considered?

The review considered both efficacy and adverse events of treatment. The authors commented that the short term adverse effects observed with infliximab were, in most cases, not very different from those observed with placebo, but the sample sizes of the primary studies may not have been adequate to detect any increased risk in adverse events in patients receiving infliximab. The follow up period was also probably too short to allow adequate assessment of potential long term adverse events.

Despite the findings of the systematic review, Dr B was aware from his reading that adverse events reported for infliximab in case reports and case series included infusion reactions, infections, and headaches. Other members of the medical team also mentioned that there had been recent case series that have described a potential association between infliximab use and serious adverse events such as tuberculosis and lymphoma in adult patients.

Are the results of the systematic review applicable to my patient?

The primary study recruited patients from Europe and North America and the main inclusion criterion was active Crohn's disease unresponsive to conventional treatment. The age range of participants was 26 to 46 years. Dr B could not see any biological reason why a drug effective in a 26 year old with Crohn's disease would not be so in a 16 year old. The study had defined "active disease" as a CDAI between 220 and 400. In order to be certain that his patient satisfied this condition, he estimated a CDAI for her and it was 300, which was within the range. He was therefore satisfied that his patient could reasonably be considered to be fairly similar to the participants of the study. Because of the reports of infliximab triggering the development of tuberculosis, Dr B screened Laura for evidence of latent tuberculosis by performing a chest x ray and a Mantoux test, the results of which were unremarkable.

GETTING EVIDENCE INTO PRACTICE

Availability and cost

At this stage, based on Laura's condition, the clinical team's judgement and the available evidence, Dr B was convinced that Laura might benefit from infliximab. He was not sure whether the drug was available from the hospital's pharmacy. He approached the pharmacy department for advice. He was told they did not have it in stock but would order it if they got approval from the hospital's medicines management committee. A single dose of infliximab for his patient would cost about £1400 and, depending on her response, she might need further infusions in the future. The chairman of the committee was initially hesitant to approve the use of the medicine because of its cost and the possibility of other physicians demanding it once it was used on one patient. Dr B subsequently contacted him and provided him with more information about the patient's condition and the available evidence. The chairman was convinced about the weight of the evidence and gave his approval.

Patient preferences

Laura and her parents were approached and given full details about the drug and its potential side effects. They asked a number of pertinent questions, mainly on side effects and availability of other options. Laura in particular was keen to

know whether this medicine would allow her to be discharged from hospital, to go back to school, and to go on holidays. They were not unduly worried about the reported side effects and were keen to try infliximab.

Outcome

She received a single dose of intravenous infliximab at a dose of 5 mg/kg. Within a week, her condition had improved dramatically and all her symptoms had begun to settle and her appetite had improved considerably. She was then discharged home. When she was reviewed in clinic three weeks later, she was extremely well and her symptoms had completely settled. No side effects had occurred. She also reported significant improvements in her quality of life as she was going out more, and in her own words "doing what normal teenagers do".

Evidence based policy

Following Laura's treatment the multidisciplinary team of consultants, junior doctors, nurses, and pharmacist had a meeting to decide on a policy for future infliximab usage. They agreed to task Dr B and another senior paediatric registrar to search the literature further, particularly for articles on children. Further searches identified no paediatric RCTs that had compared the effectiveness of infliximab with either placebo or another treatment. However, several paediatric case reports and case series were found which had suggested that infliximab might be effective in children and adolescents with Crohn's disease who are unresponsive to conventional treatment. One recent study of 21 children suggested that infliximab may be safe and effective as short term treatment of medically refractory moderate to severe Crohn's disease in a paediatric population.⁴

Based on the overall evidence from the medical literature, the team had extensive discussions with hospital managers, patient/parent groups, and pharmacists. Issues such as potential benefits versus potential harms, costs, patient preferences, and availability of alternatives were discussed. Following the discussions, an agreement was reached to consider treating patients with moderate to severe active Crohn's disease unresponsive to conventional treatment (enteral nutrition, mesalazine, steroids, and immunosuppressive agents) with a single two hour infusion of infliximab (5 mg/kg). Because of the potential risk of infliximab activating latent tuberculosis, the team agreed to screen patients for latent tuberculosis with a chest x ray and a Mantoux test before infliximab infusion.

A policy statement which included an assessment of the problem, a review of the evidence on infliximab, analysis of the evidence, and estimation of benefits and harms, assessment of patient preferences, and assessment of costs was written up. The team agreed to review the effectiveness of the policy continuously, and to update the policy in a year's time based on any changes in evidence, expected outcomes, alternative treatments, costs, or patient preferences.

DISCUSSION

Integrating available evidence, clinical expertise, and patient's preferences into the care of this patient was very useful and turned out to be beneficial for the patient. The clinical team were initially "overfocused" on the traditional treatments for Crohn's disease and did not consider the possibility of newer effective treatments until Dr B drew attention to the possibility of using infliximab. The acquired knowledge of evidence based practice equipped Dr B with the willingness and skill to search for evidence on this drug and having found it to apply it to the patient, taking into consideration her preferences and the availability and cost of the drug.

The practice of EBM involves a process of life long, self directed learning in which caring for patients creates the need for important information about clinical and other health care issues. Traditionally, the forum for training in critical appraisal has been the "journal club", but while such clubs have been shown to improve reading habits³ and confidence in appraising the medical literature,⁴ they may not improve the application of knowledge to clinical decision making or improve knowledge in a sustained fashion.⁵ The traditional journal clubs may be unsuitable for teaching evidence based decision making because of their exclusive focus on critical appraisal. Furthermore, they usually present EBM as an abstract exercise rather than an integral part of the flow of patient care.¹³ The traditional journal clubs of clinical teams should be adapted by incorporating the five step EBM model¹ including the formulation of answerable clinical questions, searching for evidence, critical appraisal, applicability of evidence, and evaluation of performance into the process of answering day to day clinical questions in order to make these skills more useful to everyday practice. The questions should reflect "real world" constraints faced by the busy clinician and should arise from actual patient care issues. The development and structure of such an EBM journal club has been discussed elsewhere.^{13 14}

The practice and teaching of EBM should be part of the day to day care of patients. When they are built into our ward rounds, outpatient clinics, departmental clinical meetings, and so on, they can allow the whole clinical team to develop an evidence based approach to clinical decision making. Developing the ability to access information from the medical literature, critically appraising it, and applying it to patient care requires skills that need to be taught. Most practising paediatricians were not taught these skills in medical school. EBM training will provide child health professionals with the tools needed to overcome some of the common barriers they

face when trying to use published reports to help solve patient problems.

Competing interests: none declared

REFERENCES

- 1 **Akobeng AK.** Evidence based child health. 1. Principles of evidence based medicine. *Arch Dis Child* 2005;**90**:837–40.
- 2 **Akobeng AK, Zachos M.** Tumor necrosis factor-alpha antibody for induction of remission in Crohn's disease. *The Cochrane Database of Systematic Reviews* 2003: issue 4.
- 3 **Targan SR, Hanauer SB, van Deventer SJH, et al.** A short term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. *N Engl J Med* 1997;**337**:1029–35.
- 4 **Baldassano R, Braegger CP, Escher JC, et al.** Infliximab (REMICADE) therapy in the treatment of pediatric Crohn's disease. *Am J Gastroenterol* 2003;**98**:833–8.
- 5 **D'haens G, van Deventer S, van Hogezaand R, et al.** Endoscopic and histological healing with infliximab anti-tumor necrosis factor antibodies in Crohn's disease: a European multicenter trial. *Gastroenterology* 1999;**116**:1029–34.
- 6 **Critical Appraisal Skills Programme.** Appraisal Tools. Oxford, UK. <http://www.phru.nhs.uk/casp/reviews.pdf> (accessed 10 Dec 2004).
- 7 **Clarke M, Oxman AD, eds.** *Cochrane Reviewers' Handbook* 4.1.4 (updated October 2001). In: *The Cochrane Library*, Issue 4, 2001. Oxford: Update software.
- 8 **Jadad AR, Moore RA, Carroll D, et al.** Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;**17**:1–12.
- 9 **Akobeng AK.** Understanding measures of treatment effect in clinical trials. *Arch Dis Child* 2005;**90**:54–6.
- 10 **Linzer M, Brown JT, Frazier LM, et al.** Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills. A randomised controlled trial. *JAMA* 1988;**260**:2537–41.
- 11 **Kellum JA, Rieker JP, Power M, et al.** Teaching critical appraisal during critical care fellowship training: a foundation for evidence-based critical care medicine. *Crit Care Med* 2000;**28**:3067–70.
- 12 **Neville AJ, Reiter HI, Eva KW, et al.** Critical appraisal turkey shoot: linking critical appraisal to clinical decision making. *Acad Med* 2000;**75**:S87–9.
- 13 **Green ML.** Evidence-based medicine training in graduate medical education: past, present and future. *J Eval Clin Pract* 2000;**6**:121–38.
- 14 **Elnicki DM, Halperin AK, Shocckor WT, et al.** Multidisciplinary evidence-based medicine journal clubs: curriculum design and participants' reactions. *Am J Med Sci* 1999;**317**:243–6.