

Original article

The promise of and problems with evidence-based medicine for paediatric asthma management

The concept of 'evidence-based medicine' has now been in widespread use in clinical practice for over a decade. There are different types of clinical study, which may provide evidence on which to base clinical decisions, but some are much less robust than others. In making decisions about treatment, one of the highest levels of evidence in primary research is the randomized controlled trial. This study design has been used in clinical research for over 50 years. Systematic reviews of randomized controlled trial are scientific studies that review, critically appraise and, where appropriate, aggregate results from a number of different randomized controlled trials. They are increasingly being used throughout health care to provide guidance about treatment. Under the auspices of The Cochrane Collaboration, systematic reviews of randomized controlled trials are being conducted across the whole of health care. One of the most active areas for this work is within the field of asthma and clinicians are now able to access a large number of different systematic reviews on The Cochrane Library. If used appropriately, they can aid the clinician in making decisions about individual patients and provide a sound evidence base from which clinical guidelines can be developed.

R. L. Smyth¹, P. M. O'Byrne²

¹University Institute of Paediatric Health, Alder Hey Children's Hospital, Liverpool, UK; ²Asthma Research Group, St. Joseph's Hospital and McMaster University, Hamilton, Ontario, Canada

Key words: asthma; evidence-based medicine; randomized controlled trial; systematic review.

Professor R. L. Smyth
University of Liverpool
Institute of Child Health
Alder Hey Children's Hospital
Liverpool L12 2AP
UK
E-mail: r.l.smyth@liv.ac.uk

The term "evidence" is an adaptation of the word *evidentiam* invented by Cicero, originating from the Greek word *energeia*, meaning brightness or clarity. Interestingly, Cicero used the word to indicate that no proof or explanation was needed because a fact was self-evident. Subsequently, Descartes used the term to mean clear and distinct, so much so that it was immune to doubt. Only in the 17th century did the term evidence come to mean something pointing beyond itself (1).

One widely accepted definition of evidence in medical sciences is "the best clinically relevant research on diagnosis, prognosis, therapy, or prevention" (2). Clinical evidence is obtained most often through a rigorous process of randomized clinical trials. A reliance on this methodology is regarded as the mainstay of good decision making in clinical practice. It is, however, a challenge for a clinician to cope with the vast increase in the amount of information generated and published about asthma and, indeed, all other clinical conditions. The development of systematic reviews over the past 20 years was undertaken in an attempt to help clinicians deal with the information overload, to assist in the effort to make the best decision in any clinical situation.

History of evidence-based medicine

In 1979, the British epidemiologist Archie Cochrane observed that "it is surely a great criticism of our

profession that we have not organized a critical summary by specialty or subspecialty adapted periodically of all relevant randomized controlled trials" (3). It was in response to this challenge that an international body known as the Cochrane Collaboration was established, whose function will be described later in this article. About the same time, the term "evidence-based medicine" was coined by David Sackett and his colleagues to mean "a conscientious, explicit and judicious use of current best evidence in making decisions in the care of individual patients" (4). The methodology for conducting meta-analysis was established and utilized in the late 1970s (5). It was, however, in the early 1990s that a working group (the Evidence Based Medicine Working Group), established at McMaster University in Canada, developed the idea and began the dissemination of the concept of a paradigm change in the practice of medicine using evidence-based medicine (6).

Levels of evidence

There are many different types of evidence which we may use to inform our decision making. When considering the evidence for the benefit or otherwise of our therapy, the different types of evidence are shown in Fig. 1, in which we have ranked the evidence according to its validity or robustness. At the bottom of the hierarchy is the anecdotal case report. On its own, a case report does

Systematic Review of Randomized Controlled Trials
 Confirmed Randomized Controlled Clinical Trials
 Single Randomized Controlled Clinical Trial
 Non-randomized Controlled Clinical Trial
 Case Controlled Observational Studies
 Analysis of Large Computer Databases
 Case Series with Historical Controls
 Case Series, Literature Control
 Uncontrolled Case Series
 Anecdotal Case

Figure 1. Research pyramid of study designs used to assess the efficacy of treatments.

not provide strong evidence for the efficacy of treatment, but may describe the outcome of treatment in an individual patient. Many case reports have been the trigger to important clinical research and therefore should not be discarded as useless (7, 8). Medical history contains many examples of case reports or case series, which have been important milestones in the development of new treatment strategy. Perhaps one of the most famous is Edward Jenner's initial case studies (9), which reported that inoculated cowpox was a safe alternative to inoculated smallpox for the prevention of smallpox. While 200 years ago the medical fraternity were prepared to base large-scale treatment strategies on uncontrolled case studies, nowadays Jenner's observations would have also been considered important but would have led quite correctly to a series of stepwise (phase 1 to phase 3) clinical trials (10). Therefore, individual observation(s) are an important element of clinical research, but should be considered as hypothesis-generating rather than hypothesis-testing.

Some case series attempt to place these reports in the context of "controls" and may report them with reference to either literature or historical controls as a comparison. There are dangers in using noncontemporaneous controls because other elements of the treatment for a particular disorder may change rapidly with time (11). There are many diseases with a strong (so-called) "cohort" effect. One such disease is cystic fibrosis (12) and comparing treatment strategies even in one 5-year period with those in another 5-year period are fraught with problems. Other such studies, which may be used to evaluate treatment, are surveys of large computer databases and analyses of individuals who have received a treatment compared to those who have not. A case-control study is a retrospective study where individuals are compared with contemporaneous controls. Cohort studies are those in which a group of patients who received a particular treatment are compared contemporaneously with a group who have not. This has the advantage in that usually other elements of the treatment are fairly similar but is subject to the bias that the selection of which patients receive the treatment and which do not may mean that the treatment group and the control group are very different.

The advantage of the randomized controlled trial (RCT) is that allocation of patients to the treatment group or the control group is entirely by chance and therefore the observed differences, at the end of the study, between the treatment and the control group should be due to the treatment alone rather than to patient characteristics, physician preference or other biases.

The RCT is a simple, powerful and revolutionary tool of clinical research (13, 14). The "interventions" which may be studied in a RCT include any clinical manoeuvre which may affect patients' health status and treatments may include preventative strategies, diagnostic tests and screening programmes. The "control" can be no intervention, alternative standard therapy or a placebo. Often more than one intervention group may be used when, for example, the investigators are interested in whether one dose of a particular drug is better than an alternative dose and both are compared to the control. Because RCTs are designed to eliminate imbalance between treatment and control groups, this aspect of trial quality can be investigated by comparing the characteristics of these groups at baseline. There are many different ways of generating a random sequence: for example, by using a random number table or computer. Sometimes investigators allocate participants according to hospital number (odd or even) or date of birth. Because these last two examples do not give each of the participants the same opportunity to be included in each of the study groups, they are not random and are often referred to as pseudo-random or quasi-random allocation.

There are some questions about a treatment, which the RCT is not the best designed to address (15). For example, many drugs may have very rare adverse effects: it would be difficult or impossible to adequately power a randomized controlled trial to detect these. Such effects are often only detected with certainty after a drug has been released onto the market and is available to many thousands of patients. Post-marketing surveillance is a very important element of investigation into drug therapy, but is an observational study design. Where diseases are extremely rare it may be difficult to recruit sufficient patients from one country or even from many different countries to investigate the efficacy of therapy. This may apply to certain inborn errors of metabolism or very rare forms of malignancy. The effect of treatments for a lifelong disease should be evaluated over the patient's lifetime and again this is difficult to carry out within the confines of an RCT.

In 1972 the British epidemiologist Archie Cochrane defined three concepts related to testing health-care interventions (16). First, efficacy assesses whether an intervention does more harm under ideal circumstances than good ("can it work?"). The effectiveness of an intervention is the extent to which it does more good than harm when provided under the usual circumstances ("does it work?"). Finally, the efficiency of an intervention is the effect of that intervention in relation to the

resources it consumes (“is it worth it?”). Trials of efficacy and effectiveness have also been referred to as explanatory and pragmatic trials, respectively (17). This area has been elegantly reviewed recently by Haynes (18). Explanatory trials typically recruit patients in whom the diagnosis has been defined carefully, where the treatment is administered in a very controlled situation and strenuous measures are taken to ensure and assess that the patients are compliant with the treatment. The outcomes that are assessed are frequently not clinical outcomes but are often laboratory-based outcomes to answer the “can it work?” question (19). For example, in a trial to investigate whether influenza vaccine is beneficial to patients with asthma, one may first of all wish to evaluate whether those patients who receive the immunization had good antibody levels in their serum to influenza after immunization. This is not because patients, or their clinicians, think that measuring antibody levels is important in assessing the benefit of treatment, but rather that if the vaccine does not produce this effect then it is unlikely to lead to other very important clinical effects such as reduction in the number of asthma exacerbations. Consequently, explanatory trials are frequently smaller and evaluate outcomes that are related to important mechanisms of action. A pragmatic clinical trial would recruit patients from a variety of settings with a range of clinical severities of a disorder. In the example given above, the patients would be randomized to either receive or not receive the vaccine and their outcomes would be analysed according to the groups to which they were randomized (known as an intention-to-treat analysis) (19).

The rationale behind the difference between the explanatory and pragmatic trials is that if the therapy is not shown to work in an explanatory trial then it usually will not work under normal conditions. In essence, many trials contain elements of both explanatory and pragmatic design and almost all trials evaluate efficacy. However, we should not lose sight of the question that what most clinicians and patients want to know is the answer to the efficiency question. That is, do the benefits, if any, justify any adverse effects, the costs or the inconvenience of the therapy?

Over the past 10 years we have seen many studies which have evaluated RCTs as the subject rather than the tool of clinical research. Many of these studies have evaluated the quality of RCTs (14). Thus, although random allocation is an important tool to minimize bias, there are many other sources of bias which may occur in a clinical trial. Selection bias may be introduced if some eligible individuals are excluded from the study because of prior knowledge of the group to which they will be allocated if they participated in this study (20). There is an opportunity for this form of bias if the randomization sequence is not concealed from the investigators at the time of obtaining consent from prospective trial participants. For example, if a doctor felt strongly that a particular therapy would be of benefit to those patients

with asthma who were having frequent exacerbations, he may prefer to enter patients into the treatment arm only if they experienced frequent exacerbations. If the investigator could predict whether or not the patient was going to be randomized to the treatment arm, he may choose not to enter such a patient to the study if they were going to be allocated to the control group. An important study by Schulz et al. (21) has shown that trials with inadequate allocation concealment can exaggerate the effects of interventions by as much as 40% on average. Allocation concealment helps to prevent selection bias before and until the interventions are given to the study participants and can always be introduced into trial design. There are many ways in which randomization can be subverted by investigators. For example, even when the allocation codes are kept in sealed opaque envelopes investigators may be able to look through the envelope using a strong light (22). However, despite its importance, other work has shown that < 10% of articles describing RCTs report allocation concealment (23).

Another type of bias, known as ascertainment bias, occurs when the conclusions of a trial are distorted by knowledge of which intervention the participants are receiving. The best way to protect a trial against ascertainment bias is by introducing masking (or so-called “blinding”) into the trial design (24). This can be performed whereby the investigators do not know the group to which the patient has been allocated (single-blind) or where both the investigators and the patients are unaware to which arm they have been allocated (double-blind). Blinding helps to prevent ascertainment bias and protects the randomization sequence after allocation, but for certain interventions (e.g. physiotherapy) cannot always be implemented in a double-blind manner. Even where it is possible, blinding is frequently not used, and may not be reported adequately (25). The importance of blinding has also been confirmed in empirical studies; for example, it has been shown that studies which are not blinded are more likely to favour experimental interventions over controls (24).

Bias can also be introduced by mishandling dropouts, withdrawals of protocol evaluation. In many trials there is missing information. The strategy which is used to try to eliminate such bias includes what is described as “intention-to-treat analysis”. This means that all study participants are analysed according to the groups to which they are randomized, regardless of whether or not they completed the study. However, in trial reports the analysis that is performed frequently is that which presents the treatment in the light that the investigators would wish it to be seen. Often patients are analysed according to “completed treatment” groups.

One of the problems with trying to assess the quality of a trial is that important elements of its design may not have been reported adequately, if at all. The 1996 guidelines, called CONSORT, were presented for the reporting of RCTs and these have been adopted by

several leading journals (26). This group developed a checklist of 21 items, including descriptions of the randomization procedure, the mechanisms of blinding/masking, number of people lost during the follow-up and some details about the analysis. If these important aspects of methodology are not described explicitly then it is unsafe to assume that they were performed adequately. The abstracts of RCT reports should ideally be structured and describe the objective, research design, clinical setting, participants, interventions, main outcome measures, results and then conclusions systematically (27). As well as describing the important methodological elements (which have been referred to previously) these studies should report carefully the inclusion and exclusion criteria as well as the details of the interventions in sufficient detail to enable the reader to repeat the study, should they wish to. Outcomes should be described and it should be stated what the primary outcome was and which outcomes were regarded as secondary. There is much scope for selective reporting of outcomes. This situation may arise where some outcomes showed statistical significance and others did not and bias may be introduced where the investigators report selectively those outcomes that were statistically significant between the groups. The power calculation for the study should be stated. This is based on what would be generally regarded as a *clinically* significant difference between the treatment and control groups for trials that are powered to show a difference between these groups. Often trials are under-powered for reasons which relate to the inability of the investigators to recruit sufficient patients. This should be evident if the investigators have presented a power calculation and stated clearly the number of patients that they recruited.

As is clear from the preceding discussion, many trials are of poor quality or recruit insufficient patients to show a difference between therapies even if such a difference exists. It may be, then, that a rigorously performed observational study may provide a higher level of evidence than a poorly performed and under-powered RCT. However, even if an RCT has been performed well, it is unlikely that a clinician would base treatment decisions on the results of a single trial. This is because that despite how interesting, relevant, unbiased and well-reported an individual trial may be, such a trial is likely to be just one among many other studies that are addressing the same question which may contradict or corroborate the findings of the trial. It is therefore important to set the results of individual trials in the context of other relevant studies. If a number of clinical trials show the effect of a treatment to be in the same direction in each (a situation which arises infrequently) that may provide overwhelming evidence on which one would feel able to base treatment decisions. More frequently, however, one needs to go to a different and higher level of evidence, which is the systematic review of RCTs.

Systematic reviews

The process of evidence-based medicine involves four essential steps (Table 1). The first step is to formulate (or ask) the clinically relevant question. Key elements of the well-formulated clinical question included a description of: 1) the patient or problem; 2) the intervention (whether by nature or clinical design); 3) a comparison intervention, when relevant; and 4) the clinical outcome or outcomes of interest (2). One of the problems is that having formulated the question, steps 2 and 3 can be extremely time-consuming. There are 2 million articles published annually in the biomedical literature in over 20 000 journals (28). Within the relatively small area of cystic fibrosis we have found that relevant RCTs may be published in nearly 100 different medical journals (29). It is important that the average clinician can access all of this information, but much of it may be invalid or irrelevant and finding and critically appraising the evidence from well-conducted studies is a challenging task. To do this efficiently and to integrate best evidence into the clinical care of patients the clinician needs to develop shortcuts. In systematic reviews, the literature searching and critical appraisal have been carried out by the reviewers according to a methodology which is described explicitly in the systematic review. Systematic reviews provide summaries or shortcuts to enable the clinician to obtain answers to his/her questions in a timely fashion. Such approaches “have the potential to rescue us from sinking in a sea of papers” (30), as T. S. Eliot has asked in his poem “The Rock”: “Where is the knowledge that we have lost in information?”

As has been discussed previously, investigators undertaking primary research studies should make strenuous efforts to ensure that bias is eliminated. It is recognized increasingly that rigorous scientific methodology is also necessary when reviewing a number of RCTs or any other source of evidence as it is in performing RCTs in the first place (31). The use of the term *systematic* (also called *scientific* reviews) is to distinguish them from more *traditional* or *narrative* reviews of topic areas conducted by experts in the field. Such narrative reviews may be very valuable in providing an overview of a field, but as pieces of scientific work they have a large potential for bias. Authors may cite studies selectively which support their own predetermined opinion and fail to cite other studies which provide evidence which conflicts with their opinion.

Guidelines and checklists have been provided to enable readers to assess the quality of a review article (32)

Table 1. Four steps in practising evidence-based medicine

1	Ask a clear clinical question from a patient's problem
2	Search the literature to identify the best external clinical evidence
3	Evaluate (critically appraise) available evidence for its validity and relevance
4	Apply the evidence in clinical practice

Table 2. Identify sources of bias in systematic reviews

Asking the question
Is the question clearly focused?
Study identification
Has a comprehensive search been performed for all relevant studies?
Study selection
Are inclusion criteria appropriate?
Study appraisal
Has the quality of included studies been adequately assessed?
Data collection
Is all relevant information included and has missing information been obtained from investigators?
Data synthesis
How sensitive are the results to changes in the way the review is performed?
Interpretation of results
Are the conclusions justified by the results of the review?
Are recommendations linked to the strength of evidence?
If there is "no evidence of effect" have the reviewers avoided interpreting this as "evidence of no effect"?
Are subgroup analyses interpreted cautiously?

(Table 2). As with any other piece of research, a reviewer needs to write a protocol which states, first, the questions or hypotheses which they wish to investigate. Secondly, a comprehensive search strategy should be described. The search strategy is likely to include searching bibliographic databases such as Medline. However, electronic database searching may not be sufficiently comprehensive to identify all potentially relevant RCTs. In the past, work has identified that Medline searches have yielded only around 50–60% of potentially relevant RCTs (33, 34). Other techniques which can be used to improve the yield of potentially relevant studies are hand-searching of health-care journals and searching for unpublished RCTs.

Searching for unpublished RCTs is important because empirical evidence has demonstrated that publication bias operates (35–37). This means that studies which are small or do not show a so-called "positive" effect are less likely to be published than those that are large or do show a positive effect (38–40). Exclusion of such studies would lead to bias in the results of the review. Other sources of bias may be introduced if the reviewers choose to search only for RCTs which are published in English language journals (23). Publication bias may be a factor in the systematic differences found between studies where the investigators have affiliations with industry and those where the investigators do not have such affiliations (41).

The inclusion of unpublished work in systematic reviews has been criticized on the basis that such studies are more likely to be of poor quality. This is a most important point, because the validity of a review depends on the quality of studies which are included. This emphasizes the importance of stating *a priori* the inclusion and exclusion criteria and applying a rigorous methodological quality assessment to all studies both published and unpublished. Such methodology ensures that unsuitable studies are not included, or if included, techniques can be introduced to

evaluate whether the inclusion of poor-quality studies has affected the results of the meta-analysis. However, it does ensure that as far as possible publication bias does not affect the results of the review.

One of the difficulties is that an abstract report may contain insufficient information to provide an assessment of methodological quality. However, if the CONSORT (26) guidelines are followed and abstracts for RCTs are structured (27), more relevant detail would be included in abstract reports to enable them to be assessed for inclusion in a review.

The protocol should state what studies will be considered eligible for inclusion in the review and this should be determined according to types of study, types of participant and types of intervention. The protocol should also state what outcomes are of interest and should state all potentially relevant outcomes, particularly those which are relevant to the patient.

There are now a number of assessment tools which have been validated for assessing the quality of RCTs (21, 42). These tools can be used to score the quality of the studies included. Having determined which study should be included, the reviewers then need to proceed to data extraction and analysis of data. Ideally, deciding study eligibility, assessing study quality and data extraction should be performed independently by two reviewers to minimize bias. Disagreement between reviewers could then be resolved by asking a third reviewer to arbitrate. Empirical research has shown that bias does not occur if the authors and institutional affiliations of the primary studies are known to the reviewers, so making strenuous efforts to blind reviewers to the names of authors and their institutions is not necessary (43).

In a review, as in a primary study, statistical methods, referred to in this case as meta-analysis, could be used to analyse and summarize the data. These can be used to provide a combined effect estimate to help answer the question "does this intervention have a beneficial effect?" and if so "what is the size of that effect?" The same questions can be asked about harmful effects. If statistical methods are not used to aggregate data in systematic reviews there may be a tendency for reviewers to "vote count", this means that they may simply add up the numbers of trials showing benefit and compare it with numbers of trials which show no benefit or which show harm from an intervention. This simplistic approach has many problems, the most obvious of which is that it applies equal weighting to all studies, ignoring differences in the size and quality of individual studies.

However, a systematic review which does not include meta-analysis may be as valuable as one which does. Often there is a lack of relevant or valid data to include in a meta-analysis. There may be such heterogeneity between the studies that it is inappropriate to combine the data. The outcome measure in different studies, although similar, may have been expressed differently or measured at different time points which renders it inappropriate to

combine them. Conducting meta-analysis when it would be inappropriate or misleading is worse than not conducting a meta-analysis but drawing quality out of the conclusion from a review.

The Cochrane Collaboration

The need to provide up-to-date summaries of evidence to enable decisions to be made about health-care has been recognized for some time and it was the British epidemiologist Archie Cochrane who observed in 1979 that “it is surely a great criticism of our profession that we have not organized a critical summary by specialty or subspecialty adapted periodically of all relevant randomized controlled trials” (3). It was in response to this challenge that an international body known as the Cochrane Collaboration was established. The Cochrane Collaboration is dedicated to producing systematic reviews of the effects of health-care interventions using RCTs as the building blocks for this form of research synthesis. One of the key elements in the Collaboration’s work is that systematic reviews should not only be prepared but should be maintained and updated so that they continue to be relevant to the needs of decision makers. This task, when considered in the context of all of health-care, is huge and has been likened in scale and importance to that of the Human Genome Project (44). It has nine key principles (Table 3). The entities within the Cochrane Collaboration which produce systematic reviews are known as Collaborative Review Groups. These entities are made up of individuals showing an interest in a particular health-care problem, e.g. epilepsy, inflammatory bowel disease, incontinence and so on. There are at least three Collaborative Review Groups whose interests are relevant to those of paediatric pulmonologists. They are the Airways Group, the Cystic Fibrosis and Genetic Disorders Group

and the Acute Respiratory Infections Group. Each Collaborative Review Group has its own international network of collaborators who participate not only in preparing Cochrane reviews but also in other activities that help the group to fulfil its aims, such as hand-searching journals, peer review of protocols and reviews.

A consumer is either someone who has personal family experience of a condition under consideration, or who as a formal representative of a consumer group has access to other consumers’ experience and knowledge. Consumers are involved in all collaborative review groups and provide an important perspective which ensures that reviews produced are relevant to the needs of people with the disease. Many groups have a panel of consumers who peer-review all protocols and reviews. Each Collaborative Review Group is coordinated by an editorial team which is responsible for updating and submitting regularly an edited module of protocols, reviews and information about the group which is published in the *Cochrane Library* (45).

In addition to Collaborative Review Groups, there are currently 15 Cochrane Centres dotted throughout the world; they help to coordinate and support the collaboration and each has a specific responsibility. For example, the Australasian Cochrane Centre has undertaken to develop and maintain the consumer network. Each centre is responsible for providing support within a specified geographic area. A further entity within the Collaboration is known as “Field”; these focus on dimensions of health-care other than health problems such as the setting of care (e.g. primary care), the type of consumer (e.g. child health) or the type of intervention (e.g. physical therapy). Individuals working in fields help to ensure that priorities and prospective in their field of interest are reflected in the work of collaborative review groups. They may compile specialist databases of clinical trials and hand search specialist journals. Finally, Methods Working Groups are entities in the Cochrane Collaboration made up of individuals who are interested in the judgement that lead to selection appraisal synthesis interpretation and dissemination of health-care information. Examples of these are the statistical methods working group, the nonrandomized controlled trials methods working group and the screening and diagnostic tests methods working group.

The *Cochrane Library* (45) is an electronic publication available on CD ROM which is updated quarterly. It contains the *Cochrane Database of Systematic Reviews* which contains all the reviews and protocols prepared by individual collaborative review groups. It also contains the *Cochrane Controlled Trials Register*, which is the largest electronic database of RCTs. The *Database of Abstract of Reviews of Effectiveness* is a database of other systematic reviews other than Cochrane reviews which have been appraised critically. This appraisal is available for readers of the database. The fourth database including the *Cochrane Library* is the *Cochrane Review Methodology Database*.

Table 3. Nine key principles of Cochrane Collaboration

- | |
|---|
| 1 Collaboration, by internally and externally fostering good communications, open decision-making and teamwork |
| 2 Building on the enthusiasm of individuals, by involving and supporting people of different skills and backgrounds |
| 3 Avoiding duplication, by good management and coordination to maximize economy of effort |
| 4 Minimizing bias, through a variety of approaches such as scientific rigour, ensuring broad participation, and avoiding conflicts of interest |
| 5 Keeping up to date, by a commitment to ensure that Cochrane Reviews are maintained through identification and incorporation of new evidence |
| 6 Striving for relevance, by promoting the assessment of health-care interventions using outcomes that matter to people making choices in health care |
| 7 Promoting access, by wide dissemination of the outputs of the Collaboration, taking advantage of strategic alliances, and by promoting appropriate prices, content and media to meet the needs of users worldwide |
| 8 Ensuring quality, by being open and responsive to criticism, applying advances in methodology, and developing systems for quality improvement |
| 9 Continuity, by ensuring that responsibility for reviews, editorial processes and key functions is maintained and renewed |

Examples of systematic reviews in paediatric asthma management

At the time of writing, the current disk issue of the *Cochrane Library* is Issue 1, 2000 (45). A search of this disk for systematic reviews in which the word "asthma" is used in the abstract yielded 32 reviews. In addition there were 58 protocols which mentioned the word "asthma". These protocols should become reviews within the ensuing 12 months.

In considering how to use systematic reviews we will illustrate this with a clinical scenario (Box 1). This scenario described a 5-year old girl with asthma who attends the accident and emergency department with acute onset of wheeze. Her peak expiratory flow rate is recorded as 60% of predicted and this is her first attempt at performing this. The attending doctor is faced with the question as to whether he should prescribe oral steroids. First, one needs to focus the question as described previously. In this particular scenario the question which may be posed is "in a child with diagnosed asthma, who presents with acute onset of wheeze and reduced lung function, does treatment with steroids, compared with no steroid treatment reduce the risk of hospital admission, reduce relapse rate, improve lung function, or lead to adverse effects?" A search of the *Cochrane Library* using the terms "asthma" and "corticosteroids" will include among the hits the review entitled "Corticosteroids for preventing relapse following acute exacerbations of asthma" (46). This review sought to include studies of patients presenting to an emergency department or other acute care setting. Studies recruiting children and/or adult participants were included and patients were randomized to receive either corticosteroids (by the oral, intramuscular or inhaled route) or placebo. Studies comparing two different corticosteroid regimens were also included. Seven studies were found to be suitable for inclusion.

The results showed that steroid treatment was associated with significantly fewer patients relapsing within 7–10 days and within 21 days and significantly fewer patients requiring subsequent hospital admission. No differences were found with pulmonary function tests and there were similar rates of side effects in both groups. The authors calculated the "numbers needed to treat" (47) for prevention of relapse and need for hospitalization. They calculated that, on average, 13 patients would require treatment with corticosteroids to prevent one relapse to additional care in the first 7–10 days after outpatient care

for an exacerbation. On average, 16 patients (95% CI, 7–125) would require treatment with corticosteroids to prevent one relapse to hospitalization after outpatient care for an exacerbation. The authors concluded that their results strongly supported the use of oral or intramuscular corticosteroids for treatment of outpatients released from the acute care setting with an exacerbation of asthma. It would appear that in the scenario used here that this child would benefit from oral steroid treatment.

In this example the reviewers used the "number needed to treat" concept as a way of expressing their results (47). Presenting results in this way has the attraction that most lay people and most clinicians can understand the statement "on average 16 patients would require treatment with corticosteroids to prevent one patient being hospitalized for relapse", whereas they may have more difficulty in understanding an odds ratio. It should be emphasized, however, that such a statement refers to the "average" patient rather than the individual patient. The precision of the estimate should be provided where possible by, for example, stating the confidence intervals for the number needed to treat (48).

In the review discussed above, it was stated at the outset that children and adults would be studied separately in a subgroup analysis. However, it would appear that there were insufficient data available to be able to analyse children separately from adults. There is now growing evidence that in children with asthma the mechanisms of drug action may be different from adults (49). The majority of clinical trials which have been undertaken in people with asthma have been conducted in adults, so we need to be careful about extrapolating directly from studies in adults when making decisions about care in children with asthma. The general practice within the *Cochrane Airways Group* for their reviews published in the *Cochrane Library* has been to consider all ages of participants within a review, but to conduct a separate subgroup analysis based on children where this is possible. This has been undertaken, for example, in a review entitled "Holding chambers vs. nebulisers for β_2 agonist treatment of acute asthma" (50). This review included 16 trials and a total of 686 children and 375 adults. The objective of the review was to assess the effects of holding chambers compared to nebulizers for the delivery of β_2 agonist for acute asthma. None of the outcomes examined was significantly worse with the holding chambers for either children or for adults. When the time spent in the emergency department was examined there was significant heterogeneity when the results from adults and children were pooled and this could not be explained on the basis of methodological quality. There was, however, no significant heterogeneity when adults and children were analysed separately. Children treated with holding chambers compared with those treated with nebulizers spent less time in the accident and emergency department and had less increase in pulse rate. All other outcomes (hospital admission rates, lung

Box 1. Clinical scenario: Clare, aged 5, attends A&E department

From the age of 3 Clare has had intermittent wheeze with occasional nocturnal cough and has been diagnosed with asthma. These episodes have been managed at home with salbutamol via a volumatic and she has not required any hospital admissions. Her symptoms have worsened in the past 2 months. On examination her chest was hyperinflated with widespread expiratory wheeze. PEFR was 60% of predicted
Would you prescribe oral steroids?

function, respiratory rate) were similar between the two groups. The authors concluded that for children, in most cases holding chambers could be substituted for nebulizers to deliver β_2 agonists in acute asthma in emergency departments.

In the first review, used as an example, it was not possible to look separately at the results for children and therefore results for children and adults were aggregated. The second review was able to examine separately the results for children and this yielded important differences. By undertaking reviews which include both children and adults it is possible to evaluate the overall effect of the therapy across the age groups. Where insufficient children have been studied in clinical trials this is the best information available. Therefore, the paediatrician can go to a review and if there are no separate data included for children s/he can access the next best thing, which is the summary of the treatment effects across the age groups. Future updates of the review will hopefully be able to assess outcomes in children when sufficient data are available.

Problems with systematic reviews

As described above, RCTs have become the mainstay of obtaining evidence to use making appropriate management decisions. There are, however, limitations in depending *only* on this generally very reliable methodology. These include:

- The importance of astute clinical observation in identifying clinical developments. A good example was the identification, in 1981, of the association of *Pneumocystis carinii* infection in five homosexual men (51), which has resulted in the search for cure for the one of the most important medical problems of the century.
- The necessity to make important clinical decisions in the absence of complete evidence. In 1993, the United States blood banks began to screen blood products from high-risk donors, despite almost no evidence, and certainly no RCT, that AIDS could be transmitted through blood products. By contrast the Canadian Blood Banks were waiting for conclusive evidence of this association before taking these steps to aggressively screen blood products (52). This delay has now been criticized very severely, and has resulted in extensive morbidity and mortality (52).
- The lack of agreement in the interpretation of even well-conducted randomized clinical trials, leading to differences in recommendations for treatment. This has resulted in differences in the consensus recommendations for several diseases such as hypertension, where differences existed in recommendations between British and American experts on the treatment of diastolic blood pressure >90 mmHg. This is also true for

paediatric asthma, where some consensus documents recommend the early use of inhaled corticosteroids for the treatment of paediatric asthma (53) and others do not (54), despite access to the same studies.

- Randomized clinical trials in asthma have provided contradictory evidence. A good example is the controversy about the safety of regular inhaled β_2 -agonist use, where several studies indicated a detrimental effect (55, 56) while others did not (57, 58). There are many reasons for this including differences in the outcome variables chosen, the patients studied, the duration of the studies, the methods of analysis and the interpretation of the results. The studies, however, have resulted in different experts providing different advice about β_2 -agonist use in asthma.

The use of systematic reviews provides an opportunity to access clinically relevant information rapidly, and also helps to clarify some of the problems with relying on RCTs alone as identified above. However, there also exist some limitations with systematic reviews at the present time which may limit their value in improving paediatric asthma management. Systematic reviews, when conducted correctly as already described in this article, make every effort to minimize bias, maximize precision and provide a careful description of how information was collected and interpreted. Unfortunately, some of the time this rigour is not applied, which can lead to even more confusion for the clinician than existed previously because of the widespread belief in the methodological stringency applied in developing systematic review, as described above.

This issue was highlighted in a recent critical evaluation of all systematic reviews and meta-analyses on the treatment of asthma. Jadad et al. (59) identified 50 systematic reviews and meta-analyses, $>50\%$ published in the past 2 years, 38 in peer-reviewed journals and 12 in the *Cochrane Library*. Forty reviews had major to extensive flaws. All six associated with the pharmaceutical industry were in this group. The most rigorous were published in the *Cochrane Library*. The flaws identified included:

- Lack of information on methods used to find evidence on the primary question.
- Lack of a comprehensive search for evidence.
- Lack of reporting of criteria used to include studies.
- Inclusion of bias in study selection.
- Lack of reporting of the assessment of the validity of studies included.
- Lack of assessment, using appropriate criteria, of the validity of the studies included.
- Lack of reporting of the methods used to combine the findings.
- Inappropriate combination of the findings of the studies included.
- Inappropriate conclusions reached from the data reported.

The overall median quality score (using a standardized index described by Oxman and Guyatt (60) of all 50 reviews was 3, from a maximum score of 9. The median for the 12 Cochrane reviews was 6, while for those published in peer-reviewed journals and associated with industry was 2. The authors concluded that these methodological flaws limit their value to guide decision making. However, unfortunately many of these flawed systematic reviews are being used to promote one approach to clinical management over another.

Another problem that exists with systematic reviews is the inability of nonexpert readers of the reviews to be able to identify the flaws that may exist, which would make the conclusions reached in the review untenable. This problem must exist even for some peer reviewers and journal editors who did not identify the limitations of those systematic reviews on asthma, published in peer-reviewed journals yet containing serious methodological flaws, resulting in a median quality score of 2 in the article by Jadad et al. (59)

Finally, easy access to the information contained in the *Cochrane Library*, and other repositories of systematic reviews, to utilize for clinical decision making, is a barrier for many clinicians, particularly those not adept in, or without easy access to computers.

The future

The development and dissemination of practice guidelines in both paediatric (53, 54), and adult asthma (61) have been useful attempts to change clinical practice. Until recently, none of the asthma consensus guidelines published from 1989 were evidence-based. The most recent iteration of the Canadian Asthma Treatment Guidelines (53) were developed along evidence-based medicine principles and the recommendations contained in the document ranked the quality of evidence used to establish the recommendation. The Global Initiative on Asthma (61) is currently revising its recommendations along similar lines. It is to be hoped that all future paediatric and adult treatment guidelines for pulmonary diseases will follow this lead.

Another important development for the future of evidence-based medicine will be the application of the

rigorous methodology described above in the development of systematic reviews and meta-analyses, and ensuring that these methods are explicitly described in any publication being submitted to the Cochrane Collaboration or otherwise for peer review. The lack of this information in many of the current systematic reviews on asthma, as identified by Jadad et al. (59), diminishes the impact and value of these more rigorously completed reviews. In addition, reviewers of these systematic reviews for peer-reviewed journals, some of whom are not expert in the methodology required to ensure a well-performed systematic review, should be provided with as much information as possible by the journal to allow as critical a review as possible. It is to be hoped that the standards that are applied by the individuals who are expert in the methodology of systematic reviews will be applied to the development of all systematic reviews in the future to ensure that these approaches are useful to clinical decision making. Unfortunately, the lack of methodological rigour in many of the systematic reviews published on asthma and its management at the present time adds only confusion, not clarity, to decision making.

One of the greatest values of systematic reviews is the identification, usually at an early part of the review process, of gaps in the knowledge base needed to arrive at a sound and clinical useful recommendation about an intervention. This allows the research community to develop and extend hypotheses and to undertake new studies, the results of which will provide further information to support or refute the original opinions of the systematic review.

The formal evaluation of evidence around clinical issues that is termed "evidence-based medicine" is here to stay. The methods used to collect and evaluate the evidence is constantly improving and the resource provided by the *Cochrane Library* is invaluable to clinicians who are unsure about the quality of evidence when making a clinical decision. The tools needed to access the databases are becoming much more available and easier to use, thereby making the information almost instantly available when needed. All future medical advances will be evaluated in this way, just as the standards of practice of the past are have been and are being currently critically appraised.

References

1. AUCLAIR F. On the nature of evidence. *Ann RCPSC* 1999;**32**:453-455.
2. SACKETT DL, RICHARDSON WS, ROSENBERG W, HAYNES RB. Evidence-Based Medicine: How to Practice and Teach EBM. Edinburgh: Churchill Livingstone, 1997.
3. COCHRANE AL. A Critical Review, with Particular Reference to the Medical Profession. In: TEELING-SMITH, G, ed. *Medicines for the Year 2000*. London: Office of Health Economics, 1979, 1931-1971.
4. SACKETT DL, ROSENBERG WMC, GRAY JAM, et al. Evidence based medicine; what it is and what it isn't. *BMJ* 1996;**312**:71-72.
5. SMITH M, GLASS G. Meta-analysis of psychotherapy outcome studies. *Am J Psychol* 1977;**32**:752-760.
6. EVIDENCE-BASED MEDICINE WORKING GROUP. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;**268**:2420-2425.
7. JAMES P. Judgement by peers. *Lancet* 2000;**355**:580.

8. CHARLETON R. Value of anecdotes. *Lancet* 2000;**355**:1372.
9. JENNER E. The effects of cutaneous eruptions on modifications of the vaccine vesicle. *Med J* 1804;**12**:97–102.
10. LEVINE M. The legacy of Edward Jenner [editorial]. *BMJ* 1999;**312**:1177–1178.
11. SACKS H, CHALMERS TC, SMITH HJ. Randomized versus historical controls for clinical trials. *Am J Med* 1982;**72**: 233–240.
12. ELBORN J, SHALE D, BRITTON J. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* 1991;**46**:881–885.
13. SILVERMAN M, CHALMERS I. Sir Austin Bradford Hill: an appreciation. *Control Clin Trials* 1992;**13**:100–105.
14. JADAD AR, RENNIE D. The randomized controlled trial gets a middle-aged checkup. *JAMA* 1998;**279**:319–320.
15. BLACK N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215–1218.
16. COCHRANE AL. Effectiveness and Efficiency: Random Reflection on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.
17. ASHBY D. Can iron supplementation improve cognitive functions? *Lancet* 1996;**348**:973.
18. HAYNES B. Can it work? Does it work? Is it worth it? [editorial]. *BMJ* 1999;**319**:652–653.
19. MURRAY G. Statistical aspects of research methodology. *Br J Surg* 1991;**78**:777–781.
20. CHALMERS TC, CELANO P, SACKS H, SMITH H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;**309**:1359–1361.
21. SCHULZ K, CHALMERS I, HAYES R, ALTMAN DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effect in controlled clinical trials. *JAMA* 1995;**273**:408–412.
22. SCHULZ KF. Subverting randomization in controlled trials. *JAMA* 1995;**274**:1456–1458.
23. MOHER D, FORTIN P, JADAD AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347**:363–366.
24. COLDITZ G, MILLER J, MOSTELLER F. How study design affects outcomes in comparisons of therapy. I. Therapy. *Stat Med* 1989;**8**:441–454.
25. SCHULZ KF, GRIMES D, ALTMAN DG, HAYES R. Blinding and exclusions after allocation in randomized controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;**312**:742–744.
26. BEGG C, CHO M, EASTWOOD S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;**276**:7–9.
27. HAYNES RB, MULROW CD, HUTH E, et al. More informative abstracts revisited. *Ann Intern Med* 1990;**113**:69–76.
28. MULROW CD. Rationale for systematic reviews. *BMJ* 1994;**309**:597–599.
29. CHENG K, SMYTH RL, MOTLEY J, et al. Randomized controlled trials in cystic fibrosis (1966–97) categorized by time, design, and intervention. *Pediatr Pulmonol* 2000;**29**:1–7.
30. BRADLEY F, FIELD J. Evidence-based medicine. *Lancet* 1995;**346**:838–839.
31. MULROW CD. The medical review article: state of the science. *Ann Intern Med* 1987;**104**:485–488.
32. OXMAN AD. Checklists for review articles. *BMJ* 1994;**309**:648–651.
33. DICKERSIN K, SCHERER R, LEFEBRE C. Identification of relevant studies for systematic reviews. *BMJ* 1994;**309**: 1286–1291.
34. CHENG K, PRESTON C, ASHBY D, et al. Time to publication as full reports of abstracts of randomized controlled trials in cystic fibrosis. *Pediatr Pulmonol* 1998;**26**:101–105.
35. EASTERBROOK PJ, BERLIN JA, GOPALAN R, MATTHEWS DR. Publication bias in clinical research. *Lancet* 1991;**337**: 867–872.
36. DICKERSIN K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;**9**:15–21.
37. RENNIE D, FLANAGIN A. Publication bias – the triumph of hope over experience. *JAMA* 2000;**267**:411–412.
38. DICKERSIN K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;**263**:1385–1389.
39. DICKERSIN K, MIN Y, MEINERT C. Factors influencing publication or research results; follow-up of applications, submitted to two institutional review boards. *JAMA* 1992;**263**:374–378.
40. CHALMERS I, ADAMS M, DICKERSON K, et al. A cohort study of summary reports of controlled trials. *JAMA* 1990;**263**:1401–1405.
41. BERO L, RENNIE D. Influences on the quality of published drug trials. *J Technol Assess Health Care* 1996;**12**:209–237.
42. JADAD AR, MOORE RA, CARROLL D, et al. Assessing the quality of reports of randomized controlled trials: is blinding necessary? *Control Clin Trials* 1996;**17**:1–12.
43. BERLIN JA, FOR THE UNIVERSITY OF PENNSYLVANIA META-ANALYSIS BLINDING STUDY GROUP. Does blinding of readers affect the results of meta-analyses? *Lancet* 2000;**350**:185–186.
44. NAYLOR CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995;**345**:840–842.
45. The Cochrane Library. Oxford: Update Software, 2000.
46. ROWE B, SPOONER C, DUCHARME F, et al. Corticosteroids for preventing relapse following acute exacerbations of asthma (Cochrane Review). Cochrane Library, Issue 1, 2000. Oxford: Update Software.
47. CHATELLIER G, ZAPLETAL E, LEMAITRE D, et al. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ* 1996;**312**:426–429.
48. ALTMAN DG. Confidence intervals for the number needed to treat. *BMJ* 1998;**317**:1309–1312.
49. BISGAARD H. Long-acting β_2 -agonists in management of childhood asthma: a critical review of the literature. *Pediatr Pulmonol* 2000;**29**:221–234.
50. CATES C, ROWE B. Holding chambers versus nebulisers for beta-agonist treatment of acute asthma (Cochrane Review). Cochrane Library, Issue 1, 2000. Oxford: Update Software.
51. CENTER FOR DISEASE CONTROL. Pneumocystis pneumonia, Los Angeles. *MMWR* 1981;**30**:250–252.
52. KREVER H. Commission of Inquiry on the blood system in Canada. Ottawa: Minister of Public Works and Government Services Canada, 1997.
53. BOULET L, BECKER A, BERUBE D, et al. 1998 revision of the Canadian Asthma Consensus Guidelines. *Can Respir J* 1999;**6**:231–232.
54. WARNER JO, GOTZ M, LANDAU LI, et al. Management of asthma: a consensus statement. *Arch Dis Child* 1989;**64**: 1065–1079.
55. SEARS M, TAYLOR D, PRINT C, et al. Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* 1990;**336**:1391–1396.
56. CRANE J, PEARCE N, FLATT A, et al. Prescribed fenoterol and death from asthma in New Zealand, 1981–83. *Lancet* 1989;**1**:917–922.
57. CHAPMAN K, KESTEN S, SZALI J. Regular vs. as-needed inhaled salbutamol in asthma control. *Lancet* 1994;**1994**: 1379–1382.
58. DRAZEN J, ISRAEL E, BOUSHEY H, et al. Comparison or regularly scheduled with as-needed use of albuterol in mild asthma. *N Engl J Med* 1996;**335**:841–847.
59. JADAD AR, MOHER M, BROWMAN G, et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ* 2000;**320**: 537–540.
60. OXMAN AD, GUYATT GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;**138**:697–703.
61. SHEFFER A, BARTAL M, BOUSQUET J. Global strategy for asthma management and prevention. *Natl Institutes Health* 1995;**3659**:70–114.