Guides for practitioners: meta-analysis

Clinical research evidence and clinical practice

John A. Collins

Departments of Obstetrics and Gynecology and Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

Address for correspondence: Room 3N52B, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5

An informal survey among colleagues turns up the far-from surprising information that the average patient contact involves at least three or four judgements and decisions: judgements about aetiology and prognosis, decisions about diagnosis and therapy, and sometimes discussions about costs and side-effects. And so it goes: 20 patients a day, 60 decisions; 100 patients a week, 300 decisions. Who makes these decisions, the doctor or the patient? What factors govern the final choice in each case? Evidence-based medicine (EBM) is the factor getting a lot of attention these days, but clinical decisions depend on many different elements. Good doctors have always made use of experience and judgement as well as the best available evidence.

Clinical reasoning

Experience and judgement help clinicians to determine the personal feelings and preferences of each patient and to place the physician’s preferences and motivations in a proper perspective. Thus each decision is the product of complex subjective personal characteristics as well as objective research evidence (Eraker and Politser, 1988). The contribution to the final decision from the research evidence may be small or large: evidence plays a small role when the clinical decision is elective and personally driven, as in the selection of a particular oral contraceptive pill; evidence plays a dominant role when the decision involves life-threatening issues, as in the detection or treatment of cancer. Although the subjective determinants of clinical decisions are in need of more research (Greer, 1988), clinicians are learning to make better use of the objective determinants, that is the results of well-conducted studies.

Clinicians’ knowledge of research evidence is derived from their education, their accumulated experience and their knowledge of current publications. The education component disappears within a few years, because biomedical knowledge is changing and expanding. For example, how much of the information learned by students in 1990 about oral contraceptive use and venous thromboembolism would be useful today without supplemental knowledge about inherited thrombophilia states (Bokarewa et al., 1995; Henkens et al., 1995)? Also, although experience is priceless and some clinical experiences are etched into our memories, experience can be misleading: the most memorable experiences are the atypical events which generate utter terror and sheer pleasure. Thus, clinical research is the most reliable source of evidence on which to base clinical reasoning and clinical judgements.

How do clinicians access research evidence? Discussions with colleagues, hospital grand rounds, clinical meetings and postgraduate sessions usually take precedence over reading the current literature. The rapid expansion of published evidence in recent years means that clinicians are hard-pressed to stay at the forefront of advances in knowledge and technology through reading. Two unfolding developments may make that task easier. Evidence-based medicine (EBM) procedures empower physicians to choose ruthlessly only the readings which will be useful in their practice. Meta-analysis techniques objectively summarize the often-confusing literature on clinical subjects. Both developments have the potential to assist clinical decision-making in reproductive medicine.

Evidence-based medicine

EBM has recently been defined as the judicious and conscientious use of current best evidence from medical care research for making medical decisions (Sackett et al., 1996). That means finding, evaluating and applying the best evidence. EBM procedures involve finding the best available evidence, rating the quality of the published information and applying the results usefully in practice. The idea was not born yesterday; for many years, publications from the group led by Ian Chalmers have made reproductive clinicians aware that for evaluating treatments, evidence from randomized controlled trials is often superior to that from cohort and case-control studies, and certainly superior to case series (Chalmers, 1989).

Treatment decisions, however, comprise only part of the reasoning that is initiated by each patient contact. Diagnostic tests are done; then, having arrived at a diagnosis, the clinician forecasts the likely prognosis; next, one or more treatments may be selected. The final choice depends on judgements about the balance between the benefits and risks of each treatment. While randomized controlled trials provide the best evidence on the benefits (and common risks) of treatment, evidence about diagnosis, prognosis and rare side-effects comes from a variety of study types, each of which is characterized by specific standards of quality. The usefulness of diagnostic tests is best determined by studies based on cohorts of typical patients with the clinical presentation, all of whom undergo both the diagnostic test being evaluated and a reference standard test for the condition. The evidence is superior if the
determination of the reference standard is blinded to the initial test result (Jaeschke et al., 1994).

The prognosis of a given clinical condition is best determined in cohorts of typical patients who are followed long enough to determine the incidence of all common outcomes. The likelihood of a given outcome should be adjusted for factors which affect the outcome, such as age, and the severity of the condition. If the cohort includes treated patients, the effect of treatment also should be taken into account (Laupacis et al., 1994). The frequency of rare side-effects of treatment can be estimated only by means of aetiological studies, which may be large, expensive and time-consuming cohort studies, or, more expeditiously, retrospective case-control studies. In cohort and case-control studies, the quality of the evidence depends on the care which is taken to exclude or account for the biases that must be expected when the groups under comparison are established by a means other than random allocation (Levine et al., 1994).

Thus EBM helps clinicians to read only the research literature that will make clinical practice more effective (Oxman et al., 1993). Sadly, two large barriers hinder the ultimate usefulness of EBM in practice: time pressure and the diversity of the published evidence. In an ideal world, the clinician presented with a novel clinical problem would have time to do the search, evaluate the data in all of the studies and convert the data into a useful form. In the real world, faced with increasingly complex decisions, physicians do not have time to follow-up every uncertainty they encounter in practice. The diversity of the clinical literature is a barrier to the usefulness of EBM because clinicians are left to wonder which of many papers is the correct one. Even randomized clinical trials addressing the same question can lead to a variety of negative and positive results. This is due, of course, to normal variability in trial settings, inclusion criteria and the play of chance, to name just a few of the factors. Thus, clinical readers who have diligently sought the best available evidence often find an apparent lack of agreement in the published evidence. Meta-analysis techniques can address both the time and diversity problems.

Meta-analysis

Meta-analysis aggregates published data in a quantitative manner. Clinicians can save time by searching for the relevant meta-analysis (Oxman et al., 1994). The quantitative techniques also help clinical readers to deal with the diversity of the published evidence. Because meta-analysis combines study data objectively, it is possible to establish guidelines for each set of procedures: search strategies, study selection, validity scoring and analysis. There is open discussion on the principles for dealing with heterogeneity and interpretation. Some critics hold that meta-analysis and statistical methods in general cannot do justice to complex clinical issues. To those critics, one can only reply as follows. Faced with a variety of published answers to a complicated problem, the typical physician must somehow, come hell or high water, sooner or later, intuitively or explicitly, qualitatively or quantitatively, reconcile the conflicting information and arrive at a synthesis which will be applied in clinical practice. Accepting that the clinician inevitably will integrate the contentious data, is it not better to do so objectively rather than subjectively? Summary and synthesis are essential steps in the application of knowledge to the clinical process, and the only question is whether this will be done by intuitive or explicit means. In either case the evidence will be collected and judged, inferences will be made and action will be taken. Meta-analysis techniques allow for that process to be orderly, objective, reportable and accountable. Statistical summaries cannot replace clinical judgement, but they provide access to the research evidence that is needed to implement good judgement. Meta-analysis is not a method to be recommended for browsing an area of clinical concern: the reliance on explicit and detailed procedures implies that a single meta-analysis cannot be expected to cover a wide range of related questions (Table I).

In view of the variety of useful types of clinical research evidence, it is not surprising that different meta-analysis procedures are employed for different types of evidence. The available meta-analysis methods range from simple graphical presentation to sophisticated multiple regression techniques. Table II lists some of the methods and clinical questions to which they may be applied. The table does not include the biases that must be expected when the groups under consideration are established by means other than random allocation (Levine et al., 1994).

<table>
<thead>
<tr>
<th>Study type</th>
<th>Meta-analysis technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prognosis studies</td>
<td>Tabulation, graphics</td>
</tr>
<tr>
<td>Diagnostic studies</td>
<td>Simple aggregation</td>
</tr>
<tr>
<td>RCTs (continuous outcome)</td>
<td>Weighted means</td>
</tr>
<tr>
<td>RCTs (event outcome); diagnostic studies</td>
<td>Weighted odds ratios, logistic regression</td>
</tr>
<tr>
<td>Case-control, cohort studies</td>
<td>Weighted odds ratios, multiple regression</td>
</tr>
</tbody>
</table>

RCTs = randomized controlled trials.

In the case of treatment studies, the relevant meta-analysis techniques generally combine the logarithm of the odds ratio

---

Table I. Comparison of narrative and numerical types of clinical reviews

<table>
<thead>
<tr>
<th>Review type</th>
<th>Major strength</th>
<th>Major drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative</td>
<td>Allows broader interpretation</td>
<td>Potential for bias</td>
</tr>
<tr>
<td>Numerical</td>
<td>Objective procedures</td>
<td>Limited scope</td>
</tr>
</tbody>
</table>

---

Table II. Types of clinical evidence and examples of applicable meta-analysis methods
studies involve ovulation induction co-interventions and differ- and methodology for the meta-analysis of epidemiological
log (typical OR)
5
studies (Collins, 1995). to studies about treatment (Sackett, 1995; Irwig
number of breast cancer cases who were hormone replacement
associated with hormone replacement therapy, arranged by the
Meta-analysis methods for other study types
Figure 1.
Published estimates of the relative risk of breast cancer
associated with hormone replacement therapy, arranged by the
(OR) for each study, weighted in each case by the inverse of
its variance. Thus larger studies which would have more
precise estimates of the treatment effect are given more
weight in the calculation. The equation takes the following
general form:
log (typical OR) = \frac{\text{sum of log (OR) for each study} \times \text{weight}}{\text{sum of weights}}

Associated equations yield the 95% confidence intervals of
the estimate, the associated \( P \) value and an estimate of
heterogeneity (Fleiss, 1981).

For example, published trials on the treatment of male
infertility were collected and analysed by means of similar
methodology in this journal (O’Donovan et al., 1993) and
these were updated recently (ESHRE Capri Workshop, 1996).
For treatment by means of intrauterine insemination (IUI)
compared with intercourse or intracervical insemination, the
typical odds ratio was 1.4 (95% CI 0.8–2.5). The estimate was
based on six studies involving 1751 cycles; the inference is
that IUI treatment for male infertility, while promising, remains
unproven. The most likely effect of IUI (the point estimate of
the typical OR) is a 40% improvement over the baseline
untreated fecundability. To apply this estimate in practice,
consider a couple with untreated fecundability of \(~2\%\) per
cycle. The best available estimate of the treated fecundability
is a rather bleak 2.8%. (2% \times 1.4) After considering the side-
effects and costs, and recognizing that the best available
evidence is unreliable, this simple calculation can help infertile
couples come to their own conclusion about whether the
treatment is worthwhile.

There are more data on IUI treatment but the additional
studies involve ovulation induction co-interventions and differ-
ent diagnostic groups. The methods referred to above do not
allow for an analysis that takes these additional clinical factors
into account. Logistic regression analysis can address this
drawback, because it allows for the estimation of the typical
odds ratio while adjusting for the presence of one or more
indicator conditions. For this purpose, the data must be reported
in separate two-by-two tables within each level of the indicator
variables, which in this case are treatment and diagnosis. In
this symposium, logistic regression has been applied to analyse
an assembly of randomized controlled trials featuring ovulation
stimulation and/or IUI therapy (Hughes, 1997). This analysis
estimated the independent effects of human menopausal
gonadotrophin, clomiphene citrate and IUI while accounting
for diagnostic category. Ideally one would like to make use of
evidence which is based solely on random allocation to uniform
interventions in similar clinical samples. Until such evidence
accumulates, a regression analysis of this kind allows for a
numerical summary of the existing heterogeneous literature,
which can be applied in a clinical setting.

Meta-analysis methods for other study types
Meta-analysis methods for studies about diagnosis and pro-
gnosis have not yet received the attention that has been given
to studies about treatment (Sackett, 1995; Irwig et al., 1995).
With diagnostic studies, the useful clinical information comes
from the predictive values and likelihood ratios, which are not
generated by standard meta-analysis procedures (Jaeschke
et al., 1994). One option is simply to aggregate the data into
a summary two-by-two table from which the desired values
can be calculated. A second option is to compute separately
the weighted averages of the sensitivity and specificity, together
with the associated 95% CI, \( P \) values and heterogeneity
statistics. To apply the data in practice one would substitute
these values into a hypothetical two-by-two table based on the
typical prevalence of the disorder and calculate the predictive
values. With prognostic studies, the use of graphics and simple
aggregation are likely to provide as much information as any
of the more sophisticated approaches.

The meta-analysis of aetiological studies has an especially
important role in reproductive medicine. For example, oral
contraceptives are used by 90% of young women (Chilvers
et al., 1989), and hormone replacement therapy is the most
common prescription purchased by middle-aged women
(Wysowski et al., 1995); both prescriptions are associated
with rare side-effects, including serious cardiovascular and
malignant conditions. Such rare risks (and rare benefits, too)
can be estimated only by means of epidemiological studies.
Case–control studies and cohort studies are more susceptible
to bias and variability than randomized trials. On any particular
question, the reported risk may be increased in some papers
and decreased in others. Nevertheless clinicians need to inter-
pret this literature, in order to respond to patients’ questions
and needs. Is it appropriate to simply say that the studies are
confusing? In the face of such uncertainty the graphical and
quantitative techniques of meta-analysis can be extremely
useful.

Schlesselman (1997) provides a case history of the approach
and methodology for the meta-analysis of epidemiological
studies. Schlesselman’s graphical presentations provide readers
with an instant appreciation of both the scatter and the
trend within the published results. His analysis makes use of
regression procedures to generate summary statistics. These
summaries are then applied to baseline risk data (the age-specific incidence of endometrial cancer) to obtain absolute risks that can be related to the questions of individual patients.

Graphical representation of published data can have special importance in aetiological questions. A key duty for meta-analysts is to attempt to explain the heterogeneity in the published results. Let’s take as an example the disparity of the published odds ratios or relative risks of breast cancer among ever-users of hormone replacement therapy. (For rare conditions, odds ratios and relative risks are approximately equivalent.) With the use of a funnel plot, the published risk estimates are plotted against sample size; in this example the number of exposed cases represents overall sample size. Figure 1 shows that the scatter of negative and positive results is a function of sample size; with larger studies the relative risk estimates lie closer to unity. The figure also indicates that studies with fewer than ~200 exposed cases are liable to yield imprecise results on either side of unity. A prudent clinician would conclude that evidence from a small study is insufficient. It would not be appropriate, however, to rely only on data from large studies, which would introduce selection bias. The best approximation of the true risk should incorporate all of the published data. Schlesselman (1997) discusses the hazards of arbitrary actions to include or exclude studies and his analysis demonstrates techniques for testing the effects of judgements about selecting studies.

Physicians involved in dozens of decisions every day need accessible, reliable, terse, cogent, numerical research evidence. EBM and meta-analysis can fill this need. As we learn about these methods, undoubtedly there will be changes in the way clinicians make use of EBM procedures and meta-analysis techniques. Let us hope that future clinicians will make use of EBM literature. It is of higher quality than the informal survey referred to above. For now, the study of the subjective components of clinical reasoning remains in its infancy, but rapid advances are occurring in the objective process of accessing, evaluating and applying the research evidence. Those who are concerned with advances in EBM and meta-analysis methods will have a large audience of willing clinicians if they can further reduce the burden of clinical reading, while increasing clinicians’ access to useful evidence.

References


Received on October 7, 1996; accepted on May 12, 1997