Levels of evidence and grades of recommendations in general thoracic surgery

Andrew J. Graham, MD, MSc, MHSc; Gary Gelfand, MD, MSc, MHSc; Sean D. McFadden, MD, MSc; Sean C. Grondin, MD, MPH

Objectives: To determine the grades of recommendations and levels of evidence available if the formal practice of evidence-based medicine is applied to general thoracic surgery. Methods: Three general thoracic surgeons, by consensus, developed a sample of 10 clinically important questions. The first 3 steps of evidence-based medicine (creation of answerable clinical questions, search for best external evidence, and critical appraisal of literature) were performed. Abstracts and appropriate articles were identified through Medline from January 1999 through December 2001. A hierarchical series of search strategies was employed to identify the best level of evidence. The best evidence found was categorized according to the Oxford Centre for Evidence-Based Medicine into 4 grades of recommendations (A–D) and 5 levels of evidence (1–5). Results: The best evidence found for the 10 sample questions was categorized as grade A recommendations in 5 and grade B, also in 5 questions. The levels of evidence found were 1a in 3 studies, 1b in 5, and 2b in 2. Conclusions: A formal evidence-based-medicine approach to general thoracic surgery found the grades of recommendation and levels of evidence for a sample of clinically important questions to be high.
Five steps are recognized in the formal practice of evidence-based medicine (formal EBM):

1. Form clinical questions so that they can be answered.
2. Search for the best external evidence.
3. Clinically appraise that evidence for its validity and importance.
4. Apply it to clinical practice.
5. Evaluate your performance as a practitioner of EBM.

The external evidence you find may be of varying quality. To identify the quality and potential limitations of evidence, several systems have been developed to determine levels of evidence and grades of recommendation.5,7

The strength of recommendations available to general thoracic surgeons following a formal EBM approach is unknown. The quality of surgical research has been criticized in the past.8 The quality of randomized clinical trials (RCTs) comparing 2 surgical procedures was found to have a lower standard than RCTs comparing medical therapies.9 A previous study examining the quality of evidence available for general thoracic surgical procedures alone (not the broader practice of general thoracic surgery) reported the quality as low.10 These studies suggest that the methodologic quality of evidence in general thoracic surgery may be poor, and therefore the strength of recommendations may also be poor. We wished to determine the grades of recommendations and levels of evidence available if formal EBM was applied to general thoracic surgery.

Methods

The first step in formal EBM is to form a clinical question. Ideally, its topic is generated by a physician–patient interaction. Within the setting of the study, we elected to create a sample of 10 clinical topics. The sample topics were chosen by the consensus of 3 thoracic surgeons to represent common clinically significant topics that covered the selection of therapy and diagnostic tests.

Formal EBM then requires that answerable questions be created. Answerable questions are constructed by specifying 4 critical elements:

- the patient or the problem
- the intervention or diagnostic test
- the comparison intervention
- the outcome of interest

The creation of answerable questions is done to maximize efficiency of searching and learning.

The next step in EBM practice is to search for the best external evidence. We developed a set of strategies to search the Medline database from January 1966 to December 2001 via PubMed. The set was designed to identify systematic reviews (SRs) first, which are the key component of the highest level of evidence. It is important to know that SRs are not the same as traditional review articles in that they require explicit reporting of methods. The best-known type of SR is the meta-analysis. If no SRs were located, a search strategy designed to identify RCTs was employed. If no such trials were found, studies generated by the main medical subject heading (MeSH) of the clinical question were the next target.

A search strategy designed to locate systematic reviews by Hunt and McKibbin11 was used. The Boolean command OR was used to combine searches of the complete Medline database for publication type—meta-analysis, publication type—review, textword—meta-anal* (truncation, indicated with an asterisk, finds all terms that begin with a given text string) and textword—Medline. The identified systematic reviews were combined with the major MeSH heading (e.g., esophageal neoplasm) using the Boolean command AND to produce the final result of the search.

Randomized clinical trials were identified by employing the Boolean command OR to combine searches of the Medline database for publication type—randomized controlled trial, and textword—random* OR randomized controlled trial, and textword—random*. The RCTs identified were then combined with the major MeSH heading with the Boolean AND to produce the final search result. To confirm location of all relevant RCTs, the CancerLit database was searched by the major MeSH heading with limits to randomized controlled trial. If no studies were located with these strategies, a search generated by the main MeSH heading limited to English abstracts was performed. (Exact search strategies employed for each question are available upon request.)

The search performed for each clinical question produced a list of references. The abstract for each reference was critically appraised by each surgeon individually, using the criteria of the Oxford Centre for Evidence-Based Medicine (Table 1, Table 2) to determine the appropriate level of evidence and grade of recommendation category. If abstracts were considered ambiguous about the level of evidence, the complete study was reviewed.

If 2 or more abstracts were located for a clinical question, that with the highest level of evidence was reported. If 2 studies at the highest level of evidence were found, both were reported. Discrepancies in categorization were resolved by consensus.

Note that we were not attempting to provide definitive clinical answers to the questions, but rather to establish the level of evidence available to general thoracic surgeons upon which to base their recommendations.
Results

Table 3 lists the answerable questions created and shows the grades of recommendation and level of evidence found for each, along with the pertinent reference. The best evidence found for the 10 sample questions was categorized as grade A recommendations in 5 and grade B recommendations in 5. The levels of evidence found were 1a in 3 studies, 1b in 5 studies and 2b in 2 studies.

Conclusions

The practice of EBM has been suggested to improve patient care and to be a very effective form of CME. Previous studies have suggested that the quality of evidence available for general thoracic surgery procedures may be low. In this study, we adapted a broader perspective than surgical procedures alone and explicitly followed the first 3 steps of formal EBM to determine the strength of recommendations available to general thoracic surgeons. We selected topics from a general thoracic surgery practice that were considered important and common clinical problems, included questions on diagnostic tests and selection of surgical procedures. We found the strength of recommendations to be high, with 5 of 10 recommendations grade A and 5 of 10 grade B.

The evidence found was generally of higher quality than previously reported. This is believed to reflect the perspective of common clinical problems as compared with a log of surgical procedures. By selecting questions based on clinical problems that are both common and important, the quality of available evidence that we found was likely to be superior, as some surgical procedures are unsuitable for assessment by RCTs or SRs. The higher quality of evidence available suggests that following an EBM approach may be more rewarding in general thoracic surgery than previously believed.

The limitations of the study included exclusive use of the Medline and CancerLit databases, without hand searches of articles and textbooks. Solomon and colleagues have shown that 46% of surgical RCTs may be missed with computer searches. This suggests that some studies may have been missed, resulting in an incorrect grade of recommendation being assigned. We believe, however, that the results reflect what is available to a practising thoracic surgeon using the readily available Medline rather than libraries of paper journals and textbooks that must be hand-searched.

A variety of grading systems for levels of evidence have been reported. We selected the Oxford Centre for Evidence-Based Medicine grading system, as it was developed from the most established grading system. The important advantage of this grading system is that it includes grades of recommendations for diagnostic tests as well as therapeutic interventions. Use of a different grading system may have produced alternate results. In general, gradings of levels of evidence would be similar in systems that use a similar hierarchy of studies. Not all grading systems use the same hierarchy of studies; for example, the National Cancer Institute in the USA does not grade SRs higher than RCTs. In these grading systems, the results may be significantly altered.

Table 1

Levels of evidence, Oxford Centre for Evidenced-Based Medicine

<table>
<thead>
<tr>
<th>Level</th>
<th>Therapy</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review with homogeneity* (SRwH) of RCTs</td>
<td>SRwH of level 1 diagnostic studies; CDR with 1b studies from different clinical centres</td>
</tr>
<tr>
<td>1b</td>
<td>Individual RCT with a narrow confidence interval</td>
<td>Validating cohort study with good reference standards, or CDR tested within 1 clinical centre</td>
</tr>
<tr>
<td>1c</td>
<td>All or none†</td>
<td>Absolute SpPins and SnNouts</td>
</tr>
<tr>
<td>2a</td>
<td>SRwH of cohort studies</td>
<td>SRwH of level 2a–c diagnostic studies</td>
</tr>
<tr>
<td>2b</td>
<td>Individual cohort study (including low-quality RCT; e.g., follow-up of &lt; 80% of patients)</td>
<td>Exploratory cohort study with good reference standards; CDR after derivation, or validated only on split sample or databases</td>
</tr>
<tr>
<td>2c</td>
<td>&quot;Outcomes&quot; research; ecological studies</td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>SRwH of case–control studies</td>
<td>SRwH of 3b-and-better studies</td>
</tr>
<tr>
<td>3b</td>
<td>Individual case–control study</td>
<td>Non-consecutive study, or one without consistently applied reference standards</td>
</tr>
<tr>
<td>4</td>
<td>Case series (and poor-quality cohort and case–control studies)</td>
<td>Case–control study, poor or non-independent reference standard</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td></td>
</tr>
</tbody>
</table>

CDR = clinical decision rule (an algorithm or scoring system that leads to a prognostic estimation or a diagnostic category); RCT = randomized controlled trial; SpPin = a diagnostic finding with a specificity so high that a negative result rules out the diagnosis; SnNout = a diagnostic finding with a specificity so high that a positive result rules in the diagnosis; SRwH = systematic review with homogeneity.

* A systematic review that is free of troubling variations in the directions and degrees of results between individual studies
† Reference standards that are independent of the test, and applied blindly objectively to all patients
§ Met when all patients died before the treatment/prescription became available, but some now survive on it; or when some patients died before the treatment became available, but none now die on it.
‡ Reference standards that are applied haphazardly, but are still independent of the test

Table adapted from: http://www.cebm.net/levels_of_evidence.asp

Table 2

Grades of recommendations, Oxford Centre for Evidence-Based Medicine

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Consistent level 1 studies</td>
</tr>
<tr>
<td>B</td>
<td>Consistent level 2 or 3 studies or extrapolations from level 1 studies</td>
</tr>
<tr>
<td>C</td>
<td>Level 4 studies or extrapolations from level 2 or 3 studies</td>
</tr>
<tr>
<td>D</td>
<td>Level 5 evidence or troublingly inconsistent or inconclusive studies of any level</td>
</tr>
</tbody>
</table>
In conclusion, we carried out a study to determine the strength of recommendations for common clinical problems. This study was also limited by the small number of questions that we examined, focused on common and important clinical problems. This may have biased the study toward finding higher levels of evidence, and precluded meaningful statistical analysis. The practice of general thoracic surgery obviously has a much broader scope than contained by these questions. Other questions in general thoracic surgery will likely have poorer quality evidence. It is, however, encouraging that clinically important questions did have high quality levels of evidence available.

In conclusion, we carried out a study to determine the strength of recommendations available to general thoracic surgeons by following a formal EBM approach associated with fewer postoperative complications than contained by these recommendations. In: *EBM* (Web site of the Oxford Centre for Evidence-Based Medicine). Available at: www.cebm.net/levels_of_evidence.asp (accessed 2004 Oct 6).


This study was also limited by the small number of questions that we examined, focused on common and important clinical problems. This may have biased the study toward finding higher levels of evidence, and precluded meaningful statistical analysis. The practice of general thoracic surgery obviously has a much broader scope than contained by these questions. Other questions in general thoracic surgery will likely have poorer quality evidence. It is, however, encouraging that clinically important questions did have high quality levels of evidence available.

In conclusion, we carried out a study to determine the strength of recommendations available to general thoracic surgeons by following a formal EBM approach associated with fewer postoperative complications than contained by these recommendations. In: *EBM* (Web site of the Oxford Centre for Evidence-Based Medicine). Available at: www.cebm.net/levels_of_evidence.asp (accessed 2004 Oct 6).

Competing interests: None declared.

References


