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.

Objectifs : Déterminer la qualité des recommandations et les niveaux de preuves disponibles si l'on applique à la chirurgie thoracique générale la pratique officielle de la médecine factuelle. **Méthodes :** Trois chirurgiens thoraciques généraux, après consensus, ont conçu un échantillon de dix questions cliniquement importantes. On a exécuté les trois premières étapes d'une médecine factuelle (création de questions cliniques avec réponse possible, recherche des meilleures preuves externes et analyse critique de la documentation). On a extrait des résumés et des articles appropriés de Medline entre janvier 1999 et décembre 2001. On a utilisé une série hiérarchique de stratégies de recherche afin de trouver le meilleur niveau de preuves. On a ensuite eu recours au système du Oxford Centre for Evidence-Based Medicine pour classer les meilleures preuves en quatre niveaux de recommandations (A–D) et cinq niveaux de preuves (1–5). **Résultats :** On a classé les meilleures preuves trouvées pour les dix questions de l'échantillon selon des recommandations de niveau A pour cinq questions et des recommandations de niveau B, aussi pour cinq questions. Les niveaux des preuves trouvées s'établissaient ainsi : 1a dans trois études, 1b dans cinq et 2b dans deux. **Conclusions :** Une approche structurée de médecine factuelle appliquée à la chirurgie thoracique générale a permis d'obtenir des niveaux élevés de recommandations et de preuves pour un échantillon de questions cliniquement importantes.

The practice of evidence-based medicine (EBM) has been defined as integrating individual clinical expertise with the best available external clinical evidence from systematic research.[1] The phrase has often been used loosely to describe a variety of approaches to obtaining information about patient care decisions. More accurately, EBM is a process of self-directed learning that follows 5 steps.[1] The advantages of the formal EBM approach are thought to include acquiring current information, direct review of the evidence and an interactive form of continuing medical education (CME). Traditional methods of acquiring new informa-

From the Division of Thoracic Surgery, University of Calgary, Calgary, Alta.

Presented at the Canadian Surgery Forum, Vancouver, BC, Sept. 18–21, 2003

Accepted for publication Mar. 22, 2004

Correspondence to: *Dr. Andrew J. Graham, Department of Surgery, Foothills Medical Centre, 1403 29th St., Calgary AB T2N 2T9; fax 403 270-8431; andrew.graham@calgaryhealthregion.ca*

tion have been review of textbooks and journals. Textbooks have been shown to be rapidly out-of-date, and ongoing review of medical journals can be a difficult task.[2,3] Direct review of evidence has been suggested to be superior to traditional review articles by experts, which have been revealed to be of low scientific quality.[4] Finally, didactic styles of CME have been found to be inferior to interactive forms of learning at changing physician performance.[5]

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.[6,7]

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[1]

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 McKibbon[11] 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 OR to combine searches of the Medline database for publication type—randomized controlled trial, and text word—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.[12–25] 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.[1] Previous studies have suggested that the quality of evidence available for general thoracic surgery procedures may be low.[10] 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.[10] 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[9] 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.[7,26] 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.[26] In these grading systems, the results may be significantly altered.

Table 1

Levels of evidence, Oxford Centre for Evidenced-Based Medicine

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

CDR = clinical decision rule (an algorithm or scoring system that leads to a prognostic estimation or a diagnostic category); RCT = randomized controlled trial; SnNout = a diagnostic finding with a *sen*sitivity so high that a *n*egative result rules *out* the diagnosis; SpPin = a diagnostic finding with a *sp*ecificity so high that a *p*ositive 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

Grade	
A	Consistent level 1 studies
B	Consistent level 2 or 3 studies or extrapolations from level 1 studies
C	Level 4 studies or extrapolations from level 2 or 3 studies
D	Level 5 evidence or troublingly inconsistent or inconclusive studies of any level

Table 3

Results

	Clinical question	Grade of recommendation	Level of evidence	Study
1.	In patients with suspected bronchogenic carcinoma, does positron-emission tomography accurately diagnose metastases to the mediastinum?	B	1b	Pieterman et al.[12] *N Engl J Med* 2000;343(4):254-61.
2.	In patients undergoing anti-reflux surgery, is a 360° fundoplication superior to a partial fundoplication in achieving reflux control?	A	1b	Watson et al.[13] *Br J Surg* 1999;86:123-30. Walker et al.[14] *Br J Surg* 1992;79(5):410-4.
3.	Do patients with locally advanced esophageal cancer that is resectable and who receive neoadjuvant treatment have a survival advantage?	A	1a	Arnott et al.[15] *Int J Radiat Oncol Biol Phys* 1998;41(3):579-83.
4.	In patients with resectable mid–distal-third esophageal cancers, is a trans-hiatal esophagectomy approach associated with fewer postoperative complications than a right thoracotomy laparotomy approach?	B	2b	Chu et al.[16] *Am J Surg* 1997;174(3):320-4. Goldminc et al.[17] *Br J Surg* 1993;80(3):367-70.
5.	Of patients with resected esophageal cancer, do those who receive adjuvant treatment survive longer than those who do not?	A	1b	Ando et al.[18] *J Thorac Cardiovasc Surg* 1997;114(2):205-9.
6.	In patients with resectable esophageal cancer, does an en-bloc resection improve survival over that of a routine resection?	B	2b	Chu et al.[16] *Am J Surg* 1997;174(3):320-4. Goldminc et al.[17] *Br J Surg* 1993;80(3):367-70.
7.	In patients with resectable non–small cell lung cancer, does a mediastinal lymphadenectomy improve survival compared with lymph-node sampling?	B	1b	Sugi et al.[19] *World J Surg* 1998;22(3):290-4, discussion 294-5. Izbicki et al.[20] *Ann Surg* 1998;227(1):138-44.
8.	In patients with undiagnosed pulmonary nodules, is positron-emission tomography an accurate diagnostic test?	A	1a	Gould et al.[21] *JAMA* 2001;285(7):914-24.
9.	In patients with stage IIIa non–small cell lung cancer, does neoadjuvant treatment preceding surgery improve survival?	B	1b	Rosell et al.[22] *Lung Cancer* 1999;26(1):7-14. Roth et al.[23] *Lung Cancer* 1998;21(1):1-6.
10.	In patients with resected non–small cell lung cancer, does adjuvant therapy after surgery improve survival?	A	1a	PORT (postoperative radiotherapy) Meta-analysis Trialists Group.[24] *Lancet* 1998;352(9124):257-63. Non–small Cell Lung Cancer Collaborative Group.[25] *BMJ* 1995;311(7010):899-909.

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. The strength of the recommendations was found to be high. We would recommend a formal EBM approach in general thoracic surgery, as it may improve patient care, provide good-quality CME and provide high grades of recommendations for common clinical problems.

Competing interests: None declared.

References

1. Sackett D, Richardson S, Rosenberg W, Haynes R. *Evidence-based medicine: how to practice and teach EBM*. 1st ed. Edinburgh: Churchill Livingston; 1998.
2. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA* 1992;268(2):240-8.
3. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. *BMJ* 1995;310(6987):1085-6.
4. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci* 1993;703:125-33; discussion 133-4.
5. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA* 1999;282(9):867-74.
6. Ball C, Sackett D, Phillips B, Haynes B, Straus S. Levels of evidence and grades of 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).
7. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323(7308):334-6.
8. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996;347(9007):984-5.
9. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery* 1994;115(6):707-12.

10. Lee JS, Urschel DM, Urschel JD. Is general thoracic surgical practice evidence based? *Ann Thorac Surg* 2000;70(2):429-31.

11. Hunt DL, McKibbon KA. Locating and appraising systematic reviews. In: Mulrow C, Cook D, editors. *Systematic reviews.* 1st ed. Philadelphia: American College of Physicians; 1998. p. 13-22.

12. Pieterman RM, van Putten JW, Meuzelaar JJ, Mooyaart EL, Vaalburg W, Koeter GH, et al. Preoperative staging of non–small-cell lung cancer with positron-emission tomography. *N Engl J Med* 2000;343(4):254-61.

13. Watson DI, Jamieson GG, Pike GK, Davies N, Richardson M, Devitt PG. Prospective randomized double-blind trial between laparoscopic Nissen fundoplication and anterior partial fundoplication. *Br J Surg* 1999;86:123-30.

14. Walker SJ, Holt S, Sanderson CJ, Stoddard CJ. Comparison of Nissen total and Lind partial transabdominal fundoplication in the treatment of gastro-œsophageal reflux. *Br J Surg* 1992;79(5):410-4.

15. Arnott SJ, Duncan W, Gignoux M, Girling DJ, Hansen HS, Launois B, et al. Preoperative radiotherapy in esophageal carcinoma: a meta-analysis using individual patient data (Oesophageal Cancer Collaborative Group). *Int J Radiat Oncol Biol Phys* 1998;41(3):579-83.

16. Chu KM, Law SY, Fok M, Wong J. A prospective randomized comparison of transhiatal and transthoracic resection for lower-third esophageal carcinoma. *Am J Surg* 1997;174(3):320-4.

17. Goldminc M, Maddern G, Le Prise E, Meunier B, Campion JP, Launois B. Œsophagectomy by a transhiatal approach or thoracotomy: a prospective randomized trial. *Br J Surg* 1993;80(3):367-70.

18. Ando N, Iizuka T, Kakegawa T, Isono K, Watanabe H, Ide H, et al. A randomized trial of surgery with and without chemotherapy for localized squamous carcinoma of the thoracic esophagus: the Japan Clinical Oncology Group Study. *J Thorac Cardiovasc Surg* 1997;114(2):205-9.

19. Sugi K, Nawata K, Fujita N, Ueda K, Tanaka T, Matsuoka T, et al. Systematic lymph node dissection for clinically diagnosed peripheral non–small-cell lung cancer less than 2 cm in diameter. *World J Surg* 1998;22(3):290-4; discussion 294-5.

20. Izbicki JR, Passlick B, Pantel K, Pichlmeier U, Hosch SB, Karg O, et al. Effectiveness of radical systematic mediastinal lymphadenectomy in patients with resectable non–small cell lung cancer: results of a prospective randomized trial. *Ann Surg* 1998;227(1):138-44.

21. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001;285(7):914-24.

22. Rosell R, Gomez-Codina J, Camps C, Javier Sanchez J, Maestre J, Padilla J, et al. Preresectional chemotherapy in stage IIIA non–small-cell lung cancer: a 7-year assessment of a randomized controlled trial. *Lung Cancer* 1999;26(1):7-14.

23. Roth JA, Atkinson EN, Fossella F, Komaki R, Bernadette Ryan M, Putnam JB Jr, et al. Long-term follow-up of patients enrolled in a randomized trial comparing perioperative chemotherapy and surgery with surgery alone in resectable stage IIIA non–small-cell lung cancer. *Lung Cancer* 1998;21(1):1-6.

24. Postoperative radiotherapy in non–small-cell lung cancer: systematic review and meta-analysis of individual patient data from nine randomised controlled trials. PORT Meta-analysis Trialists Group. *Lancet* 1998;352(9124):257-63.

25. Chemotherapy in non–small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. Non–small Cell Lung Cancer Collaborative Group. *BMJ* 1995;311(7010):899-909.

26. National Cancer Institute. Levels of evidence for cancer treatment studies: definition and use (PDQ®). Available at: www.cancer.gov/cancer_information/doc.aspx?viewid=2B9AC8C6-7202-4728-9DD0-77CA57170044 (accessed 2004 Oct 6).