

Canadian Association of General Surgeons Evidence Based Reviews in Surgery. 2

Jeffrey S.T. Barkun, MD; Roger G. Keith, MD; for Members of the CAGS Evidence Based Reviews in Surgery Group*

CAGS Evidence Based Reviews in Surgery

In September 2000, the Canadian Association of General Surgeons (CAGS) initiated a program entitled "CAGS Evidence Based Reviews in Surgery" (CAGS-EBRS) to help practising clinicians improve their critical appraisal skills. During the academic year, 8 clinical articles are chosen for review and discussion. Both methodologic and clinical reviews of the article are performed by experts in the relevant areas. The *Canadian Journal of Surgery* will publish 4 of these reviews each year. Each review will consist of an abstract of the selected article and a summary of the methodologic and clinical reviews. We hope that readers will find these useful and learn skills that can be used to evaluate other articles. For more information about the CAGS-EBRS or information about participating in the program, send an email to mmckenzie@mtsinai.on.ca.

Selected article

Cuschieri A, Lezoche E, Morino M, Croce E, Lacy A, Toouli J, et al: E.A.E.S. multicenter prospec-

tive randomized trial comparing two-stage vs single-stage management of patients with gallstone disease and ductal calculi. *Surg Endosc* 1999;13(10):952-7.

Abstract

Question: To compare the effectiveness of endoscopic stone extraction (ESE) followed by laparoscopic cholecystectomy (LC) to laparoscopic cholecystectomy and extraction of common bile-duct (CBD) stones. **Design:** A randomized controlled trial. **Setting:** European Association of Endoscopic Surgery (EAES) at 9 centres. **Patients:** Three hundred fit patients with proven or suspected ductal calculi based on the following criteria: jaundice, a recent episode of acute pancreatitis, elevated liver function test and ultrasonographic results. **Intervention:** Group A patients ($n = 150$) received preoperative endoscopic retrograde cholangiopancreatography (ERCP) with ESE followed by LC during the same admission. Group B patients ($n = 150$) received single-stage laparoscopic management with laparoscopic cholecystectomy, intraoperative cholangiography and extraction of stones through the CBD or cystic duct. **Main outcome measures:** Total hospital stay, efficacy (as mea-

sured by the ability to clear the CBD of stones) and complications. **Results:** See Table 1. **Conclusions:** The results demonstrate equivalent success rates and patient morbidity for the 2 management options but a significantly shorter hospital stay with the single-stage laparoscopic treatment.

Commentary

This report presents the results of the first randomized controlled trial comparing preoperative ERCP with ESE followed by LC to a fully minimal access surgery (MAS) approach for CBD stones. The management of patients in Group A of this trial mimicked the current practice of diagnostic and therapeutic ERCP in Canada. Group B involved cholangiography at the time of LC and either transcystic exploration or laparoscopic CBD exploration. Despite a disclaimer by the authors, international experts at tertiary referral centres conducted this study. In fact the MAS expertise required, unlike ERCP, is currently still not available in most centres in North America, especially Canada.

This is an efficacy trial, and the authors sought to compare the efficacy of one clinical approach with another. Such a goal usually proves that one treatment is "better" than another,

*The CAGS Evidence Based Reviews in Surgery Group comprises Drs. J.S.T. Barkun, G.W.N. Fitzgerald, H.J.A. Henteleff, H.M. MacRae, R.S. McLeod, C.S. Richard, M.C. Taylor, E.M. Webber and M.E. McKenzie.

Correspondence to: Ms. Marg McKenzie, RN, Administrative Assistant, CAGS-EBRS, Mount Sinai Hospital, 1560-600 University Ave., Toronto ON M5G 1X5; fax 416 586-5932

except occasionally when an investigator tries to demonstrate that 2 approaches are similar (“equivalency trial”). With this goal in mind, the trial is said to be “positive” if the much overstated “*p* value is less than 0.05.” Conversely, a trial is “negative” if the *p* value is above this conventional threshold. In fact, well after the results are in and all the *p* values have been counted, there are numerous questions that should be asked. Thus, we will attempt to determine if the methodologic approach and the statistical conclusion make the trial clinically definitive or not.

The primary study outcome was the duration of hospital stay from the start of treatment. The authors do not state whether a formal sample size was calculated before conducting the trial and what the related assumptions were. Secondary endpoints included success rates, conversion rates, morbidity and mortality. It is unlikely that the trial was powered to detect significant differences with respect to these secondary variables.

The authors assigned patients to the 2 groups according to random numbers generated through a centralized office. This limits the possibility of improper randomization at any “overenthusiastic” participating centre and is a definite strength. However, we do not know if each centre randomized a balanced number of patients to each of the 2 treatment arms (“blocked randomization”). This is often done in multicentre trials to offset different “effects” of treatments across centres.

When this is not done, and if a given centre contributes more patients to 1 group than the other, the effect related to centre may only be corrected during analysis through regression techniques. It should be noted that even a properly designed randomization process might not guarantee that patients in both groups are identical for all clinically significant baseline variables. Table 1 thus summarizes the major clinical baseline characteristics across groups, and there are no obvious significant differences in what few baseline variables have been described. One important variable not recorded, but which may be a significant confounder, is the presence of a large CBD stone (> 8 mm, for example). Such cases are more difficult to deal with either by ERCP or laparoscopic approaches. An unbalanced representation of this type of patient in either group could thus influence outcome negatively (i.e., leading to more complications or greater length of stay).

Apart from randomization, another major strategy to bolster the internal validity of the trial is blinding (or “masking”). There are at least 3 types; they involve making the patient, the physician and the outcome measurer unaware of the group to which the patient was randomized. The purpose is to decrease bias across groups. Blinding is often difficult in surgical trials. Here, because the primary outcome is the duration of hospital stay, lack of blinding is a central issue. Indeed, the unblinded surgeons (and patients) essentially deter-

mine the duration of hospital stay. We also do not know who recorded the study endpoints. One solution could have been to have a partner-blinded surgeon determine the eligibility for discharge, as has been done in other trials. Another option would have been to establish “discharge criteria” that patients must fulfil. This solution would have also standardized discharge practices, which are very different across centres in various countries. One could argue that these differences are so significant that pooling length of stay data from all centres may not have been valid (issue of statistical homogeneity).

Patients were entered into the trial when they were suspected of having a CBD stone rather than when the presence of a stone was confirmed (e.g., by cholangiography). This is a complex yet appropriate design, and the entry criteria were surprisingly good as reflected by an 84% prevalence of CBD stones confirmed by cholangiography. Since the natural history of CBD stones is that many pass with increasing time from diagnosis, one could infer that the patients in both groups were treated very early. All patients are very well accounted for, including “misappropriations.” Ten percent of patients underwent major “protocol violations” (i.e., crossover treatment). In a pure efficacy trial, this would be at the upper end of what is acceptable to maintain the internal validity of the trial. However, we can consider this to be a pragmatic trial (approaching normal clinical practice); it would thus be usual to first attempt ERCP and ESE, then, if this failed, proceed to LC and CBDE. Thus, 2 different clinically meaningful “approaches,” rather than procedures, were compared.

Perhaps the main weakness of the paper lies in the authors’ interpretation of the data. They conclude: “equivalent patient morbidity and success rates ... but significantly shorter hospital stay with the single stage approach.” There are at least 3

Table 1

Summary of the Results of Two-Stage (Group A) Versus Single-Stage (Group B) Management of Patients With Gallstones and Ductal Calculi

Outcomes	Group A	Group B	<i>p</i> value
Crossovers, no./total no. (%)	14/150 (9)	17/150 (11)	
Conversions, no./total no. (%)	8/133 (6)	17/136 (13)	0.08
Effectiveness (all pts), no./total no. (%)	129/136 (95)	128/133 (96)	NS
Efficacy (pts. with stones) no. /total no. (%)	82/98 (84)	90/109 (83)	NS
Complications, no./total no. (%)	19/147 (13)	24/149 (16)	NS
Deaths, no. (%)	2 (1.5)	1 (0.75)	NS
Median hospital stay, d	9	6	< 0.05

main outcomes that are measured: total hospital stay, efficacy (as measured by the ability to clear the CBD of stones) and complications. We will only look at the last and first of these in turn. The authors chose to express their results strictly according to *p* values. This is an incomplete way of expressing results whereby a *p* value of less than 0.05 is the clinician's mantra for success. This only gives a yes or no answer; it does not allow interpretation of the results quantitatively, such as with clinical judgement. Rejecting the null hypothesis does not necessarily mean that in fact there is no difference. The use of confidence intervals (CI) (traditionally 95%) gives us more information: a 95% CI indicates that we can be 95% certain that the true difference between the 2 treatments lies in this range. Such a way of reporting results is more informative because one can extrapolate the notion of a clinically significant result. This is a quantitative result that, by individual or common standards, has led surgeons to change their practice. Here, the complication rates in the groups were 17 of 136 (12.5%) for Group A and 21 of 133 (15.8%) for Group B. The difference is 3.3% and the 95% CI is -12.4% to +5.8%. Most surgeons would agree that a 10% difference (for example) in complications is clinically significant. That number is, however, included in the 95% CI: the possibility of having a 10% difference in complications between groups cannot therefore be excluded from the results of the current trial. Contrary to what is claimed, this study does not have the "power" to reject a clinically significant difference, and it is misleading to state with confidence that there is no difference between groups with respect to complications.

Another way of evaluating this point is by use of the conventional

calculation of the power of the study, which is based on rejecting the null hypothesis (this does not consider CIs). Power is usually considered in calculating sample size. The power is the fraction of similar trials that have a *p* values less than 0.05 and thus are deemed "statistically significant." The chance of finding a difference of 10% in complications between the 2 groups can be calculated to be between 60% and 70% (in other words 30%–40% of the time we will not detect a 10% difference when in reality there is one). That last number represents the power of the study, and it is not high (usually one expects a power of at least 80%).

The study conclusions with respect to the difference in median days of hospitalization are also suspect. This was the only significant difference between the 2 groups, favouring the fully laparoscopic group B by 3 days (6 v. 9 d). However, by North American standards, the duration of hospital stay for Group A patients appears long. There is no information relating to factors that led to delays between the times of: (1) diagnosis to ERCP or (2) ERCP/ESE and LC. A breakdown of the duration of hospitalization before and after surgery would have allowed for a better interpretation of results according to cultural patterns but is absent. There is also no information related to management algorithms for patients in either group as would be expected in a methodologically strong multicentre trial. Although the authors are an excellent clinical group, we are thus left wondering whether differences in outcome may be affected by differential management schemes or cultural practices. Indeed, a more "optimal" timing of ERCP/ESE and LC, such as what is the current standard in North America, could reverse the

published results of length of stay in favour of a 2-stage treatment.

There are no long-term data on the presence of residual stones. The ESE group would parallel the established data indicating subsequent passage of residual stones in the majority. CBD exploration with the use of MAS might have a higher (or at least equivalent) residual stone frequency comparable to historical data on open CBD exploration. The MAS CBD group would likely present for late ERCP/ESE. Such data would be of interest — if ever published.

Conclusions

This is an important randomized controlled trial comparing current practice and proposed alternative MAS for CBD stones. It represents a major undertaking, which helps to better situate the relative roles of CBD clearance strategies. It confirms that laparoscopic duct clearance strategies are viable in some surgical hands. Increased efficiency of same admission 2-stage treatment or outpatient ERCP/ESE followed by 24-hour stay for LC would reverse the significance of length of stay data in this article. A change to laparoscopic CBD exploration as a single-stage routine approach to CBD stones therefore cannot be strongly recommended. Moreover, to apply such a change to North American general surgery practice would require either extensive training of residents in this technique, or the referral of all patients with CBD stones to tertiary referral centres — recognizing a 20% failure rate plus the possibility of worrisome bile-duct complications.

Acknowledgements: The Canadian Association of General Surgeons Evidence Based Reviews in Surgery is supported by an unrestricted educational grant from Ethicon Inc., Ethicon Endo-Surgery Inc.