

Users' guide to the surgical literature: how to assess an article on health-related quality of life

Achilleas Thoma, MD; Sylvie D. Cornacchi, MSc; Peter J. Lovrics, MD; Charlie H. Goldsmith, PhD;
for the Evidence-Based Surgery Working Group

In addition to relieving clinical symptoms and prolonging survival, the primary objective of any health care intervention is to enhance quality of life (QOL) and well-being.¹ The broader term "quality of life" can be defined as "the adequacy of people's material circumstances and their feelings about these circumstances."² This encompasses indicators of life satisfaction, personal wealth and possessions, level of safety, level of freedom, spirituality, health perceptions and physical, psychological, social and cognitive well-being.²

Health-related quality of life (HRQL), a subcomponent of QOL, comprises all areas specific to health, that is, physical, emotional, psychological, social, cognitive and role functioning, as well as abilities, relationships, perceptions, life satisfaction and well-being. It refers to patients' appraisals of their current level of functioning and satisfaction with it, compared with what they perceive to be ideal.^{3,4} Impaired HRQL is a major reason why patients seek surgical

care. Over the past 20 years, investigators have developed sophisticated instruments to measure such patient-centred outcomes as HRQL in addition to such clinical outcomes as mortality and morbidity.^{5,6} Surgeons increasingly use HRQL assessments to monitor patients over time, help select patients for surgery, determine the effects of surgical treatments and assist in policy decisions.⁷ HRQL outcomes are particularly relevant to surgical investigations because they can capture the impact of, and recovery from, surgery.

This paper's purpose is to help surgeons appraise articles that appear in the surgical literature and purport to measure HRQL outcomes.

Clinical scenario

A recently retired steel worker, aged 69 years and having a 10-year history of treated hypertension, was found during routine examination by his family doctor to have an asymptomatic abdominal aortic aneurysm

(AAA). On subsequent ultrasonography, the aneurysm was found to be 6 cm in diameter. The family doctor referred the patient to you, a vascular surgeon, for treatment. You recommended repair with the open method. After explaining to the patient what the surgery entails, you sent him home to think about it and let you know of his decision within a week. On the second visit, the patient brings along his daughter, an intensive care unit nurse who has done her "Internet research" and would like to know whether you would perform the endovascular aortic aneurysm repair. She has read that patients who have this newer technique have better QOL, and she brings you a copy of an article by Lottman and colleagues⁸ that deals with this issue.

You inform her that you are uncertain about whether this is true and that you will confirm your preferred procedure on the third visit, after you have reviewed this particular article and performed your own independent literature review on the subject.

From the Surgical Outcomes Research Centre, Department of Surgery, St. Joseph's Healthcare and McMaster University, Hamilton, Ont.

The Evidence-Based Surgery Working Group members include Stuart Archibald, MD;†‡ Frank Baillie, MD;†‡§ Mohit Bhandari, MD;†‡§ Margherita Cadeddu, MD;*†‡ Claudio Cinà, MD;†‡§¶ Sylvie D. Cornacchi, MSc;†‡ Forough Farrokhyar, PhD;†‡§ Anna Garnett, BA;†‡ Charles H. Goldsmith, PhD;‡§ Ted Haines, MD;‡§ Robert Hansebout, MD;*†‡ Carolyn Levis, MD;*†‡ Marko Simunovic, MD, MPH;†‡§¶ Ved R. Tandan, MD;†‡§ Achilleas Thoma, MD.*†‡§*

*From the *Department of Surgery, St. Joseph's Healthcare, †Department of Surgery, ‡Surgical Outcomes Research Centre and §Department of Clinical Epidemiology and Biostatistics, McMaster University, and ¶Hamilton Health Sciences, Hamilton, Ont.*

Accepted for publication May 22, 2007

Correspondence to: Dr. A. Thoma, Division of Plastic Surgery, Department of Surgery, St. Joseph's Healthcare, McMaster University, 101-206 James St. S, Hamilton ON L8P 3A9; fax 905 523-0229; athoma@mcmaster.ca

Literature search

The ideal article addressing this question would be a meta-analysis of randomized controlled trials (RCTs) that compares QOL after open AAA repair with QOL after endovascular AAA repair. If not available, you will look for an RCT comparing the 2 techniques. From a computer in the hospital library, you search the Cochrane Database of Systematic Reviews and find that no meta-analysis has been published on this topic. You then use the Internet PubMed search engine that contains the MEDLINE database from the National Library of Medicine (www.ncbi.nlm.nih.gov/PubMed). Keywords to use in your search are derived from your clinical question (refer to the "Users' guide to the surgical literature: how to perform a literature search"⁹ for detailed information on how to develop a clinical question and conduct a successful literature research). You enter the search terms "endovascular aortic aneurysm repair" AND "quality of life"; this yields 21 articles. Because RCTs are considered the optimal study design for evaluating the effects of different surgical interventions, you limit your search to "randomized controlled trials" and to recent articles published between January 2000 and June 2006. These limits narrow your search to 6 articles. You review the titles and abstracts of these articles, and you indeed find the article by Lottman and colleagues,⁸ which was published in 2004.

Of the remaining 5 articles, 1 discusses the background, design and methods for an RCT comparing endovascular AAA repair and open AAA repair,¹⁰ 1 discusses patients at low surgical risk,¹¹ and 1 discusses patients who are unfit to have open repair.¹² The remaining 2 articles are RCTs that compare HRQL in patients randomized to either endovascular AAA repair or open AAA repair.^{13,14} One, by Prinssen and colleagues,¹³ is an RCT including 153 patients and 1-year follow-up.

The second is based on a large multicentre RCT (34 centres) with a sample size of 1082 and a 4-year follow-up (the EVAR trial^{14,15}). You decide to critically review and compare the results from the Lottman and colleagues study and the study with the larger sample size (the EVAR trial).

Several searching hedges have been developed for various topics such as therapy, diagnosis, etiology, prognosis, clinical practice guidelines, systematic reviews, economics and costs, qualitative studies and others.¹⁶ We are unaware of any specifically developed for QOL. Surgeons interested in locating relevant QOL articles on a particular surgical topic can use several different search terms depending on how broad or specific they would like the search to be. For example, using the terms "abdominal aortic aneurysm" AND "quality of life" yields 89 articles, whereas using the term "health-related quality of life" yields 16 articles and the acronym QOL generates 12 articles. As indicated in our original search above, using a very specific term such as "endovascular AAA repair" rather than "AAA" will also produce a more specific search. Terms such as "cost-effectiveness" and "cost-utility" can also generate QOL-related articles.

As with previous guides to surgical literature articles,¹⁷⁻¹⁹ we will use a similar framework (Box 1) to appraise the validity of the study, interpreting the results and applying the results to our patient. The key characteristics of the Lottman and EVAR studies are outlined in Table 1.

Primary guides

Have the investigators measured aspects of patients' lives that patients consider important?

Every day, surgeons make decisions for their patients by asking themselves whether the surgical intervention will make their patient better. Typically, we answer this question on the basis of clinical observations and changes in physiological variables. In

addition, clinicians and patients also expect that interventions will make patients feel better. Surgeons are limited to determining and ameliorating abnormalities of structure and function and hope that through surgical and medical interventions a patient's symptoms and HRQL will be improved. How can we be sure that the investigators have considered the aspects of health that patients consider valuable?

One way that study investigators can show they obtained appropriate HRQL outcomes is to ask patients directly. Another method is to use HRQL questionnaires. Investigators should provide a detailed description of the questionnaires used to assess HRQL in their studies. This method allows readers to judge whether the investigators have measured the aspects of HRQL important to patients: HRQL issues that surgeons think are intrinsically important to patients may not be so. We prefer that investigators use HRQL tools developed with patient input.

Investigators may choose to design their own questions or to use stan-

Box 1. Guidelines for how to use surgical articles about health-related quality of life
A. Are the results valid?
Primary guides
i) Have the investigators measured aspects of patients' lives that patients consider important?
ii) Have important aspects of HRQL been omitted?
iii) Are the HRQL instruments chosen valid, reliable and responsive?
Secondary guides
i) Were HRQL assessments appropriately timed to evaluate the effects of the surgical intervention?
ii) If there were trade-offs between quantity and quality of life, did the investigators use an economic analysis?
B. What were the results?
i) What was the magnitude of the effect on HRQL?
C. Will the results help me in caring for my patients?
i) Will the information from this study help me inform my patients?
HRQL = health-related quality of life.

standardized questionnaires. Streiner and Norman²⁰ provide an excellent guide on how to design health-measurement scales. Urbach and colleagues²¹ describe the steps they took to design and validate a new instrument to measure HRQL following abdominal surgery. They developed a conceptual framework for the constructs underlying HRQL after abdominal surgery by using semistructured interviews with 21 inpatients who were recovering from abdominal operations. They also employed focus groups involving nurses, surgeons, physiotherapy and occupational therapy specialists and

acute pain clinical nurse specialists. Finally, they reviewed existing health status measures. Using structural equation modelling, they designed a reliable and valid 18-item, 6-subscale instrument for use as an outcome measure in studies comparing laparoscopic and conventional surgery.²¹

Standardized HRQL measures of health status may not always capture individual patient concerns. However, in routine clinical practice, most physicians do not rely on health status questionnaires but ask patients directly whether they are better after treatment.²² Alternative options to

standardized health status measures are to have patients complete instruments that allow clinicians to measure and quantify their individual concerns in evaluating medical or surgical interventions. Such patient-specific outcome measures have been developed, some of which include the Patient-Specific Index,^{22,23} the Patient-Specific Symptom Distress Index,²⁴ the McMaster Toronto Arthritis Patient Preference Disability Questionnaire²⁵ and the Schedule for the Evaluation of Individual Quality of Life.²⁶ Wright²² outlines some aspects of patient-specific instruments: they commonly identify patients' individual concerns, complaints or treatment goals; they rate these concerns for severity; they weigh the concerns for relative severity; and they re-evaluate these concerns after intervention. HRQL questionnaires also provide a forum for patients to discuss their expectations of therapy with their clinicians. Potential disadvantages include increased time and effort needed to implement these measures and the potential to miss the patient's overall state of health by concentrating too much on individual concerns. Patient-specific measures may be useful in conjunction with standardized health status measures, but to date, they have not been widely implemented in clinical research.²²

Turning to our articles,^{8,14} we need to assess whether the investigators have measured aspects of patients' lives that patients consider important. AAA repair is associated with high operative mortality and morbidity. In the open method, mortality has been reported to be between 3% and 8% and morbidity to be between 10% and 23%.^{27,28} Although endovascular AAA repair is a new technique, the operative mortality and morbidity have been shown to be lower in some studies.^{29,30} We will assume that mortality and morbidity are important issues to measure and would be of concern to our patient.

A major disadvantage of endovascular AAA repair is lifelong surveil-

Table 1

Key methodological features of the Lottman and colleagues⁸ and EVAR trials^{14,15}

Feature	Lottman		EVAR	
	EAAAR	OAAAR	EAAAR	OAAAR
Sample size	57	19	543	539
Mean age, y	69	68	74	74
Mean aneurysm size, mm	56	52	65	65
Mean hospital stay, d	5	11	7	12
Outcomes				
EQ-5D Index, 0-3 mo	—	—	0.73†	0.67†
SF-36 physical component score, 0-3 mo	—	—	37.82†	36.14†
SF-36, 1 mo postoperatively				
Physical functioning	61*	44*	—	—
Pain	70*	45*	—	—
Vitality	55*	39*	—	—
Physical limitations	44*	13*	—	—
EQ-5D usual activities	46*	12*	—	—
Follow-up for HRQL				
	Baseline, 1 and 3 mo		Baseline, 1, 3 and 12 mo	
Mortality, no. (and %), 30 d	1 (1)	1 (5)	19 (3.5)†‡	34 (6.3)†‡
Mortality, no. (and %), to 4 y postoperatively	—	—	100 (18)§	109 (20)§
Complications, no. (and %)				
Endoleaks	12 (21)	0	118 (22)	1 (0.2)
Graft rupture	—	—	9 (1.7)	0
Graft infection	—	—	1 (0.2)	2 (0.4)
Graft migration	—	—	12 (2.3)	0
Graft thrombosis	—	—	12 (2.3)	1 (0.2)
Graft stenosis	—	—	2 (0.4)	1 (0.2)
Iliac dilatation	—	—	1 (0.2)	5 (1)
Reinterventions	—	—	81 (15)	36 (7)
Follow-up for complications				
	At 3 mo		To 4 y postoperatively	
Recruiting period	1996-1999		1999-2003	

EAAAR = endovascular abdominal aortic aneurysm repair; EQ-5D = EuroQol; HRQL = health-related quality of life; OAAAR = open abdominal aortic aneurysm repair; SF-36 = 36-Item Short-Form Health Survey; — = data not provided.
 *Significant at $p < 0.05$.
 †Significant at $p < 0.01$.
 ‡Aneurysm-related deaths.
 §Number of deaths from all causes.

lance and a high rate of complications and secondary interventions.²⁹⁻³¹ Lottman and colleagues⁸ set the “time horizon” at 3 months after surgery, reporting endoleaks in the endovascular AAA repair group (21%). They provide no information concerning reinterventions or other complications such as graft-related problems. Such a short follow-up period would have missed important secondary events that would have an impact on HRQL. For example, the endovascular AAA repair also requires a computed tomography scan or ultrasound every 6 months to ensure that there has been no progress in the aneurysm, a procedure not required in the standard open technique. The repeat radiologic assessments may be associated with some radiation exposure. In contrast, the EVAR trial measured HRQL up to 12 months postoperatively and provided follow-up information on postoperative complications and reinterventions between 1 and 4 years after surgery (complete 4-year follow-up for all patients pending completion of trial) (Table 1).¹⁴

In terms of HRQL assessment, both Lottman and colleagues⁸ and the EVAR study¹⁴ administered the 36-Item Short-Form Health Survey (SF-36) and the EuroQol (EQ-5D) questionnaires before and after surgery.

Different types of standardized HRQL instruments have been developed, including generic measures (which have both health status and utility instruments) and disease-specific measures. Box 2 outlines some of the generic health status and utility instruments most commonly used in surgical trials. Generic HRQL instruments provide an overall assessment of HRQL, with questions covering many health-related domains such as physical, social, emotional and cognitive functioning, as well as mental health, pain and general health.^{6,7,32} These instruments include both descriptive health status questionnaires (e.g., the SF-36) and health utility measures (e.g., the

EQ-5D).³⁷ Utility measures provide preference-weighted outcome measures that represent patients’ preferences for a given health state relative to death (represented by 0) or perfect health (represented by 1). There are various methods of measuring utilities, including the visual analogue scale, the standard gamble, the time trade-off and standardized questionnaires such as the EQ-5D and the Health Utilities Index.^{36,37}

Generic HRQL instruments allow comparison among patients with different types of diseases but may not be sensitive enough to detect small differences in patient groups with specific disabilities.³⁸ By comparison, disease-specific (condition-specific) HRQL measures focus on specific symptoms and impairments relevant to a particular surgical intervention or disease state. Examples include the European Organization for Research into Treatment of Cancer Quality of Life Questionnaires, which have scales for specific cancer disease sites,³⁸ or the 12-item Hip Score Questionnaire³⁹ for symptoms specific to patients undergoing total hip replacement. Box 3 outlines some commonly used disease-specific instruments used in surgical trials. Disease-specific measures may be more sensitive for detecting small differences in specific symptoms in patients with a particular disease but are less useful for comparing different populations and disease groups.

The particular HRQL domains relevant to a particular study will be determined by the objectives of the

study as well as the disease state, patient group and type of intervention. One should select measures that are sensitive enough to capture both positive and negative effects of surgery and also to detect changes in HRQL in both treatment and control groups.⁵⁰

Have important aspects of HRQL been omitted?

In assessing HRQL, investigators must not be biased, examining some issues while ignoring others. Assessments need to be comprehensive. Depending on the type of disease you are dealing with, certain aspects of HRQL may be more relevant than others; similarly, disease-specific instruments may be more relevant than generic ones and vice versa. For example, an orthopedic surgeon performing total knee replacements for patients with osteoarthritis may focus

Box 3. Disease-specific scales commonly used in surgical studies

- The European Organization for Research in Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), with disease-specific modules for breast, head and neck, esophageal, gastric, ovarian, cervical, prostate, pancreatic and others³⁸
- The Functional Assessment of Cancer Therapy (FACT), with disease-specific modules for breast, head and neck, colon, cervical, brain, bladder and others⁴⁰
- Auckland QOL Questionnaire/Quality of Life Instrument for Head and Neck Cancer (QL-H&N)⁴¹
- Head and Neck Quality of Life Instrument/University of Michigan Head and Neck Quality of Life (HNQOL)⁴¹
- Gastroesophageal Reflux Disease Health-Related QOL Symptoms (GERD-HRQL)⁴²
- Gastrointestinal Quality of Life Index (GIQLI)⁴³
- Abdominal Surgery Impact Scale²¹
- Epilepsy Surgery Inventory⁴⁴
- Western Ontario and McMaster University Osteoarthritis Index (WOMAC)⁴⁵
- Arthritis Impact Measurement Scale⁴⁶
- 12-item Hip Score Questionnaire and 12-item Knee Score Questionnaire^{39,47}
- Vascular Disease Quality of Life Questionnaire (VascuQoL)⁴⁸
- Claudication Scale (CLAU-S)⁴⁹

Box 2. Examples of generic health-related quality of life scales used in surgery

Health status instruments

- Short Form-36 (SF-36)³²
- Nottingham Health Profile³³
- McMaster Health Index Questionnaire³⁴
- Sickness Impact Profile³⁵

Health-utility instruments

- Quality of Well-Being Scale²
- Health Utilities Index, Mark 2 and 3³⁶
- EuroQol (EQ-5D)³⁷

on pain and physical functioning by using the Western Ontario and McMaster University Osteoarthritis Index (WOMAC) but may not measure emotional functioning.⁵¹ Surgeons evaluating the literature need to decide what aspects of HRQL are important and whether omissions of certain factors are critical for the management of their patients. Guyatt and colleagues¹⁹ recommend that all studies contain 3 types of HRQL instruments — a generic measure, a disease-specific measure and a utility measure — to capture both overall and specific HRQL concerns.

Lottman and colleagues⁸ and the EVAR trial¹⁴ emphasized morbidity and mortality as outcomes of concern. Pain, problems with physical functioning and feeling unwell are issues that affect surgical patients. The EQ-5D and SF-36, 2 generic scales used in these studies, are appropriate for use in a cohort of patients undergoing AAA repair. If there were particular issues that affected only AAA patients, a disease-specific instrument, if available, would have been appropriate. Since the publication of the above studies, Borchard and colleagues⁵² have developed the Australian Vascular Quality of Life Index, an index specific for patients with vascular disease and appropriate for patients with AAA in the clinical setting. They have conducted a study to show the validity and reliability of their new instrument, although more research is needed to further evaluate it (e.g., additional validity and reliability studies, studies of its responsiveness, studies with a larger patient cohort). Two other disease-specific QOL instruments that have been used with patients suffering from peripheral vascular disease are the Claudication Scale⁴⁹ and the Vascular Disease Quality of Life Questionnaire.⁴⁸

Review articles outlining HRQL assessments and instruments used in various areas of surgery are available as follows: orthopedics,⁵³ surgical oncology,^{38,50} AAA repair⁵⁴ and cosmetic surgery.⁵⁵

Are the HRQL instruments chosen valid, reliable and responsive?

Measuring surgical patients' HRQL can be challenging. The onus is on the investigators to demonstrate that the instruments they chose in a particular study allow strong inferences to be made about the effect of treatment on HRQL.^{1,56}

HRQL instruments need to satisfy 3 important measurement properties: reliability, validity and responsiveness.

Reliability is assessed by tests of repeatability or reproducibility; these are classified as interobserver reliability (degree of agreement between different observers) and intraobserver or test–retest reliability (agreement between observations made by the same observer).^{20,56} To estimate test–retest reliability, the same HRQL instrument is completed by the same patient on 2 different occasions. The assumption is that there is no substantial change in the health status of the patient being measured on the 2 separate occasions. Test–retest reliability is important because we are interested in determining that the pre- and postoperative scores reported by the patient reflect a real change in the patient's health that is due to the surgical intervention.

Validity represents the extent to which an instrument measures what it intends to measure. Validating a scale is the process whereby we determine the degree of confidence we can place on inferences we make about people based on their scores from that scale.²⁰ An instrument has face validity if it appears to be measuring what it is intended to measure. Content validity is concerned with a test's ability to include or represent all the content of a particular construct.²⁰ Construct validity involves comparisons between measures, and the examination of the logical relations that should exist between a measure and characteristics of patients and patient groups.^{19,20} We refer readers to Streiner and Norman's²⁰ in-depth discussion of reliability and validity because a full

discussion is beyond the scope of this paper.

It is also imperative to consider the responsiveness of a HRQL instrument, that is, its ability to detect changes in general, to detect clinically important change and to detect real changes in the concept being measured.⁵⁷ Responsiveness has 2 major aspects: internal responsiveness characterizes the ability of a measure to change over a prespecified timeframe, and external responsiveness reflects the extent to which change in a measure relates to a corresponding change in a reference measure of clinical or health status.⁵⁸ Terwee and colleagues⁵⁷ provide a comprehensive list of measures used to evaluate the responsiveness of HRQL instruments.

When evaluating the properties of a HRQL scale, floor and ceiling effects must also be considered. This occurs when most of the scores are very near the top or bottom of the scale, meaning that it is almost impossible to detect any improvement or decline.²⁰ For example, Baron and colleagues⁵⁹ evaluated the psychometric properties of the SF-36 in a sample of patients undergoing spine surgery. They found that, in this specific patient group, 3 of the concepts measured (physical role limitations, emotional role limitations and physical functioning) all showed significant floor and ceiling effects, representing cohorts of patients whose individual scores might not be accurate measurements of their true level of functioning, possibly leading to an underestimation of treatment effectiveness.⁵⁹ The alternative explanation is that the floor and ceiling effects they found represent a cohort of patients whose functioning could not be improved. The authors suggest that their findings do not preclude the use of the SF-36 in cervical spine surgery, but they emphasize the importance of using a scale that matches the spectrum of health covered by the study sample.⁵⁹

Both Lottman and colleagues and the EVAR study chose to use the

SF-36 and EQ-5D. These are commonly used generic instruments, and their validity and reliability have been widely demonstrated in many clinical and normative populations.^{32,37} The responsiveness of the SF-36 has been demonstrated in a cohort of patients with AAA repair,⁶⁰ and it is the most widely used instrument in studies evaluating surgical interventions for AAA.^{8,13,14,54,60,61} For interested readers, we recommend McDowell,² who provides a list (Table 20.1, p 254) of commonly used HRQL instruments (both generic and disease-specific) and ratings of their reliability and validity.

Secondary guides

Were HRQL assessments appropriately timed to evaluate the effects of the surgical intervention?

The timing of an assessment is another critical factor in evaluating the effects of surgery on HRQL. The measures should be taken before and after surgery to assess changes. The acute effects of the postoperative convalescent phase will be predominant at early postoperative evaluations; reductions in overall HRQL and in specific domains such as physical and social functioning, along with increases in pain, should be expected immediately after surgery. Assessments should also be administered at various times after surgery to adequately capture early, intermediate and late results, surgical complications and any long-term benefits or iatrogenic events affecting HRQL.

Lottman and colleagues⁸ measured HRQL before surgery and at 2 times (1 and 3 mo) after surgery. Because complications and reinterventions are important factors in endovascular AAA repair, Lottman and colleagues should have extended their follow-up to capture the effects of these complications on patients' HRQL. The EVAR study,¹⁴ on the other hand, provided both early (1 and 3 mo) and

intermediate (12 mo) outcomes for HRQL measures. Mortality and complications were also tracked for up to 4 years postoperatively for some patients, making this study more clinically relevant, although long-term HRQL follow-up to 4 years postoperatively for all patients would have been ideal.

In long-term studies of HRQL, response shift becomes an important issue. This refers to a patient's ability to adapt to his or her illness and to alter his or her perspective and expectations about health and HRQL as experience with disease accumulates; for example, a terminally ill patient may rate severity of pain differently after experiencing very severe pain or may rate HRQL as similar to that of healthy people as the definition of perfect health alters with chronic illness.^{2,62} Response shift should not be an issue in the Lottman and colleagues and EVAR studies because patients with AAA were randomized to surgical treatments, and therefore, we would expect response shift changes to occur in both groups over time and cancel each other, decreasing the bias associated with this phenomenon.

If there were trade-offs between quantity and quality of life, did the investigators use an economic analysis?

When 2 surgical interventions are compared, it is possible that the surgeon may have the option of prolonging the life of the patient but not necessarily improving HRQL. For example, a patient may live a certain number of years but be in pain during this time. An alternative intervention may improve the HRQL by controlling pain but may not prolong survival. As well, although a particular surgical procedure may provide an excellent QOL outcome, it may be associated with a rare and serious complication that could adversely affect HRQL, whereas an alternative intervention might result in lower QOL

yet carry no risks of serious complications. In an RCT, such events would be captured in the calculation of the means of the competing interventions. In the Lottman and colleagues study⁸ and the EVAR trial,¹⁴ these means were provided for the SF-36 and EQ-5D (Table 1). Surgical patients are usually aware of these issues and may exercise their autonomy when faced with these options.

New surgical interventions may be more effective but also more expensive. Before adopting new and expensive technologies, surgeons should look for articles dealing with cost-effectiveness analyses that compare new and standard technologies.¹⁷

By using the EQ-5D, the Lottman and colleagues study and the EVAR trial calculated utilities. These utilities can then be converted into an outcome metric called Quality Adjusted Life Years (QALYs)^{63,64} by multiplying the utility of a health state by the duration of that health state. A QALY is a measure of health outcome that captures gains from reduced mortality and morbidity and combines them into a single measure. The gain in QALYs between 2 surgical procedures provides us with a more patient-centred measure of the success of the surgery. In addition, QALYs provide a common metric in the comparison of disparate interventions, e.g., carpal tunnel release versus coronary bypass.

Because endovascular AAA repair is a more expensive technology, the authors should have integrated the costs and QALYs in a cost-utility analysis. This would have provided us with some idea as to whether endovascular AAA repair is a cost-effective procedure.¹⁷ The EVAR trial measured costs per patient of the initial surgery and the hospital admission on an intention-to-treat basis, finding that costs were higher in the endovascular AAA repair group than in the open AAA repair group (£13 325 v. £9945). This is only a partial economic analysis because they did not integrate the costs and effectiveness of the competing interventions. Neither Lottman and

colleagues nor the EVAR trial calculated QALYs, nor did they perform a cost-utility analysis, a topic beyond the scope of this article. We refer readers interested in cost-utility and economic analyses to Thoma and colleagues' paper¹⁷ on how to evaluate an article on economic analysis.

What were the results?

What was the magnitude of the effect on HRQL?

When investigators report the results of a trial that compares a novel intervention with some standard surgical intervention, they should report the difference in HRQL they obtain. This difference needs to be interpreted in a meaningful way and applied in the context of clinical practice. Is the effect (change in the score) small, medium or large? Is the change obtained by the trial large enough to persuade us to change our practice and adopt the novel procedure? What does a 6-point difference in the SF-36 mean? What does a 0.2-point difference in the EQ-5D mean?^{65,66}

Lottman and colleagues⁸ found that SF-36 physical limitation scores were lower and pain scores were higher at 1 month after surgery when compared with preoperative scores for both the endovascular and open AAA repair groups (Table 1). This is understandable because the patients are likely still in the recovery period. When the 2 surgical groups were compared, the open AAA repair group showed lower scores on vitality, physical functioning and role limitations due to physical health and higher scores for pain than did the endovascular AAA repair group at 1 month after surgery. At 3 months after surgery, they found no difference between the competing interventions. Let us take the 1-month assessment period and examine it critically. The authors did not provide us with a composite score comparing the 2 interventions (i.e., the physical and mental component summary

scores for the SF-36 or the Index score for the EQ-5D). Why not? Were there no differences?

A statistically significant difference was found when the physical limitations score for open AAA repair (mean 13, standard deviation [SD] 25) was compared with the score for endovascular AAA repair (mean 44, SD 42). However, the investigators did not tell us whether this was a clinically important difference. They could have provided us with some evidence from the literature that this 31-point difference is clinically relevant to our patients. A 10-point change in scores has been suggested as a rule of thumb to apply on 100-point QOL scales such as the SF-36.⁶⁷ They also found a statistically significant difference in the EQ-5D item "usual activities no problem," with the endovascular AAA repair group scoring 46 and the open AAA repair group scoring 12. They concluded that endovascular AAA repair patients achieve better physical functioning and are in less pain 1 month after the operation, which may result in an improvement in their vitality and ability to conduct usual activities.

In contrast, the EVAR study¹⁴ reported summary scores, physical component scores, mental component scores and Index EQ-5D but did not report scores for individual domains such as pain. Consequently, it was difficult to compare the EVAR results directly with the results of the Lottman study (Table 1). The EVAR study grouped HRQL scores on the basis of time from randomization (0–3 months, 3–12 months and 12–24 months), although actual HRQL assessments were taken at 1, 3 and 12 months after surgery. The authors used both sets of groupings to analyze their data and reported that they obtained similar findings for both (although they do not show the data for the time from surgery analysis). They found that the open AAA repair group had lower mean SF-36 physical component scores, compared with the endovascular

AAA repair group (36.14 and 37.82, respectively) and lower mean EQ-5D Index scores (0.67 and 0.73, respectively) at 0–3 months after randomization. No between-group differences were found at later times. Once again, the authors did not discuss whether these differences were clinically important. Drummond⁶⁸ reported that differences of 0.03 or greater in mean utility scores are definitely clinically important. According to this guideline, the EVAR trial's 0.06 difference in mean EQ-5D Index score at 0–3 months after randomization is clinically relevant. A subgroup analysis comparing the HRQL scores for those who had complications or reinterventions with those who did not would have been useful for understanding the impact of these secondary events on HRQL.

Another method of understanding the magnitude of HRQL effects is to relate the HRQL scores to patients' global ratings of the magnitude of change they have experienced or to assess whether patients rate themselves as improved when compared with other patients.^{19,65,66}

Although not the focus of this article, the magnitude of effect for mortality is also an important issue for AAA repair. In terms of mortality, Lottman and colleagues⁸ found that there was a 4% difference in mortality in favour of endovascular AAA repair. They did not specify whether they performed a sample size calculation or post hoc power calculation; therefore, we do not know how to interpret their results. Let us assume for a moment that the methodology of this RCT was valid and that the study had adequate power.⁶⁹ From the absolute risk reduction, we can calculate the number needed to treat, which is 25. This means that, for every 25 AAA repairs, we can prevent 1 death by performing endovascular rather than open AAA repair.

The EVAR trial¹⁴ found that all-cause mortality was similar in the 2 groups over 4 years of follow-up, although there was a persistent

reduction in aneurysm-related deaths in the endovascular AAA repair group compared with the open AAA repair group (3.5% and 6.3%, respectively). Within 4 years of randomization, 41% of interventions in the endovascular AAA repair group had complications, compared with 9% in the open AAA repair group ($p < 0.001$). In addition the percentage of patients with at least 1 reintervention by 4 years was 15% in the endovascular group and 7% in the open group.¹⁴

Will the results help me in caring for my patients?

Will the information from this study help me inform my patients?

Most surgeons know from their clinical experience that patients with the same condition sometimes respond differently to their illness and their subsequent surgery. We must determine whether the cohort of patients in the Lottman and colleagues and EVAR studies is similar to the patient described in our scenario before we can generalize the results to our specific patient.

Our patient is similar in age to the mean age of both study cohorts and has an aortic aneurysm of similar size. He is a chronic smoker, which may affect his postoperative recovery. Information concerning smoking habits was not described in the Lottman and colleagues study, whereas the EVAR study reported that about 20% of both the endovascular and open AAA repair groups were current smokers. Knowing whether the results of the HRQL studies are relevant to our patient requires understanding what is involved in the endovascular AAA repair procedure. It all depends on how our patient feels about the issue of pain and physical functioning at 1 to 3 months after surgery. Also of concern is the need for long-term surveillance, the potential for secondary surgery and the possibility of endovascular AAA repair rupture. Our patient may be

inclined to take his chances with a single definitive operation and choose the open AAA repair. If we believe that our asymptomatic patient is similar to the study's patients then we may be inclined to think that he would experience outcomes and benefits comparable to those of the study patients.

Surgical interventions affect HRQL by alleviating symptoms. They can also introduce new problems. Our patient has been diagnosed with a life-threatening condition, and his HRQL has decreased as a result of this information. Lovrics and colleagues⁷⁰ found that the diagnosis of breast cancer had a negative impact on Health Utilities Index and SF-36 scores, with women having an initial diagnosis showing lower scores in comparison with normative scores for women of the same age without breast cancer. In addition, knowledge of possible future complications and reinterventions associated with the endovascular AAA repair may also affect the patient's HRQL.

Resolution of the scenario

Many factors affect the process of deciding on which surgical intervention to use. These factors include:

- The best available evidence from clinical research.
- Patient preferences and actions (Is the patient averse to hospital and doctor visits?).
- Available health care resources (endovascular AAA repair is only available in tertiary academic medical centres).
- The patient's clinical state (comorbidities), setting (rural v. urban) and circumstances.
- Available surgical expertise.⁷¹

You see the patient and his daughter the following week, and you inform them that you reviewed the Lottman and colleagues and EVAR articles. You describe to them the findings of the 2 studies. Our patient is averse to hospitals and doctors' visits, but at the same time, he

is aware that a "time bomb" may be ticking in his abdomen, and he wants his aortic aneurysm dealt with. When he hears that HRQL is only better for the first few months after endovascular AAA repair, that more frequent follow-up visits are required with endovascular AAA repair and that overall mortality at 4 years is the same for the 2 surgical groups in the EVAR study, together with the evidence that open AAA repair has a lower risk of complications and reinterventions, he decides to have the open AAA repair.

Competing interests: None declared.

References

1. Berzon RA. Understanding and using health-related quality of life instruments within clinical research. In Staquet MJ, Hays RD, Fayers PM, editors. *Quality of life assessment in clinical trials: methods and practice*. Oxford: Oxford University Press; 1998.
2. McDowell I. *Measuring health: a guide to rating scales and questionnaires*. 3rd ed. Oxford: Oxford University Press; 2006.
3. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life: basic sciences review. *Ann Intern Med* 1993; 118:622-9.
4. Cella DF, Tulsky DS. Measuring quality of life today: methodological aspects. *Oncology (Williston Park)* 1990;5:29-38.
5. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27-36.
6. Wood-Dauphinee S. Assessing quality of life in clinical research: From where have we come and where are we going? *J Clin Epidemiol* 1999;52:355-63.
7. Wood-Dauphinee S. Quality of life assessment: Recent trends in surgery. *Can J Surg* 1996;39:368-72.
8. Lottman PEM, Laheij RJF, Cuyper PWM, et al. Health-related quality of life outcomes following elective open or endovascular AAA repair: a randomized controlled trial. *J Endovasc Ther* 2004;11:323-9.
9. Birch DW, Eady A, Robertson D, et al. Users' guide to the surgical literature: how to perform a literature search. *Can J Surg* 2003;46:136-41.
10. Prinssen M, Buskens E, Blankensteijn JD. The Dutch Randomised Endovascular Aneurysm Management (DREAM) trial. Background, design and methods. *J Cardiovasc Surg (Torino)* 2002;43:379-84.
11. Soulez G, Therasse E, Monfared AA, et al.

- Pain and quality of life assessment after endovascular versus open repair of abdominal aortic aneurysms in patients at low risk. *J Vasc Interv Radiol* 2005;16:1093-100.
12. EVAR trial participants. Endovascular aneurysm repair and outcome in patients unfit for open repair of abdominal aortic aneurysm (EVAR trial 2): randomised controlled trial. *Lancet* 2005;365:2187-92.
 13. Prinssen M, Buskens E, Blankensteijn JD. DREAM trial participants. Quality of life endovascular and open AAA repair. Results of a randomized trial. *Eur J Vasc Endovasc Surg* 2004;27:121-7.
 14. EVAR trial participants. Endovascular aneurysm repair versus open repair in patients with abdominal aortic aneurysm (EVAR trial 1): randomised controlled trial. *Lancet* 2005;365:2179-86.
 15. EVAR trial participants. Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR trial 1), 30-day operative mortality results: randomized controlled trial. *Lancet* 2004;364:843-8.
 16. Health Information Research Unit (HIRU) Evidence-based Health Informatics. Search Strategies for MEDLINE in Ovid Syntax and the PubMed translation. Available: <http://hiru.mcmaster.ca/hedges/> (accessed 2007 Apr 16).
 17. Thoma A, Tandan V, Sprague S. Users' guide to the surgical literature: how to use an article on economic analysis. *Can J Surg* 2001;44:347-54.
 18. Thoma A, Farrokhyar F, Bhandari M, et al. Users' guide to the surgical literature: how to assess a randomized controlled trial (RCT) in surgery. *Can J Surg* 2004;47:200-8.
 19. Guyatt GH, Naylor CD, Juniper E, et al. Users' guides to the medical literature XII. How to use articles about health-related quality of life. *JAMA* 1997;277:1232-7.
 20. Streiner DL, Norman RG. *Health measurement scales: a practical guide to their development and use*. 2nd ed. Oxford: Oxford University Press; 1998.
 21. Urbach DR, Harnish JL, McLroy JH, et al. A measure of quality of life after abdominal surgery. *Qual Life Res* 2006;15:1053-61.
 22. Wright JG. Evaluating the outcome of treatment: Shouldn't we be asking patients if they are better? *J Clin Epidemiol* 2000a;53:549-53.
 23. Wright JG, Young NL, Waddell JP. The reliability and validity of the self-reported patient-specific index for total hip arthroplasty. *J Bone Joint Surg Am* 2000;82:829-37.
 24. Testa MA, Dimonson DC. Management of hypertension and nephropathy in diabetes. Measuring quality of life in hypertensive patients with diabetes. *Postgrad Med J* 1988;64:50-8.
 25. Tugwell P, Bombardier C, Buchanan WW, et al. The MACTAR Patient Preference Disability Questionnaire — an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol* 1987;14:446-51.
 26. O'Boyle CA, McGee H, Hickey A, et al. Individual quality of life in patients undergoing hip replacement. *Lancet* 1992;339:1088-91.
 27. Blankensteijn JD. Mortality and morbidity rates after conventional abdominal aortic aneurysm repair. *Semin Interv Cardiol* 2000;5:7-13.
 28. May J, White GH, Waugh R, et al. Improved survival after endoluminal repair with second-generation prostheses compared with open repair in the treatment of abdominal aortic aneurysms: a 5-year concurrent comparison using life table method. *J Vasc Surg* 2001;33:S21-6.
 29. May J, White G, Yu W, et al. Concurrent comparison of endoluminal versus open repair in the treatment of abdominal aortic aneurysms: analysis of 303 patients by life table method. *J Vasc Surg* 1998;27:213-21.
 30. Schurink GWH, Aarts NJM, van Bockel JH. Endoleak after stent-graft treatment of abdominal aortic aneurysm: a meta-analysis of clinical studies. *Br J Surg* 1999;86:581-7.
 31. May J, White GH, Yu W, et al. Endoluminal repair of abdominal aortic aneurysms: strengths and weaknesses of various prostheses observed in a 4.5-year experience. *J Endovasc Surg* 1997;4:147-51.
 32. Ware JE Jr. The SF-36 health survey. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd ed. Philadelphia: Lippincott Raven Publishers; 1996. p. 337-45.
 33. Hunt SM, McKenna SP, McEwen J, et al. The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* 1981;15(3 Pt 1):221-9.
 34. Chambers LW, MacDonald LA, Tugwell P, et al. The McMaster Health Index Questionnaire as a measure of quality of life for patients with rheumatoid disease. *J Rheumatol* 1982;9:780-4.
 35. Bergner M, Bobbitt RA, Pollard WE, et al. The sickness impact profile: validation of a health status measure. *Med Care* 1976;14:57-67.
 36. Horsman J, Furlong B, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties, and applications. *Health Qual Life Outcomes* 2003;1:54.
 37. Brooks R. EuroQol: The current state of play. *Health Policy* 1996;37:53-72.
 38. Viola KV, Ariyan C, Sosa JA. Surgical perspectives in gastrointestinal disease: a study of quality of life outcomes in esophageal, pancreatic, colon, and rectal cancers. *World J Gastroenterol* 2006;12:3213-8.
 39. Dawson J, Fitzpatrick R, Carr A, et al. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78:185-90.
 40. Cella DF, Bonomi AE. The Functional Assessment of Cancer Therapy (FACT) and Functional Assessment of HIV Infection (FAHI) quality of life measurement system. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd ed. Philadelphia: Lippincott Raven Publishers; 1996. p. 203-14.
 41. Ringash J, Bezjak A. A structured review of quality of life instruments for head and neck cancer. *Head Neck* 2001;23:201-13.
 42. Velanovich V, Vallance SR, Gusz JR, et al. Quality of life scale for gastroesophageal reflux disease. *J Am Coll Surg* 1996;183:217-24.
 43. Nieveen van Dijkum EJ, Kuhlmann KF, Terwee CB, et al. Quality of life after curative or palliative surgical treatment of pancreatic and periampullary carcinoma. *Br J Surg* 2005;92:471-7.
 44. Vickrey BG, Hays RD, Graber J, et al. A health-related quality of life instrument for patients evaluated for epilepsy surgery. *Med Care* 1992;30:299-319.
 45. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
 46. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980;23:146-52.
 47. Dawson J, Fitzpatrick R, Murray D, et al. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;80:63-9.
 48. Morgan MBF, Crayford T, Murrin B, et al. Developing the Vascular Disease Quality of Life Questionnaire: a new disease-specific quality of life measure for use in lower limb ischemia. *J Vasc Surg* 2001;33:679-87.
 49. Spengel FA, Brown TM, Dietze S. The Claudication Scale (CLAUS): a new disease-specific quality of life instrument in intermittent claudication. *Dis Manag Health Outcomes* 1997;2:S65-70.
 50. Avery K, Blazeby JM. Quality of life assessment in surgical oncology trials. *World J Surg* 2006;30:1163-72.
 51. Aarons H, Hall G, Hughes S, et al. Short-term recovery from hip and knee arthroplasty. *J Bone Joint Surg Br* 1996;78:555-8.
 52. Borchard KA, Hewitt PM, Wotherspoon S, et al. Australian Vascular Quality of Life Index (AUSVIQOL): a pilot study of a disease-specific quality of life measure. *ANZ J Surg* 2006;76:208-13.
 53. Ethgen O, Bruyere O, Richy F, et al. Health-related quality of life in total hip and knee arthroplasty. A qualitative and

- systematic review of the literature. *J Bone Joint Surg Am* 2004;86-A:963-74.
54. Tambyraja AL, Fraser SCA, Murie JA, et al. Quality of life after repair of ruptured abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg* 2004;28:229-33.
 55. Ching S, Thoma A, McCabe RE, et al. Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. *Plast Reconstr Surg* 2003;111:469-80.
 56. Furlong W, Barr R, Feeny D, et al. Patient-focused measures of functional health status and health-related quality of life in pediatric orthopedics: a case study in measurement selection. *Health Qual Life Outcomes* 2005;3:3.
 57. Terwee CB, Dekker FW, Wiersinga WM, et al. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res* 2003;12:349-62.
 58. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459-68.
 59. Baron R, Elashaal A, Germon T, et al. Measuring outcomes in cervical spine surgery: Think twice before using the SF-36. *Spine* 2006;31:2575-84.
 60. Mangione CM, Goldman L, Orav EJ, et al. Health-related quality of life after elective surgery: measurement of longitudinal changes. *J Gen Intern Med* 1997;12:686-97.
 61. Perkins JMT, Magee TR, Hands LJ, et al. Prospective evaluation of quality of life after conventional abdominal aortic aneurysm surgery. *Eur J Vasc Endovasc Surg* 1998;16:203-7.
 62. Razmjou H, Yee A, Ford M, et al. Response shift in outcome assessment in patients undergoing total knee arthroplasty. *J Bone Joint Surg Am* 2006;88:2590-5.
 63. Torrance GW, Furlong WJ, Feeny DH, et al. Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics* 1995;7:503-20.
 64. Feeny D, Furlong WJ, Boyle MH, et al. Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics* 1995;7:490-502.
 65. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
 66. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;49:1215-9.
 67. King MT. The interpretation of scores from the EORCTC quality of life questionnaire QCQ-C30. *Qual Life Res* 1996; 5:555-67.
 68. Drummond M. Introducing economic and quality of life measurements into clinical studies. *Ann Med* 2001;33:344-9.
 69. Lochner HV, Bhandari M, Tornetta P III. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am* 2001;83:1650-5.
 70. Lovrics PJ, Cornacchi SD, Barnabi F, et al. The feasibility and responsiveness of the health utilities index in patients with early-stage breast cancer: a prospective longitudinal study. *Qual Life Res* 2008;17:333-45.
 71. Haynes B, Sackett DL, Guyatt GH, et al. *Clinical epidemiology: how to do clinical practice research*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2006.

Correspondence Correspondance

The treatment of colorectal cancer in young patients

Regarding the article "Colon cancer presenting as an appendiceal abscess in a young patient"¹ published online in the *Canadian Journal of Surgery*, colorectal cancer in young patients is very aggressive with a poor prognosis. The second surgery in the patient should have been supported with a transoperative histopathologic study to demonstrate its benignancy. Otherwise, such malignancy is treated with a radical surgery: right hemicolectomy with or without an ileotransverse colonic anastomosis, depending on the conditions of the abdominal cavity and the experience of the surgeon. Only the benign pathology is treated with minor surgical procedures. Further, a 6-month follow-up is too short.

Luis Enrique Romero-Morales
Internship Pregraduate Student

HGR 36, Instituto Mexicano del Seguro Social
Puebla, México

Competing interests: None declared.

References

1. Arjona Sánchez A, Tordera Torres EM, Cecilia Martínez D, et al. Colon cancer presenting as an appendiceal abscess in a young patient. *Can J Surg* 2008;51:E15-6.

(Dr. Arjona Sánchez replies)

First, I would like thank you for your interest in our case report published in the *Canadian Journal of Surgery*. I agree with your questions, but the aim of this case report was to demonstrate the rare possibility that we may find a malignant pathology in the context of a young patient with no significant medical history who presents with acute appendicitis. We did not consider a malig-

nant pathology in the urgent treatment because we had made a less aggressive intervention instead of a right hemicolectomy.

I agree that the follow-up was short, but we thought that the interesting aspect of this case was the rare disease with which this patient presented. He received adjuvant chemotherapy for 1 year and then presented with a mass in the abdominal wall. Positron emission tomography confirmed a recurring mass in the right peritoneum. The patient underwent peritonectomy followed by intraperitoneal chemotherapy with mitomicin C. His postoperative course was smooth, and he is currently being followed by our Oncology Unit.

Alvaro Arjona Sánchez, MD
Department of Surgery
H.U. Reina Sofia
Cordoba, Spain

Competing interests: None declared.