Health-technology assessment in surgery

Barnaby Reeves

The term health technology is intended to include "all methods used by health-care professionals to promote health, to prevent and treat disease, and to improve rehabilitation and long-term care. The broad definition of a health technology[1] means that a wide range of health-care measures can be considered to be surgical technologies. Health-technology assessment in surgery therefore includes comparisons of surgery with no surgery or best medical treatment, of alternative surgical procedures, and of alternative non-surgical adjuvant therapies (panel 1).

The need for health-technology assessment is widely acknowledged. Describing the effects of both new and established health-care interventions is important because they have financial costs, and sometimes side-effects or complications, as well as benefits. The balance between benefits and costs influences whether or not an intervention is adopted widely. New techniques are sometimes widely implemented and only subsequently found to have no advantage, or even to be less effective, than those that they were intended to supplant.

A general lack of high-quality evidence about the effectiveness of surgical techniques means that examples of ineffective or harmful ones that were adopted without evaluation are harder to identify than in other areas of health care. One example is the introduction of first-trimester amniocentesis instead of chorionic villus sampling (CVS) for fetal karyotyping. Equivocal evidence suggested that second-trimester amniocentesis resulted in better pregnancy outcomes than did first-trimester CVS, and first-trimester amniocentesis was introduced in the late 1980s. However, in 1994, a randomised controlled trial reported no difference in total deaths (3·6% fewer for CVS, 95% CI −0·8% to 8·0%), but significantly fewer spontaneous deaths with CVS (4·6% fewer for CVS, 95% CI 1·4 to 8·0).[2]

A second example is the use of the gamma nail, which was introduced in the late 1980s for fixation of extracapsular hip fractures. This nail was thought to have theoretical advantages over the established fixation device, the sliding hip screw, but a systematic review of ten randomised controlled trials has shown the nail to be associated with an increased risk of operative fracture of the femur (odds ratio 4·48, 95% CI 2·12 to 9·48), and with later fracture of the femur and re-operation.[3]

Large variations in hospital-admission rates and surgical practice are a general indication of differences in opinion about the effects of an intervention. Although other factors, such as a tendency to practise defensively because of fear of litigation, may contribute to such variation, collective uncertainty about the effectiveness of an intervention is almost certainly a major factor. A commonly cited example of this point is the three-fold difference between and within countries in the frequency of caesarean section.[4] Although collective uncertainty can be taken as an indication of the need for an assessment of the effects of an intervention, surgeons rarely admit to uncertainty individually—a reluctance that can be a major obstacle to persuading surgeons to participate in randomised controlled trials (see below).

Panel 1: Comparisons in surgery

Surgery *vs* best medical treatment or no treatment	Surgical discectomy *vs* chemonucleolysis *vs* placebo for lumbar-disc prolapse Ovarian ablation in early breast cancer *vs* no ablation Surgery *vs* conservative treatment for meniscal injuries of the knee in adults
Comparison of alternative surgical procedures	Partial meniscectomy *vs* total meniscectomy for meniscal injuries of the knee in adults Replacement arthroplasty *vs* internal fixation for extracapsular hip fractures Alternative surgical treatments for cervical intraepithelial neoplasia
Adjuvant surgical technologies	Adjuvant chemotherapy *vs* no chemotherapy for localised resectable soft-tissue sarcoma of adults Interventions for preventing blood loss in the treatment of cervical intraepithelial neoplasia Pre-operative traction for fractures of the proximal femur

Reviews of examples cited are in Cochrane Database of Systematic Reviews[7]

Scope of health-technology assessment

Health-technology assessment includes a wider range of activities than simply primary evaluations of defined techniques. First, it is necessary to prioritise technologies for assessment, since there are insufficient resources for the assessment of all unevaluated and novel technologies. Second, several primary evaluations of a technology may be required to provide a clear and comprehensive picture of its effects.[1] Individual studies commonly include too few patients to produce a definitive answer, and the effects of a technology are often smaller than anticipated, yet clinically important. It is risky to generalise from a single study, particularly in surgery. Individual studies are not always able to assess the full range of clinical and patient-related outcomes, economic outcomes, short-term and long-term effects, and possible harm as well as benefit (panel 2). Weighing up the importance of different outcomes, such as quality and length of life, is a continuing challenge.[5] Assessment of a new health technology also often involves other issues, such as humanity, equity, and ethics, and, in some cases, legal considerations.

Health-technology assessment therefore needs to include systematic review and synthesis of a range of types of evidence of the effects of an intervention.[6] If health-technology assessment is to improve health care for patients, there must also be institutions responsible for disseminating high-quality evidence to relevant target audiences, to promote the uptake of effective measures

Lancet 1999; **353** (suppl I): 3–5

Clinical Effectiveness Unit, Royal College of Surgeons, London WC2A 3PN, UK (B Reeves D Phil)

Panel 2: **Effects of health technology to be assessed**	
Clinical outcomes: benefit	Short and long-term cure, absence/reduction in clinical signs, return of biochemical and physiological measures to normal values
Clinical outcomes: harm	Mortality, complications, and other adverse events attributable to the technology being assessed
Patient outcomes: benefit	Short-term and long-term absence/reduction in symptoms, increased ability to perform activities of daily living, increased quality and length of life
Economic outcomes	Use of hospital, primary care, social services resources and costs, and resource use and costs that fall on patients
Impact on other services	Consequences of implementing new health technologies on other health services

and discontinuation of ineffective or harmful ones. In the UK, the Centre for Reviews and Dissemination at the University of York, funded by the National Health Service Research and Development programme, fulfils this role. A database of abstracts of reviews of effectiveness prepared by the Centre is available in the Cochrane Library,[7] which is available on the Internet and on CD-ROM. Internationally, members of the Cochrane Collaboration are continually updating systematic reviews of the effects of health technologies. Such reviews can contribute to the establishment of clinical guidelines, many of which have been produced by organisations such as surgical specialty associations, which make explicit the quality of the evidence on which specific recommendations are based.[8]

Problems that surgical procedures pose for health-technology assessment

Problems can arise at each stage in the assessment of surgical procedures. At the very start of the process, it is difficult to know when to give a new procedure priority for evaluation. If an assessment is done too early, before surgeons have mastered the technique, there is a risk of rejection of an effective procedure. If too late, the technique may have diffused and become established, by which time surgeons will consider it unethical to withhold the procedure. The uptake of minimally invasive surgical techniques provides an example of some of the problems that can arise. Laparoscopic cholecystectomy was adopted in preference to minicholecystectomy by many surgeons without evidence of its effectiveness and is believed by many to have resulted in a higher rate of bileduct injuries while surgeons were learning the technique. Formal evaluations were hampered by widespread optimism about the effectiveness of the minimally invasive approach, which was subsequently found to be exaggerated.[9]

The preferred study design for primary assessments of the effects of a procedure is the randomised controlled trial;[10] this design is the most likely to result in similarity of the groups being compared and so to minimise confounding by differences in known and unknown prognostic factors. This property is important for two main reasons. First, small effects of similar size to those arising from bias and confounding are often clinically important.[11] Second, quantiyfing with adequate precision the effect of an intervention, rather than merely reporting whether or not it is significantly different from an alternative, is essential for weighing up the relative magnitudes of the benefits and the costs.

Randomised controlled trials are more straightforward to conduct for the assessment of therapies adjuvant to surgery than for the comparison of alternative surgical procedures, for the assessment of complex means of promoting recuperation (such as therapeutic or educational interventions to improve mobilisation or to reduce anxiety), or for the assessment of features of service delivery (such as the extent to which infrequently done or technically advanced surgical procedures should be restricted to specialist centres).[12] The infrastructure required for a randomised controlled trial is generally expensive to set up, and this rigour of design is not always necessary; the effects of some surgical procedures may be large and unlikely to be confused with sources of bias or confounding.

When randomised trials are impracticable, reliance may have to be placed on evidence of effectiveness from non-randomised study designs. Such studies need to be carried out and interpreted with extreme care, since they are highly susceptible to confounding and bias.[13] Nevertheless, there is evidence that non-randomised designs can provide valid estimates of effectiveness if standard epidemiological principles are applied to the design of studies and analyses of data.[14] The discrepancies that have been observed between randomised and non-randomised studies can in many cases be attributed to differences in the study populations and residual confounding,[15] or to the precise nature of the technique being evaluated, rather than to biases.

The possibility that some surgeons may have better outcomes with one procedure and other surgeons with an alternative procedure—ie, there is an interaction between surgeon and technique—creates a particular difficulty. The theoretically ideal solution is to randomise eligible patients of each participating surgeon to one or other procedure. There are practical drawbacks with this approach since surgeons who prefer one procedure may be unwilling to participate. More importantly, if there is an interaction between surgeon and procedure, pooling the results across surgeons will give a misleading answer, and quantifying the interaction requires a very large sample size. Randomising patients to surgeons who use different procedures, or studying the patients of different surgeons observationally, may represent a pragmatic alternative but addresses a different question—namely, what are the effects of the alternative procedures when carried out by surgeons who prefer them?

A lack of randomised controlled trials makes quantitative synthesis of evidence by meta-analysis dangerous, since it is difficult to control for differences in the study populations, the precise nature of the technique evaluated, or the outcomes reported when the results of non-randomised studies are pooled.[16] The paucity of randomised controlled trials of new surgical procedures means that in the latest update of the Cochrane library there are few systematic reviews of the effects of surgical procedures,[7] although there are many more reviews of adjuvant surgical therapies.

There can be difficulties in weighing up the benefits and costs of new surgical procedures. Adopting a new surgical technique may not be straightforward, since it is likely to require a surgeon to acquire new practical skills and to develop competence over a number of cases. The costs of mastering a new procedure are likely to be substantial for patients, surgeons, and health services, since surgeons who are learning a new technique typically take longer to carry

out a procedure and have a higher rate of complications than do experienced ones. Since the gradient of the learning curve may vary considerably between surgeons, the inclusion of general recommendations in high-quality evidence that is being disseminated can be difficult.

Changing of attitudes to health-technology assessment in surgery

The surgical community has been accused of not recognising the need for high-quality evaluation.[17] Reviews of surgical journals in 1990[18] and 1996[17] revealed that randomised controlled trials accounted for only 7% of clinical studies, and that the most commonly reported study designs were uncontrolled case-studies or series (84%[18] and 46%[17]). Historically, surgery has been largely unregulated, and there have been few obstacles, other than the obtaining of consent of the patients for the operation, to prevent surgeons from introducing innovative practices. By contrast, a scientific evaluation almost always requires approval by an ethics committee, which may seek assurances about the inclusion of a control group, adequacy of the proposed sample size, data collection, and monitoring.

The UK Department of Health has recently outlined a process for assessment of new health technologies. The process includes horizon scanning, selection of the most significant technologies for assessment, submission of evidence from sponsoring companies where relevant, and critical review of the evidence by regulatory agencies and the National Institute for Clinical Excellence.[19] However, there is no assurance that this process will solve all of the problems of underevaluation of surgical procedures. What makes a surgical technique new is not always easy to define because surgical procedures generally evolve in small steps, which makes it difficult to decide when a procedure has changed sufficiently to justify formal evaluation.

Even when new technologies are given priority for evaluation, the required studies are often difficult to establish. There are now several methods for treating benign prostate hyperplasia, including transurethral resection, and laser, ultrasonographic, microwave, and pharmacological treatments. However, a randomised controlled trial set up in the UK to compare the effectiveness and cost-effectiveness of some of the alternative techniques was halted because sufficient patients could not be recruited.

The difficulty of assessing a moving target is illustrated by the changing way in which minimally invasive coronary artery surgery is being used. Evaluation of this procedure was given priority by the UK National Health Service in 1997, after early case-series had indicated the success of a minithoracotomy approach for bypass grafting without the need for extracorporeal circulation for patients with single-vessel disease, who are usually treated by angioplasty.[20] The most important question about this generic technology can be said to have now changed, since some surgeons prefer to use a median sternotomy incision, even for patients with single-vessel disease, and are grafting multiple vessels without the use of extracorporeal cardiopulmonary circulation for patients who would otherwise have undergone standard coronary bypass surgery.

The introduction of new prosthetic joints in orthopaedics illustrates another problem—ie, the need to assess long-term outcomes for some procedures. Clinically and economically important differences in the failure rate of alternative prostheses are unlikely to emerge for several years and manufacturers are understandably reluctant to invest in the long-term and expensive evaluations that are needed to show benefit.

Devising ways of encouraging surgeons to recognise uncertainty about the effects of surgical procedures and to be less susceptible to the lure of new and expensive technology that has not been fully evaluated probably represents the greatest challenge to health-technology assessment in surgery. A greater awareness of the need to assess surgical technologies should lead to more and higher-quality evaluations of effectiveness, the opportunity to synthesise evidence from individual studies in systematic reviews, and the incorporation of high-quality evidence into guidelines. There also needs to be wider acknowledgment of the difficulty of carrying out randomised trials in some circumstances and a greater appreciation of the potential value of assessments with non-randomised designs when randomised trials prove to be impracticable.

References

1 NHS Management Executive. Assessing the effects of health technologies. London: Department of Health, 1992.
2 Nicolaides K, de Lourdes Brizot M, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10–13 weeks' gestation. *Lancet* 1994; **344:** 435–39.
3 Parker MJ, Handoll HHG, Robinson CM. Gamma nail versus sliding hip screw for extracapsular hip fractures. Cochrane Library. Oxford: Update Software, Issue 1, 1999.
4 Lomas J, Enkin M. Variations in operative delivery rates. In: Chalmers I, Enkin M, Keirse MJNC, eds. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, 1989: 1182–95.
5 Billingham LJ, Abrams K, Jones DR. Quality of life assessment and survival data. In: Black N, Brazier J, Fitzpatrick R, Reeves BC, eds. Health services research methods: a guide to best practice. London: BMJ Books, 1998: 163–72.
6 Chalmers I, Hetherington J, Elbourne D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJNC,eds. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, 1989: 39–65.
7 The Cochrane Library. Oxford: Update Software, Issue 1, 1999. Website address - http://www.update-software.com/ccweb/cochrane/cdsr.htm
8 Guidelines for the management of colorectal cancer. London: Royal College of Surgeons of England and Association of Coloproctology of Great Britain and Ireland, 1996.
9 Downs SH, Black NA, Devlin HB, Royston CMS, Russell RCG. Systematic review of the effectiveness and safety of laparoscopic cholecystectomy. *Ann Roy Coll Surg Engl* 1996; **78:** 241–323.
10 Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993; **270:** 2598–601.
11 Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *Ann N Y Acad Sci* 1993; **703:** 314–40.
12 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312:** 1215–18.
13 Davey Smith G. Cross design synthesis: a new strategy for studying medical outcomes? *Lancet* 1992; **340:** 944–46.
14 Reeves BC, MacLehose RM, Harvey IM, Sheldon TA, Russell IT, Black AMS. Comparisons of effect sizes derived from randomised and non-randomised studies. In: Black N, Brazier J, Fitzpatrick R, Reeves BCA , eds. Health services research methods: a guide to best practice. London: BMJ Books, 1998: 73–85.
15 Hlatky MA, Facc MD, Califf RM, et al. Comparison of predictions based on observational data with the results of RCT of coronary artery bypass surgery. *J Am Coll Cardiol* 1988; **11:** 237–45.
16 Egger M, Schneider M, Davey Smith G. Meta-analysis: spurious precision? Meta-analysis of observational studies. *Br Med J* 1998; **316:** 140–44.
17 Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996; **347:** 984–85.
18 Solomon MJ, McLeod RS. Clinical studies in surgical journals: have we improved? *Dis Colon Rectum* 1993; **36:** 43–48.
19 A first class service. London: Department of Health, 1998.
20 Calafiore AM, Angelini GD. Left anterior small thoracotomy (LAST) for coronary artery revascularisation. *Surgery* 1997; **15:** 61–63.