Achilles Thoma, Feng Xie, Jenny Santos
and Charles H. Goldsmith

Surgeons of all specialties are constantly introduced to new surgical techniques or approaches to solve surgical problems. These innovations are disseminated via conferences, workshops, or publications in specialty journals. The typical surgeon faces difficulty in deciding whether to adopt a new surgical innovation when conflicting opinions by experts are presented. A surgeon needs to consider the opportunity cost when adopting a novel surgical intervention and abandoning one that they usually use. Opportunity cost is defined as "the

value of the forgone benefits" because the resource is not available for its alternative use [1].

New innovations in surgery are often touted by their proponents as being cost-effective with the recommendation to adopt them in our practice and patients. Unfortunately, the term cost-effective is more often than not misused in the surgical literature. For example, Ziolkowski et al. [2] found that most economic evaluations published in plastic surgery and touted to be cost-effectiveness studies were simply cost comparisons.

A surgical technique or surgical approach to be considered as "cost-effective" must have integrated the costs and effectiveness [3, 4]. Economic analysis is a set of formal, quantitative methods used to compare alternative strategies with respect to their resource use and their expected outcomes [4]. Economic evaluation is a unique study design just as randomized controlled trial and case-control studies are. As most surgeons do not have background training in health economics, one can understand how important terminology such as "cost-effective", is misused. This misuse, however, may have direct consequences if surgeons adopt new techniques or surgical approaches, which are touted "cost-effective", when in truth they are not; inefficient use of scarce healthcare resources is one consequence. OECD data estimate that 20% of health expenditure worldwide is wasted, resulting in minimal-if any-improvement of health outcomes [5]. Although we do not have specific breakdown figures for surgery, we have

A. Thoma (✉) · J. Santos
Department of Surgery, Division of Plastic Surgery,
McMaster University, Hamilton, ON, Canada
e-mail: athoma@mcmaster.ca

J. Santos
e-mail: santosj8@mcmaster.ca

F. Xie
Faculty of Health Sciences, McMaster University,
Hamilton, ON, Canada
e-mail: fengxie@mcmaster.ca

A. Thoma · F. Xie · C. H. Goldsmith
Department of Health Research Methods, Evidence,
and Impact, McMaster University, Hamilton, ON,
Canada

C. H. Goldsmith
Faculty of Health Sciences, Simon Fraser University,
Burnaby, BC, Canada

C. H. Goldsmith
Department of Occupational Science and
Occupational Therapy, Faculty of Medicine,
The University of British Columbia, Vancouver,
BC, Canada

no reason to believe that surgery is immune to wasteful practices.

This chapter has three objectives. First, it will introduce surgeons to the terminology used in the economic evaluations and demystify this important study design. Second and most importantly, it will help the reader appraise and understand an article that purports to be an economic evaluation in surgery. Third, it will hopefully stimulate surgeons to consider "piggy-backing" economic evaluations to their effectiveness studies. We believe that an economic evaluation alongside a robust randomized control trial that compares a novel surgical technique to a standard technique provides the best level of evidence to adopt or reject a novel surgical intervention. We will attempt to keep the mathematics to the bare minimum and make the chapter understandable and hopefully fun to read.

Explanation of the Types of Economic Evaluations and Terminology Used

There are four types of health economic evaluations; Cost analysis (CA): this is cost comparison study and usually not considered a full economic evaluation, it is often called a partial economic evaluation, Cost-Effectiveness analysis (CEA), Cost–Utility Analysis (CUA), and Cost–Benefit Analysis (CBA) [4]. The main difference in these analyses is how the outcome or consequences of the treatments under comparison are measured. The distinguishing features of these four economic evaluations have been summarized in Table 23.1.

There are two main types of methodologies used in economic evaluations. The first type is the model-based and the second is trial-based economic evaluations. In the first, the model-based evaluation (also known as *Deterministic Analysis* or probabilistic analysis), a model is built which in its simpler form is the decision tree which is explained in Chap. 22 of this book. Primary data are usually derived from the literature, for example, by pooling the evidence (preferably through a systematic review of the literature). From these pooled data, one derives probabilities of complications or positive health outcomes labeled as "health states". These health states are then entered into a decision analysis tree, which in its most basic form will look like the one illustrated in Fig. 23.1a.

Clinical investigators then proceed to estimate the expected costs and expected benefits of the "health states" of the interventions under study

Table 23.1 Distinguishing features of the four types of economic evaluations

Type of analysis	Valuation of costs	Identification of outcomes (Consequences)	Metric used in the analysis
Cost analysis (CA)	Monetary units (i.e., $, £, €)	None	None. This analysis is only a comparison of costs
Cost-effectiveness analysis (CEA)	Monetary units (i.e., $, £, €)	Common effect of interest. Common outcomes to the competing surgical interventions but with different probability of success (i.e., lives saved, successful hernia repairs, viable flaps, viable replants, sick days averted, hospital days averted)	$ per natural unit (i.e., $ per successful replant, $ per life saved, $ per hospital day averted)
Cost–utility analysis (CUA)	Monetary units (i.e., $, £, €)	Single or multiple effects that are not necessarily common to both interventions. Outcomes are measured in health utilities that are used to calculate Quality-Adjusted Life Years (QALYs)	$, £, or € spent per QALY
Cost–benefit analysis (CBA)	Monetary units (i.e., $, £, €)	Single or multiple effects not necessarily common to both surgical procedures and are calculated in $, £, or €	Monetary units (i.e., $, £, €)

by multiplying the costs and consequences (outcomes) by their probability of occurrence. The decision analysis tree model becomes more complex as we add more branches to the main pathways. This type of analysis is based on modeling. An example of a more complex model is shown in Fig. 23.1b, representing the cost and utility associated with various complications of a Deep Inferior Epigastric Perforator (DIEP) flap in postmastectomy breast reconstruction [6]. A similar analytic tree was also constructed for the other intervention for breast reconstruction being studied; the free Transverse Rectus Abdominis Myocutaneous (TRAM) flap.

The second type, the trial-based economic evaluation (also known as *Stochastic Analysis*), derives data directly from the patients in the trial. It incorporates sampling uncertainty that is inherent probabilistic by nature. Parallel to a prospective, and preferably, a randomized controlled trial the researchers capture not only clinical outcome data but also direct and indirect costs related to the comparative interventions.

Formulas Used in Economic Evaluations that Inform Adoption of Novel Surgical Interventions

For Cost Analysis

$$\text{Cost Analysis} = \text{Mean Cost}_{\text{novel surgical intervention}} - \text{Mean Cost}_{\text{comparative surgical intervnetion}}$$

(This is a partial economic evaluation. It assumes that the outcomes are the same, which may not be correct unless one measures them and finds them to be so).

If the outcome is similar, we adopt the novel intervention if it is found to be less costly.

For Cost-Effectiveness Analysis

Here we calculate the Incremental Cost-Effectiveness Ratio (ICER)

$$\text{ICER} = \frac{\Delta C}{\Delta E} = \frac{\left(\begin{array}{c}\text{Mean Cost}_{\text{novel surgical intervention}} \\ -\text{Mean Cost}_{\text{comparative surgical intervention}}\end{array}\right)}{\left(\begin{array}{c}\text{Mean Effectiveness}_{\text{novel surgical intervention}} \\ -\text{Mean Effectiveness}_{\text{comparative surgical intervention}}\end{array}\right)}$$

The adoption of a novel surgical intervention will be based on some threshold, which is arbitrarily set by clinical investigators or health economists based on consensus or in the case of CBA on showing a Net Social Benefit [4].

An important outcome that unfortunately is not commonly used in surgical comparative studies is Quality-Adjusted Life Years (QALYs) [7]. This is an outcome that captures both quality of life and quantity of life in a single measure. It can be measured with preference-based instruments, such as the EuroQOL 5-Dimension (EQ-5D) and the Health Utilities Index [7].

(Although a full economic evaluation, because the outcomes are measured differently, it cannot be used across disparate surgical interventions that measure outcomes differently).

The main limitation of this type of analysis is that it is difficult to compare disparate surgical interventions. Imagine the hypothetical scenario where two specialties, plastic surgery and orthopedic surgery, present each to a third-party payer a CEA and ask that their novel surgeries be funded but there is funding for only one. If the ICER submitted by each specialty were as follows:

Fig. 23.1 **a** Decision analytic tree illustrating possible health state (pathway) probabilities for free TRAM versus DIEP. **b**. DIEP flap in postmastectomy breast reconstruction. *P* probability of each pathway; figure from Thoma et al. [6]
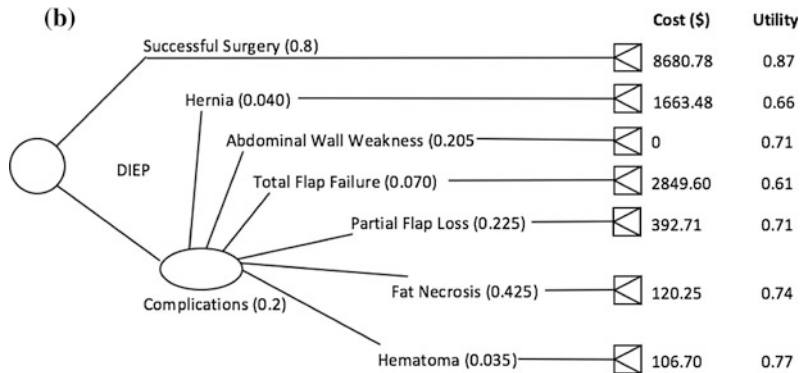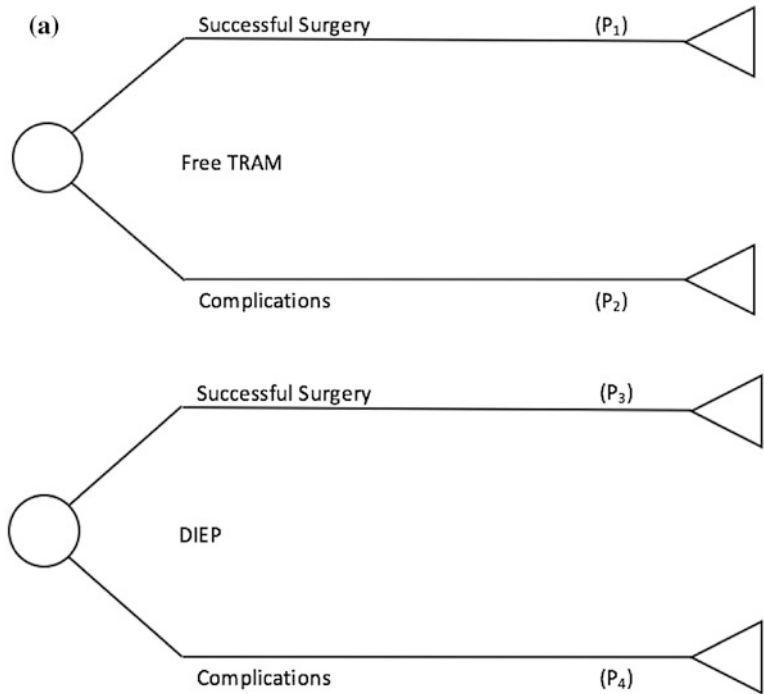
- ICER = $40,000/successful breast reconstruction after mastectomy
- ICER = $50, 00/successful knee replacement

How should the third-party payer decide? This is akin to comparing apples and oranges. This dilemma is obviated by the CUA where a common outcome (QALYs) is used

by both specialties and the decision is easier to make.

For Cost–Utility Analysis

This is a full economic evaluation, recommended to be used to inform reimbursement policy

making [3, 4]. Here, we calculate the Incremental Cost–Utility Ratio (ICUR).

Hypothesis testing is done via the *p*-value. In economic evaluations, one performs a sensitiv-

$$
\begin{aligned}
\text{ICUR} &= \frac{\Delta C}{\Delta U} \\
&= \frac{\left(\begin{array}{c} \text{Mean Cost}_{\text{novel surgical intervention}} \\ -\text{Mean Cost}_{\text{comparative surgical intervention}} \end{array}\right)}{\left(\begin{array}{c} \text{Mean QALY}_{\text{novel surgical intervention}} \\ -\text{Mean QALY}_{\text{Comparative intervention}} \end{array}\right)}
\end{aligned}
$$

(If the ICUR < \$50,000/QALY, this is a strong indication that the novel intervention should be adopted [8]). As a general guidance, it is probably better to use a Willingness-to-Pay (WTP) threshold as the \$50,000 figure has no theoretical basis and it is mostly used in the USA.

For Cost–Benefit Analysis

In this type of analysis, we attach a monetary value to the consequence of an intervention using a Willingness-to-Pay (WTP) approach.

$$\text{NSB}_i = \sum_{t=1}^{n} \frac{b_i(t) - c_i(\text{t})}{(1+r)^{t-1}}$$

where NSB = Net Social Benefit, t = year, $b_i(t)$ = benefits derived in year t, $c_i(t)$ = costs derived in year t and r = discount rate, n = lifetime of study.

(This type of economic evaluation is not used much in health care as it attaches a monetary value to the effectiveness of a health state, which health economists believe discriminates against the poor) [4].

Sensitivity Analysis

In clinical effectiveness studies, the results are reported by a point estimate such as a mean and a validity estimate such as a standard deviation.

ity analysis to determine how robust the conclusions are [4]. After the main analysis is performed based on the point estimate, the investigators redo the analysis based on one or two standard deviations around the point estimate of the costs or effectiveness or both, to see whether or not the conclusions of the study change. Probabilistic analyses with nonparametric "bootstrapping" with replacement are used in more advanced economic analyses [4, 9]. The conclusions are reported as robust if all calculations, main analysis, and sensitivity analyses favor the novel surgical intervention.

Now that you have mastered the important terminology and principles, we will help you understand how to appraise and understand a published economic evaluation report, by taking you through the following clinical scenario.

Clinical Scenario

At the hospital orthopedic rounds, a young surgeon asks his chief of service for an additional OR room for arthroscopic knee surgery. He claims there is an increasing demand for this useful procedure. His chief is skeptical because what he read that the benefits of arthroscopic surgery is controversial. He doubts he can persuade the hospital administration to invest money in this endeavor and asks his junior colleague to present supporting evidence that arthroscopic

surgery is cost-effective compared to just physical therapy from the patient's point of view.

Finding the Evidence

To identify the best evidence and inform his colleagues, the surgeon begins by conducting a literature search, as described in Chap. 3, in this book and according to the "Users" guide to the surgical literature: how to perform a high-quality literature search" [10].

The effectiveness of a surgical intervention can be found in a high-quality Randomized Controlled Trial (RCT) or a meta-analysis of number of RCTs. In this case, an RCT that compares arthroscopic surgery to optimal nonoperative therapy, in which the investigators coupled an economic evaluation, would provide the best evidence of cost-effectiveness. The surgeon follows the PICOT format, as described in Chap. 4, for the identification of important key words used in the search process:

Population: Patients with knee arthritis
Intervention: Arthroscopic surgery
Comparison: Physical Therapy (Physiotherapy or Optimal nonsurgical therapy)
Outcome: Cost-effectiveness
Time horizon: At least a year follow-up

We performed a literature search by choosing the filtered database, COCHRANE reviews and an unfiltered database, PubMed. With the PICOT format in mind, we used the search strategy: "osteoarthritis (P) AND arthroscopic surgery (I) AND physical therapy (C) AND cost-effectiveness analysis (O)". Between COCHRANE and PubMed, we identified 7 articles. We excluded articles because they did not include cost-effectiveness analyses, if they were not recent (last 10 years), or if they did not measure QALY.

One by Marsh et al. [11], which was listed in both databases, caught our attention. We read the abstract and we found it relevant to our purpose. This article includes both costs and effectiveness measured in QALYs and it is a full economic evaluation. It compares arthroscopic knee surgery and nonoperative methods. Now, we proceed to appraise the economic evaluation to determine whether the results are valid and whether they are relevant to our practice.

Are the Results Valid?

Did the Analysis Provide a Full Economic Comparison of Healthcare Strategies?

An economic analysis compares two or more healthcare interventions; in our case two surgical interventions to a surgical problem. Specifically, it compares the costs and the consequences (outcomes) of these interventions. The Marsh et al. [11] article is a full economic evaluation as they compared the costs in dollars and the effectiveness of the interventions, arthroscopy and a nonoperative method, using the Western Ontario McMaster Ostearthritis Index (WOMAC) and Quality-Adjusted Life Years (QALYs). It seems this is a full economic evaluation and we proceed to the next question.

Were Relevant Viewpoints Considered?

In an economic evaluation, there are a few possible perspectives (viewpoints). By this, we mean who bears the costs associated with the use of the new surgical procedures. In appraising an economic evaluation specifically, we ask who is benefiting from this study? Is it the patient, the hospital, the third-party payer or the society or others? There are occasions where a novel intervention may be found to be cost-effective from one perspective (i.e., the hospital) but not necessarily from another (i.e., the patient). For example, the hospital may save money by discharging patients earlier after a surgical intervention compared to an older program when the patients are hospitalized for a longer period but from the patient's point of view, this may be more costly if the patient's spouse has to take time off work to care for them during recovery at home.

In the Marsh et al. [11] study, we are told that the investigators conducted the cost-effectiveness analysis from the perspective of the Canadian healthcare payer and the societal perspectives. These perspectives are legitimate; however, we were concerned that they did not mention the patient's perspective. As the societal perspective also includes any out of pocket costs to the patient such as physical therapy, medications or assistive devices not covered by the provincial insurance plan and indirect costs such as time off employment, we wonder why they did not include explicitly the patient perspective. Their claim of taking a societal perspective may be incomplete, as other aspects, such as childcare and other familial obligations, could be different for the patient between the two interventions.

The perspective taken in an economic evaluation may depend on the question asked. The panel in cost-effectiveness in health strongly recommended that the societal perspective is the most important and should be considered, if possible [3]. Marsh and colleagues did this but as mentioned above, there may be some limitations to it [11].

Were All Relevant Clinical Strategies Compared?

When conducting an economic evaluation, it important that the investigators compare all relevant strategies for the condition under investigation. For example, the investigators should be comparing a novel procedure to a standard one and not one that is rarely used. In addition, investigators should be considering patients of different baseline risks. For example, if a general surgeon is performing a CEA on a novel hernia repair in the military population, this may not be generalizable to a retirement community population. If a CEA is performed comparing two approaches to hand surgery, one should consider both the Workers Compensation Benefit (WCB) patient and the non-WCB patient population, as the WCB patients are considered high-risk patients.

This is accomplished by doing a literature review of the population at risk. It seems that Marsh et al. considered different baseline risks as they performed a subgroup analysis [11]. This included patients with less severe radiographic disease (KL grade 2) and patients reporting mechanical symptoms of catching or locking. The clinical strategies they considered, arthroscopic versus nonoperative seems appropriate.

Were the Costs and Outcomes Properly Measured and Valued?

Was Clinical Effectiveness Established?

As mentioned in the introduction, there are two methodologically distinct types of economic evaluations, the model-based (deterministic) method, and the trial-based (stochastic) method. The preferred method is trial-based. In this type, patient-derived data are extracted from a well-executed RCT, in which the investigators collect costs from various perspectives parallel to the RCT. This type of economic evaluation provides high internal validity but at the expense of external validity as the subjects in the study may not be typical of community patients. If multiple RCTs exist, one can then pool the results in meta-analysis thus increasing generalizability because the pooled estimate of the effectiveness is derived from a wider spectrum of patients. To improve generalizability, one may relax the inclusion criteria in the RCT thus including patients that represent the whole population thus making this a pragmatic RCT and pragmatic economic evaluation.

The clinical effectiveness in the Marsh et al. article was measured by the WOMAC scale, which is a validated osteoarthritis instrument with total scores varying from 0 to 2400, higher scores indicating more pain and stiffness and reduced physical function [12–14].

Marsh et al. [11] also uses QALYs for performing a cost–utility analysis, the preferred type of economic evaluation by policy makers. To use QALYs, health utilities (such as HRQL or weights) are required [4, 7, 15]. Marsh et al. [11] used the Standard Gamble Technique to estimate health utilities. A health utility score is anchored

at 0 (death) and 1 (perfect health). The other measure taken into account to measure QALYs is the duration of the corresponding health state. We believe that the investigators measured the effectiveness of the competing interventions appropriately.

Were Costs Measured Accurately?

How the cost of the interventions is reported depends on the perspective taken, as certain aspects such as costs of physiotherapy will not be relevant if comparing a hospital perspective, as these services is usually provided outside the hospital. Reporting healthcare resource use in natural units and unit costs separately allows for appraisal and replication by others.

As mentioned previously, Marsh et al. [11] include two perspectives, the Canadian healthcare payer perspective and the societal perspective, in which they breakdown further considering both short-term and long-term costs (physical therapy vs. time off employment) directly or indirectly associated with the two treatments.

For the operative intervention, Marsh et al. [11] took the average procedure costs from the Ontario Case Costing Initiative, which includes things such as medical tests, operating room costs, equipment, and laboratory testing. For surgeon billing fee for each procedure, Marsh et al. [11] used the Ontario Schedule of Benefits. The investigators did not mention about anesthesia costs here, which we consider an important component of the procedure cost.

Direct medical cost estimates of the nonoperative intervention used information such as the number of physical therapy sessions attended by each patient, as well as medication (pain medication, anti-inflammatory medications, and hyaluronic injections) and device use (cost per unit obtained from the Ontario Drug Benefit Formulary) were considered [11].

Patient out of pocket costs include medications or devices that are not covered by any type of insurance program, Indirect costs cover items such as time off employment and caregiving

activities. The indirect costs and out of pocket costs combined are included in the "societal perspective". The out of pocket costs fall into the patient perspective, so we are perplexed as to why Marsh et al. [11] were not explicit on using the patient perspective separately.

Were Data on Costs and Outcomes Appropriately Integrated?

Studies that claim to be full economic evaluations often compare direct medical costs to each other, which does not necessarily mean that whichever one is the least costly is the most cost effective. Another common mistake is to take a ratio of cost and effect of the novel intervention and compare it to that of the comparative intervention.

To determine whether a novel surgical treatment is cost-effective, one needs to calculate an incremental cost–utility ratio (ICUR). This integrates the costs and effectiveness of the competing interventions, telling us what the extra unit of benefit is for each extra unit of cost. This is precisely what the investigators in the Marsh et al. study did. They calculated the marginal cost per marginal unit of utility [11]. This measure divides the difference in the mean cost of the novel and comparative treatment by the difference in the mean effectiveness, which in this case is QALY.

The investigators also report an Incremental Cost-Effectiveness Ratio (ICER), to look at the other outcome, WOMAC. This equation is identical to the one used for ICUR (described above) other than the denominator:

$$\text{ICER} = \frac{\left(\begin{array}{c}\text{Differences in costs between}\\\text{Intervention 1 and Intervention 2}\end{array}\right)}{\left(\begin{array}{c}\text{Differences in health effects between}\\\text{Intervention 1 and Intervention 2}\end{array}\right)}$$

These measures are critical because they represent the treatment which has the greatest incremental cost per unit gained in either QALY or pain and mobility measures from the WOMAC index. In other words, the higher the

value of ICER/ICUR, the greater the cost to improve the outcome (patient health). Acceptance or rejection of novel surgical technologies is based on this. It also depends on the patient and the circumstances at hand, such as frequency of the intervention and ability of the healthcare system to support it. Ultimately, some type of threshold of acceptability will be decided upon by a consensus of experts.

In the Marsh et al. [11] study, there was also an estimate of the total cost for each patient over the 2-year follow-up, therefore discounting (accounting for the difference in cost presently versus the cost in the future) was not necessary [11]. Therefore, it seems that costs and outcomes were appropriately integrated.

Was Appropriate Allowance Made for Uncertainties in the Analysis?

It is imperative to determine whether the ICER/ICUR values actually represent cost-effectiveness, as there are many values that can fall around the mean. This can be accomplished by recalculating the ICER/ICUR using both the best- and worst-case scenarios, referred to as a sensitivity analysis. If the conclusion on cost-effectiveness stays the same, it can be decided with more confidence that the treatment is truly cost-effective. Marsh et al. included a sensitivity analysis, in which they used either extreme of their 95% CI surrounding the mean differences in WOMAC scores and QALY. They estimated ICER and ICUR values that assumed the highest possible treatment effect observed in their sample, favoring either added arthroscopy or nonoperative treatments only, followed by Cost-Effectiveness Acceptability Curves (CEAC). We believe they satisfied this criterion.

The economic evaluation by Marsh et al. also includes the CEACs, which indicate the probabilities of an intervention being cost-effective at various Willingness-to-Pay (WTP) values [11]. WTP values represent the amount one is willing to spend per one unit increase in WOMAC or QALY value [16]. Marsh et al. [11] performed a Net Benefit Regression model (NBR) for each

outcome (WOMAC and QALY), by WTP (up to the clinically relevant threshold of $100,000 per unit gained), and stratified by perspective (Health care and Societal).

Are Estimates of Costs and Outcomes Related to the Baseline Risk in the Treatment Population?

As patients who are considered "high risk" are generally more likely to benefit from a treatment than those who are considered "low risk" , it is imperative to divide the population to reflect these groups to determine if there is in fact a difference in cost and benefit between groups.

Marsh et al. [11] included a table outlining the baseline characteristics of the population by intervention group (operative vs. nonoperative) and they were very similar between groups. However, the investigators could have categorized the continuous variables (i.e., age and BMI). This would provide a better sense of the different baseline risks or benefits of one treatment over the other for older individuals or those considered "over-weight" or "obese". Those who have less severe disease (KL grade 2) and patients reporting mechanical symptoms of catching or locking may not be the only groups to consider [11].

What Are the Results?

What Were the Incremental Costs and Outcomes Between the Two Strategies?

Marsh et al. [11] found a statistically significant difference in mean cost between groups, health-care payer perspective and societal perspective (Table 23.2). In terms of outcomes, there were differences in both outcomes WOMAC (favoring surgery) and QALY (favoring nonoperative) between intervention groups. However, these were not statistically significant ($p = 0.87$ and $p = 0.72$, respectively; Table 23.2).

The net benefit regression models (WOMAC and QALY) did not indicate arthroscopic surgery

Table 23.2 Cost and effect of outcomes (WOMAC and QALY)

	Surgery[a]	Nonoperative[a]	Incremental difference[b]
WOMAC			
Baseline	1222.91 (478.16)	1355.26 (548.92)	−132.35 (−24.58 to 289.29), 0.10
24-month	1526.45 (623.83)	1510.77 (570.21)	15.69 (−198.35 to 166.98), 0.87
Utility			
Baseline	0.79 (0.22)	0.80 (0.21)	0.01 (−0.06 to 0.07), 0.85
24-month	0.84 (0.23)	0.86 (0.16)	0.02 (−0.04 to 0.08), 0.47
QALY	1.64 (0.40)	1.66 (0.30)	−0.02 (−0.09 to 0.13), 0.72
Cost[c]			
Healthcare payer perspective	2633.25 (574.43)	737.40 (542.93)	1895.85 (1716.13–2075.57), <0.01
Societal perspective	3825.60 (1443.48)	1614.22 (1784.94)	2211.38 (1716.04–2706.51), <0.01

Note Table 23.2 was taken directly from Marsh et al. [11]
QALY quality-adjusted life yearn, *WOMAC* Western Ontario and McMaster Universities Osteoarthritis Index
[a]Mean (SD)
[b]Mean difference between groups (95% CI), *p* value
[c]2014 Canadian dollars

as a cost-effective alternative to nonoperative methods from either perspective at all levels of WTP (Table 23.3).

Do Incremental Costs and Outcomes Differ Among Subgroups?

In terms of cost and outcomes in the subgroup analysis, surgery was also not cost-effective in terms of either WOMAC or QALY. This was seen in both subgroups chosen by the investigators (patients with less severe disease and patients reporting mechanical symptoms of catching or locking) at all the levels of WTP.

How Much Does Allowance for Uncertainty Change the Results?

Marsh et al. stated that the ICER value was $140.94 from the societal perspective and $120.83 from the healthcare payer perspective, per one-point improvement on the 2400 point WOMAC total score, translating to $28,188 (societal) and $24,166 (payer) for a clinically important improvement (200 points). Additionally, the ICUR was −$110,569 from the societal

perspective and −$94,792.50 from the healthcare payer perspective per QALY gained, where the negative value indicates paying more and not getting a better outcome (surgery costs more but is less effective than nonoperative care). As reporting of economic evaluations with negative values is not intuitive, health economists prefer to use terms such as "dominant strategy", "win-win "scenario, "lose-lose" scenario. In the Marsh et al. [11] study, the ICUR from the societal perspective should have been reported as a "lose-lose" strategy as surgery was more costly and less effective (Figs. 23.2 and 23.3).

When we compare a novel surgical intervention to a prevailing technology, there are nine possibilities, which are illustrated in Fig. 23.2. The novel intervention may be more, same or less effective than the prevailing intervention illustrated on the horizontal axis of Fig. 23.2. The vertical axis illustrates whether the novel intervention is more, same or less costly than the prevailing intervention. If a novel intervention falls in *cell 1*, we adopt the new surgical technology, as it is more effective and less costly. Using the same reasoning, if it falls in cell 2, we reject it, as it is less effective and more costly. Most novel interventions fall in cell 7, where new technologies are more

Table 23.3 Net-benefit regression analysis results

(a) WOMAC

WTP[a]	Healthcare payer[b]		Societal[b]	
	Incremental net benefit	95% CI. p-value	Incremental net benefit	95% CI. p-value
0	−1179.20 (386.56)	−1942.58 to −415.82, <0.01	−1670.507 (66.18)	−2978.30 to −362.71, 0.01
1500	−352,418.73 (332804.27)	−1,009,643.31 to 304,805.85, 029	−369,151.74 (334,686.56)	−1,030,156.42 to 291,852.95, 0.27
2000	−469,498.57 (443753.31)	−1,345,826.20 to 406,829.05, 0.29	−491,645.48 (446,248.44)	−1,372,984.45 to 389,693.49. 0.27
2500	−586,578.42 (554702.36)	−1,682,009.13 to 508,852.29, 029	−614,139.22 (557,810.37)	−1715812.60 to 487534.16, 0.27
5000	−1,171,977.63 (1109447.72)	−3,362,923.98 to 1,018,968.72, 0.29	−1,226,607.94 (1,115,620.38)	−3,429,953.97 to 976,738.10, 0.27
10,000	−2,342,776.07 (2218938.56)	−6,724,753.89 to 2,039,201.76, 0.29	−2,451,545.36 (2,231,240.70)	−6,858,237.29 to 1,955,146.56, 0.27
20,000	−4,684,372.93 (4437920.28)	−13,448,413.80 to 4,079,667.94, 0.29	−4,901,420.22 (4,462,481.46)	−13,714,804.20 to 3,911,963.77, 0.27
30,000	−7,025,969.80 (6656902.01)	−2,0172,073.70 to 6,120,134.15, 0.29	−7,351,295.06 (6,693,722.26)	−20571371.20 to 5,868,871.04, 0.27
40,000	−9,367,566.66 (8875883.74)	−26,895,733.70 to 8,160,600.36, 0.29	−9,801,169.94 (8,924,963.07)	−27,427,938.20 to 7,825,598 33, 0.27
50,000	−11,709,163.50 (11094865.47	−33,619,393.60 to 10,201,066.58, 0.29	−12,251,044.80 (11,156,203.89)	−3,428,450,520 to 9,782,415.63. 0.27
60,000	−14,050,760.40 (13313847.20)	−40,343,053.60 to 12,241,532.79, 0.29	−14,700,919.70 (13,387,444.70)	−41,141,072.20 to 11,739,232.93, 0.27
70,000	−16,392,357.30 (15532828.94	−47,066,713.50 to 14,281,999.01, 0.29	−17,150,794.50 (15,618,685.51)	−47,997,639.20 to 13,696,050.23, 0.27
80,000	−18,733,954.10 (17751810.67)	−53,790,373.50 to 16,322,465.23, 0.29	−19,600,669.40 (17,849,926.33)	−54,854,206.30 to 15,652,867.53, 0.27
90,000	−21,075,551.00 (19970792.40)	−60,514,033.40 to 18,362,931.44, 0.29	−22,050,544.20 (20,081,167.14)	−61,710,773.30 to 17,609,684.84, 0.27
100,000	−23,417,147.90 (22189774.14)	−67,237,693.40 to 20,403,397.66, 0.29	−24,500,419.10 (22,312,407.96)	−68,567,340.30 to 19,566,502.14, 0.27

(b) QALY

WTP[c]	Healthcare payer[b]		Societal[b]	
	Incremental net benefit	95% CI. p-value	Incremental net benefit	95% CI, p-value
0	−2020.18 (558.61)	−3123.38 to −916.98, <0.01	−2048.89 (946.17)	−3917.66 to −180.11, 0.03
1500	−2226.59 (608.74)	−3427.75 to −1023 38, <0.01	−2250.38 (1006.93)	−4239.15 to −261.59, 0.03
2000	−2294.02 (645.94)	−3569.69 to −1018.34, <0.01	−2317.54 (1039.73)	−4371.10 to −263.98, 0.03
2500	−2362.48 (691.03)	−3727.19 to −997.77, <0.01	−2384.70 (1077.94)	−4513.73 to −255.67, 0.03
5000	−2704.77 (991.35)	−4662.60 to −746.96, 0.01	−272052 (1331.67)	−5350.69 to −90.34, 0.04
10,000	−3389.38 (1735.70)	−6817.21 to 38.45, 0.05	−339214 (2007.52)	−7357.19 to 572 91, 0.09
20,000	−4758.57 (3339.13)	−11,353.01 to 1835.87, 0.16	−4735.40 (3560.87)	−11768.45 to 2297.66, 0.18
30,000	−6127.77 (4972.26)	−15,947.50 to 3691.96, 0.22	−6078.65 (5181.15)	−16311.90 to 4154.60, 0.24

(continued)

Table 23.3 (continued)

(b) QALY

WTP[c]	Healthcare payer[b]		Societal[b]	
	Incremental net benefit	95% CI. p-value	Incremental net benefit	95% CI, p-value
40,000	−7496.98 (6613.14)	−20,557.26 to 5563.32, 0.26	−7452.72 (6820.82)	−20,893.66 to 6049.85, 0.28
50,000	−8866.17 (8257.13)	−25,173.19 to 7440.86, 0.29	−8765.16 (8468.63)	−25,491.49 to 7961.16, 0.30
60,000	−10,235.27 (9902.70)	−29,792.22 to 9321.50, 0.30	−10,108.42 (10,120.60)	−30,097.54 to 9880.70, 0.32
70,000	−11,604.56 (11549.16)	−34,413.03 to 11,203.90, 0.32	−11,451.68 (11,774.98)	−34,708.34 to 11,804.99, 0.33
80,000	−12,973.76 (13196.19)	−39,034.94 to 13,087.42, 0.33	−12,794.93 (13,430.88)	−39,322.15 to 13,732.30, 0.34
90,000	−14,342.96 (14,843.59)	−43,657.60 to 14,971.67, 0.34	−14,138.19 (15,087.81)	−43,937.98 to 15,661.62, 0.35
100,000	−15,712.16 (16,491.26)	−48,280.77 to 16,856.45, 0.34	−15,481.44 (16,745.45)	−48,555.23 to 17,592.53, 0.36

Note Table 23.3 was taken directly from Marsh et al. [11]

QALY quality-adjusted life year, *WOMAC* Western Ontario and McMaster Universities Osteoarthritis Index, *WTP* willingness-to-pay

[a]WTP for a one-point improvement on the WOMAC total score

[b]Incremental net benefit (SD)

[c]WTP for an additional QALY

Fig. 23.2 Nine possible outcomes when comparing the new surgical technique and the conventional technique (the numbers within each cell are illustrative only). *Note* This figure was adapted, with permission, from Thoma et al. [17]
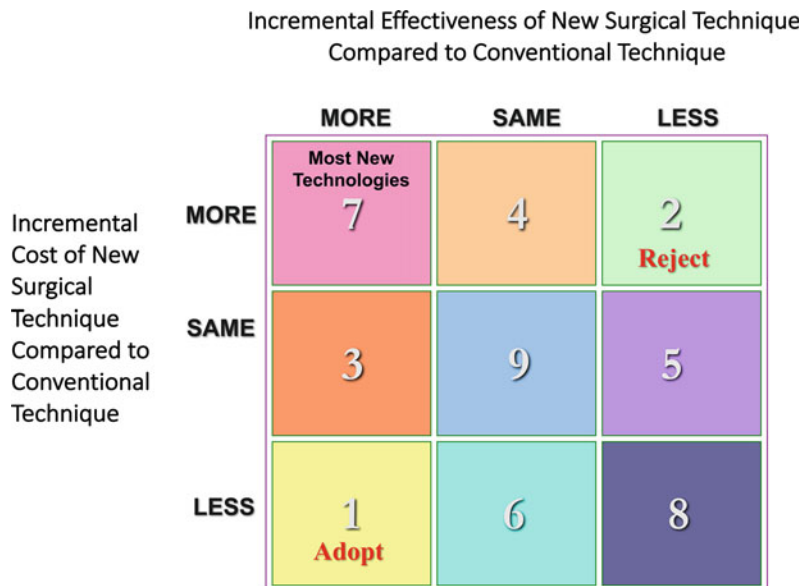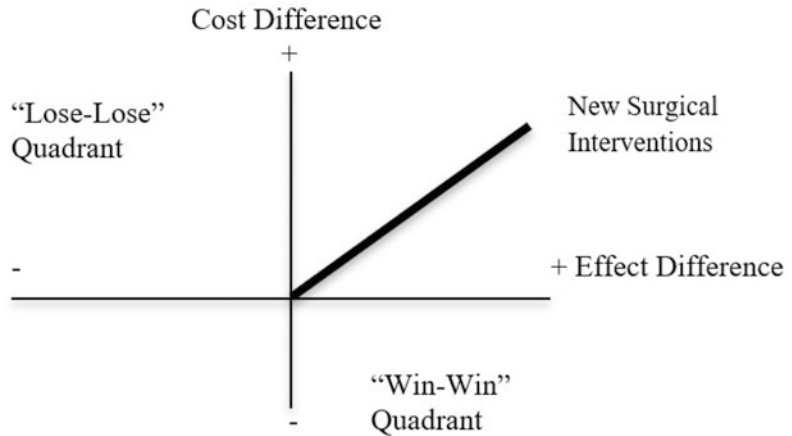
Fig. 23.3 Cost effectiveness plane. *Note* This figure was adapted, with permission, from Thoma et al. [18]

effective but also more costly. It is here that economic evaluations needs to be performed to find if indeed the novel interventions are cost-effective.

Another way of explaining these possibilities is the cost-effectiveness plane shown in Fig. 23.3, with the effectiveness on the *x*-axis and the cost on the *y*-axis. If the novel intervention falls in the right lower quadrant, we have a win–win situation meaning it is more effective and less costly. Alternatively, if it falls in the left upper quadrant, we have a lose–lose situation meaning it is less effective and costs more. The slope of this line will determine its acceptability or not. This slope is a ratio, which is precisely what the ICER/ICUR represents.

Will the Results Help Me in Caring for My Patients?

Are the Treatment Benefits Worth the Harms and Costs?

When we compare two surgical interventions (novel versus standard), surgeons need to understand that there are nine possibilities, as explained in Fig. 23.2. If the novel intervention falls in cell 1, we accept the novel intervention, as it is more

effective and less costly. If it falls in cell 2, we reject it, as it is less effective and more costly. We use a similar reasoning for the other scenarios. In most cases of modern surgery however, most new innovations are more costly and at the same time more effective. This is similar to a new intervention falling in the right upper quadrant of the cost-effectiveness plane (Fig. 23.3). It is under these circumstances that we need to perform a CEA or CUA. In the past, if a new innovation had an ICUR of $20,000/QALY the recommendation was given to accept it [19]. In recent years, this figure has increased to $50,000/QALY [8]. The above were proposed thresholds in the literature. A more official threshold is the one proposed by The National Institute for Health and Clinical Excellence (NICE), which suggests anywhere from $27,000 to $41,000 to accept a new innovation [20].

Marsh et al. [11] found that arthroscopy was, in fact, less effective and more costly than nonoperative approaches. Surgery in itself also carries risks not encountered in nonoperative approaches, such as anesthetic complications, deep venous thrombosis, and pulmonary embolism [11]. Although these are uncommon, they can have lethal outcomes. In deciding on which approach to take, one may consider these risks.

Could a Clinician's Patients Expect Similar Health Outcomes?

In looking at Table 1 of the Marsh et al. article, and in particular, the demographic characteristics of their patients pertaining to age, BMI, and Kellgren–Lawrence osteoarthritis severity grade, we believe that their patients are similar to ours and therefore have no reason to dispute their findings. If on the other hand our patients were much older and the majority of our patients had a different ratio of severity, we may be more skeptical of the findings shown here.

Can I Expect Similar Costs?

The costs of arthroscopy and physical therapy may differ in different jurisdictions (provinces, states, and countries) and these should be considered seriously. If you believe that in your specific jurisdiction the costs are similar, then you should consider adopting the findings from this study. If on the other hand, the medical costs related to surgery or physical therapy are different in your geographic area, then you should recalculate the costs based on the health-related resource units provided by the authors of this paper. For example, you can calculate the cost of physical therapy for the average patient by multiplying the number of physical therapy visits by the cost in dollars per visit. It is therefore imperative for investigators not to just present the costs in an article but also state the resource units consumed for the two comparative interventions. From these data, we can recalculate the ICUR in our setting and decide for ourselves if the surgical option is cost-effective or not.

Resolution of the Scenario

Based on the baseline ICER and ICUR, we see that the surgical approach fell into the lose–lose quadrant of the cost-effectiveness plane and therefore we are not prepared to accept surgery over physical therapy for mild arthritis of the knee. We also believe that the young orthopedic

surgeons' recommendation to spend more resources for this procedure is not supported by the evidence.

Final Thoughts and the CHEERS Statement

The EQUATOR network was established to enhance the quality and transparency of health research [21)]. They provide guidelines or checklists for various types of research including randomized trials, systematic reviews, and economic evaluations.

Clinical investigators who perform economic evaluations are encouraged to report their studies by following the CHEERS statement [21]. The CHEERS statement is a guideline specifically a checklist that covers all methodological aspects of an economic evaluation (title, introduction, methods, results, discussion, etc.). Although this guideline is not an appraisal instrument in itself, it does ensure that investigators cover all the elements of an economic evaluation. Familiarization with this guideline will ensure better quality study and eventual report.

References

1. Chatterjee A, Payette MJ, Demas CP, Finlayson SRG. Opportunity cost: a systematic application to surgery. Surgery. 2009 July;146(1):18–22.
2. Ziolkowski NI, Voineskos SH, Ignacy TA, Thoma A. Systematic review of economic evaluations in plastic surgery. Plast Reconstr Surg. 2013;132(1):191–203.
3. Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press; c1996.
4. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for economic evaluation of health care programmes. 3rd ed. United Kingdom: Oxford University Press; 2005.
5. OECD. Tackling wasteful spending on health. [Internet]. Paris: OECH Publishing; 2018 [cited 2018 April]. Available from: http://dx.doi.org/10.1787/9789264266414-en.
6. Thoma A, Veltri K, Khuthaila D, Rockwell G, Duku E. Comparison of the deep inferior epigastric perforator (DIEP) and free transverse rectus abdominis myocutaneous (TRAM) flap in post-mastectomy

reconstruction: a cost-effectiveness analysis. Plast Reconstr Surg. 2004;113:1650–61.

7. Thoma A, McKnight L. Quality adjusted life year (QALY) as a surgical outcome measure. A primer for plastic surgeons. Plast Reconstr Surg. 2010;125 (4):1279–87.

8. Neumann PJ, Cohen JT, Weinstein MC. Updating COST-Effectiveness—the curious resilience of the $50,000-per-QALY threshold. N Engl J Med. 2014;371(9):796–7.

9. Thoma A, Kaur MN, Tsoi B, Ziolkowski N, Duku E, Goldsmith CH. Cost-effectiveness analysis parallel to a randomized controlled trial comparing vertical scar reduction and inverted T-shaped reduction mammaplasty. Plast Reconstr Surg. 2014;134(6):1093–107.

10. Waltho DA, Kaur MN, Haynes RB, Farrokhyar F, Thoma A. Users' guide to the surgical literature: how to perform a high-quality literature search. Can J Surg. 2015;58:349–58.

11. Marsh JD, Birmingham TB, Giffin JR, Isaranuwatchai W, Hoch JS, Feagan BG, et al. Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. BMJ Open. 2016;5:1–10.

12. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.

13. Ehrich EW, Davies GM, Watson DJ, Bolognese Ja, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol. 2000;27:2635–41.

14. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care Res. 1999;12:172–9.

15. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96(1):5–21.

16. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002;11:415–30.

17. Thoma A, Sprague S, Tandan V. Users guide to the surgical literature: how to use an article on economic analysis. Can J Surg. 2001;44:347–54.

18. Thoma A, McKnight L, Knight C. The use of economic evaluation in hand surgery. Hand Clin. 2009;25:113–23.

19. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. CMAJ. 1992;146(4):473–81.

20. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics. 2008;26(9):733–44.

21. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Eur J Health Econ. 2013;14 (3):367–72.