SCIENTIFIC REVIEW

Cost-Effectiveness in Global Surgery: Pearls, Pitfalls, and a Checklist

Mark G. Shrime[1,2] · Blake C. Alkire[1,2] · Caris Grimes[3] · Tiffany E. Chao[4] ·
Dan Poenaru[5,6,7] · Stéphane Verguet[8]

Abstract

Introduction Cost-effectiveness analysis can be a powerful policy-making tool. In the two decades since the first cost-effectiveness analyses in global surgery, the methodology has established the cost-effectiveness of many types of surgery in low- and middle-income countries (LMICs). However, with the crescendo of cost-effectiveness analyses in global surgery has come vast disparities in methodology, with only 15% of studies adhering to published guidelines. This has led to results that have varied up to 150-fold.

Methods The theoretical basis, common pitfalls, and guidelines-based recommendations for cost-effectiveness analyses are reviewed, and a checklist to be used for cost-effectiveness analyses in global surgery is created.

Results Common pitfalls in global surgery cost-effectiveness analyses fall into five categories: the analytic perspective, cost measurement, effectiveness measurement, probability estimation, valuation of the counterfactual, and heterogeneity and uncertainty. These are reviewed in turn, and a checklist to avoid these pitfalls is developed.

Conclusion Cost-effectiveness analyses, when done rigorously, can be very useful for the development of efficient surgical systems in LMICs. This review highlights the common pitfalls in these analyses and methods to avoid these pitfalls.

✉ Mark G. Shrime
   shrime@mail.harvard.edu

1   Program in Global Surgery and Social Change, Harvard
    Medical School, Boston, MA, USA

2   Office of Global Surgery and Health, Massachusetts Eye and
    Ear Infirmary, Boston, MA, USA

3   Kings Centre for Global Health and Kings Health Partners,
    Kings College, London, UK

4   Department of Surgery, Massachusetts General Hospital,
    Boston, MA, USA

5   MyungSung Medical College, Addis Ababa, Ethiopia

6   Montreal Children's Hospital, Montreal, Canada

7   Department of Surgery, McGill University, Montreal, Canada

8   Department of Global Health and Population, Harvard T.H.
    Chan School of Public Health, Boston, MA, USA

Introduction

Cost-effectiveness analysis (CEA) can be a powerful policy-making tool. By promoting value-based decision-making, CEAs have helped guide policies ranging from human papilloma virus vaccinations [1] to mammography [2] to drug approval decisions [3].

In global surgery, CEAs are proliferating. Beginning in the 1990s with analyses in cataract [4] and trachoma surgery [5], continuing with McCord and Chowdury's seminal evaluation of a Bangladeshi hospital [6], through Gosselin et al.'s [7] evaluation of surgical services in a small hospital in Sierra Leone and Kahn et al.'s [8] study in male circumcision, and culminating in two systematic reviews of the cost-effectiveness of surgery in low- and middle-income countries (LMICs) [9, 10], this body of work has firmly established that surgery can be a cost-effective intervention [9, 10]. By expanding CEA beyond just medical interventions to system-wide analysis, CEAs also

served as a cornerstone for the economic argument for surgery in the *Lancet* Commission on Global Surgery [11] and the third edition of the World Bank's *Disease Control Priorities in Developing Countries* [12].

With the crescendo in CEAs in global surgery, however, a review of the standards for analysis is timely. A recent systematic review of CEAs in global surgery found that, out of 26 published articles, only 4 were able to fulfill all the criteria presented in one commonly used guideline [9]. Discrepancies in methodology lead to discrepancies in results—sometimes varying as much as 150-fold—limiting generalizability and preventing policymakers from making value-based comparisons of surgical interventions. This review will give an overview of CEA methodology, highlight common pitfalls in performing these analyses, attempt to reconcile sometimes divergent guideline recommendations [13–16], and establish standards for rigorous analyses in global surgery.

Theoretical basis of CEA

It is often helpful to begin at the end. When multiple intervention options face a limited budget, CEA provides a framework with which to assess which option (or options) will maximize benefit per dollar spent. The most basic analysis compares the costs and outcomes of at least two hypothetical scenarios: continuing the status quo and implementing one or more interventions.

The fundamental output of any CEA, through which this comparison is made, is the incremental cost-effectiveness ratio (ICER):

$$ICER = \frac{c_a - c_b}{e_a - e_b} \qquad (1)$$

where c_a and c_b represent the costs of scenarios A and B and e_a and e_b their outcomes. In all analyses, the cost and outcome of the status quo (scenario B above) must be assessed. The importance of assessing the incremental (or marginal) benefit of an intervention cannot be stressed enough; failing to do so leads to incorrect conclusions (see "Appendix").

The ICER represents how much it would cost society to buy one unit of benefit from scenario A, over the benefit already obtained in scenario B. It can be compared to how much society is *willing* to pay for an additional unit of benefit to determine whether intervention A is "cost-effective." ICERs may be used to compare multiple interventions against each other. In this case, the interventions are ordered from least expensive to most expensive, and the ICER is calculated for each sequential intervention. More detail on multiple interventions is given in the "Appendix".

The lower the ICER, the more cost-effective an intervention will appear, putting a heavy onus on the analyst to avoid anything that biases the ICER downward [17]. To protect against this bias, as well as to maintain transparency and improve reproducibility, all assumptions that underlie the analysis should be made explicit.

This review will highlight methods to minimize common pitfalls in CEA leading to downward bias. These pitfalls occur in five broad categories, each of which is necessary for undertaking a complete cost-effectiveness analysis: the analytic perspective, cost measurement, effectiveness measurement, probability estimation, valuation of the counterfactual, and heterogeneity and uncertainty.

One note on what this review will not cover. Other methods of cost analysis have been proposed, including cost-minimization analyses and benefit–cost analyses (BCAs). The former are less useful: They simply focus on the numerator of Eq. (1), thereby making the implicit assumption that the denominator is 1. BCAs, on the other hand, are common in global surgery, and the recommendations discussed below apply. However, BCAs take the additional step of valuing the denominator monetarily— that is, assigning a dollar value to a health benefit. Doing requires additional assumptions around, for example, willingness to pay for each life saved [18]. Discussing these assumptions is beyond the scope of this review.

Analytic perspective and intervention definition

Analyses may be undertaken from multiple perspectives— that of the patient, the hospital, the payer, or society as a whole. Every published cost-effectiveness guideline recommends the latter [13–16],[1] because adopting a societal perspective maximizes inter-analysis comparability and generalizability of findings.

Although other perspectives may be informative for specific questions, they should only be done secondarily and, if used, should be explicitly identified.

This has an important corollary. In surgery, especially, interventions are not delivered in isolation. A CEA on a surgical mission to fix hernias in Madagascar does *not* establish the cost-effectiveness of hernia surgery. It establishes the cost-effectiveness of a surgical mission to fix hernias in Madagascar, when compared to current hernia treatment strategies in Madagascar. Generalizing from the specific intervention platform to surgery as a whole is impossible.

Measuring cost

Although accurate costing is imperative for generalizability, it is in costing that the societal perspective is most commonly lost. Measures of effectiveness are designed to

---

[1] Revised United States Preventative Services Task Force recommendations are due in late 2017; whether this recommendation persists in the United States remains to be seen.

be distributions across populations and time. When costs are calculated for smaller analytic frames—for the hospital [7] or for the mission providing the surgery [19, 20]—the numerator in Eq. (1) is artificially decreased, introducing a downward bias into the ICER.

From a societal perspective, costs accrue to the health-care system, the provider, and the patient. These must all be included.[2]

Health system costs

Any health system costs specific to an intervention must be included. A new task-sharing program requires training. Including only the costs of already-trained task sharers, without taking into account the costs of training, would be an underestimate.

Provider costs

Two types of provider costs exist. Fixed costs are those that must exist for an intervention to occur, and do not vary whether one surgery is done or many. Variable costs increase as the number of surgeries increases.

Fixed costs

Capital costs—for example, the cost of a hospital building, an operating room, an anesthesia machine—must be allocated to the evaluated program. For example, if an anesthesia machine was being used exclusively for Cesarean sections, a new laparoscopic surgery program would have to include its cost, to the extent that it is used for laparoscopic surgery. Any capital should be annualized across the lifetime of the program, taking into account its resale value (if resale is possible) and discounting over time [14].

Labor costs are often large, so ignoring them biases the results. In general, medical professionals' salaries and benefits are an appropriate approximation, despite the fact that, in many LMICs, local medical professionals are likely underpaid. In global surgery, an additional wrinkle exists in the fact that many interventions are provided by volunteer organizations. As such, the opportunity cost of the volunteer surgeons and staff—effectively, their foregone salary for being on the mission—must be included.

---

[2] One important exception exists to this rule. If a cost is completely identical in both the intervention and its comparator, it can be ignored. This is because the numerator is a subtraction. As an example, if patients undergoing the intervention incur a $100 cost for a consultation, which is identical to a $100 cost for consultation in patients undergoing the comparator, then the numerator becomes:

$100 + c_{intervention} - (\$100 + c_{comparator}) = c_{intervention} - c_{comparator}$

and the $100 drops out.

Variable costs

Variable costs include anything used on a per-case basis: medications, supplies, and operating room time, for example, even if donated. If per-hour operating room time cost is available, it should be used. If not, the total aggregate cost of the operating room can be allocated proportionally to the time spent performing the evaluated interventions.

Patient costs

Patients face three sets of costs: direct medical, direct non-medical, and indirect. Direct costs are any for which a patient must reach into her pocket, while indirect costs are losses in opportunity. All direct costs must be included; the WHO recommends excluding indirect costs in the main analysis, but including them, if available, in secondary analyses [15]. Care should be taken not to assume that "free care" is always free to the patient [21].

Direct medical costs

In addition to the costs of intraoperative medications and supplies, direct medical costs include pre- and post-operative medications, laboratory and radiographic testing, blood transfusions, and any other cost a patient faces because she had surgery. In high-income countries, per-day hospital admission costs are often available from the literature [22]; for LMICs, the WHO's *Choosing Interventions that are Cost-Effective* (WHO-CHOICE) provides country-specific costs for inpatient and outpatient bed days [23].

Direct non-medical costs

Patients also have to pay to *get to* care. These direct non-medical costs can sometimes be larger than the medical costs themselves [24]. They include, at minimum, transportation, food, and lodging (if not provided by the hospital), and the "informal payments" often required for care to be delivered [25].

Indirect costs

Time is the most commonly cited indirect cost. An hour spent lying in a hospital bed is an hour spent *not* doing something else. The value of that "something else" is the opportunity cost of the hour spent in the hospital bed.

In high-income countries, time is often valued in terms of prorated wages [16]. In countries without a fully developed labor market, income is not an appropriate measure of economic worth. Economists have used household expenditures as its proxy [26].

Caregiver costs

One final note: Direct and indirect costs accrue to both patients and their caregivers. If it is common for caregivers to accompany patients, as is often the case for surgery, costs for the caregiver's transportation, food, and lodging must also be included. Their indirect costs should also be included, if this secondary analysis is being performed.

Standardizing costs

To maximize generalizability, costs may be reported in local currency units but should also be reported in a standardized currency comparable across analyses. The reference cost for most CEAs is the inflation-adjusted, standardized US dollar. Getting from the local currency unit to this standardized cost measure requires conversion across both time and space. Cost conversions can be complicated, and any uncertainty would be best resolved by consultation with a global health economist.

Adjusting across time (inflation)

In the USA, adjustment for inflation can be performed with the medical portion of the consumer price index [28]. For global health CEAs, the WHO recommends GDP deflators instead [29]. To adjust 2010 to 2014 costs, the deflators are used as follows

$$c_{2014} = c_{2010} \cdot \frac{G_{2014}}{G_{2010}} \tag{2}$$

where c_y represents cost in year y and G_y represents that year's GDP deflator.

Adjusting across space (purchasing power)

When converting out of local currency, it is tempting to use prevailing market exchange rates. This is not completely correct.

For costs of items that can be traded across borders (e.g., medications, instruments, machines), conversion from the local currency unit to US dollars is correctly done using the prevailing exchange rate at the time the purchase was made. For non-tradable goods, however (e.g., salary, a day in the hospital), costs must be converted using the purchasing power parity exchange rate, which the World Bank defines as the rate at which a dollar "would buy in the cited country a comparable amount of goods and services a US dollar would buy in the USA" [27]. For example: $1 in Madagascar buys far more than $1 in the USA, and the analyst must take this "purchasing power" into account. The World Bank provides purchasing power parity (PPP)

conversion factors, [29] of which the PPP factor for private consumption (as opposed to for GDP) should be used.

Adjusting across space and time

When adjusting across both space and time, use local GDP deflators to adjust for inflation first, then convert to international dollars using the target year PPP conversion factor.

Discounting

All future costs and benefits must be discounted, to reflect the relative importance for individuals of present costs and benefits over future costs and benefits (i.e., it is not a correction for inflation). Discounting a future stream of costs is calculated as follows:

$$\text{Total Cost} = \sum_t \frac{c_t}{(1 + r)^t} \tag{3}$$

where c_t represents the cost in year t ($t = 0$ for the current year) and r represents discount rate. Although there is significant controversy about whether costs and health benefits should be discounted at the same rate, common practice is to set $r = 0.03$ [14] and to vary r in sensitivity analyses to determine the sensitivity of the results to discounting.

How long to discount

Many analyses discount over the lifetime of the patient. Note that because average life expectancy is the life expectancy at birth, it is an underestimate for a patient who has survived past infancy. Country-specific life tables [30] give life expectancies *by age group* and should be used instead.

Measuring effectiveness

The most commonly utilized measures of effectiveness in global surgery are lives saved, years of life lost, quality-adjusted life years (QALYs) gained, or disability-adjusted life years (DALYs) averted. Although there are theoretical reasons to question these common measures [31], and although DALY calculations have changed in the last two decades, discussions of this controversy are beyond the scope of this paper.

The QALY was developed by Pliskin, Shepard, and Weinstein in 1980 [32] as a summary measure for length and quality of life:

$$\text{QALY} = qT \tag{4}$$

Table 1 Disability weight estimates [6]

DW	Severity of disease/condition
0.9	Needs assistance with activities of daily living such as eating, personal hygiene, or toilet use
0.8	Needs assistance with instrumental activities of daily living such as meal preparation, shopping, or housework
0.6	Limited ability to perform most activities in all of the following areas: recreation, education, procreation, or occupation
0.4	Limited ability to perform activities in two or more of the following areas: recreation, education, procreation, or occupation
0.2	Limited ability to perform most activities in one of the following areas: recreation, education, procreation, or occupation
0.1	Limited ability to perform at least one activity in one of the following areas: recreation, education, procreation, or occupation

where q represents a health state valuation, scored on a scale of 0 (death) to 1 (perfect health) and T represents time spent in that health state. Discounting a stream of future QALYs can be done similarly to Eq. (3).

In contrast, the DALY, developed in 1994 [33], is a "gap" measure, quantifying a loss from perfect health and perfect longevity. It is a sum of full years of life lost (YLL) due to premature mortality and full years of life lived in disability (YLD):

$$DALY = YLL + YLD \tag{5}$$

YLL is straightforward. For the purposes of CEA, YLDs should be calculated from an incidence perspective [34], such that:

$$YLD = dT \tag{6}$$

Note that this is different from the prevalence-based calculations used in burden of disease studies. For DALYs, d ranges from 0 (perfect health) to 1 (death)—the exact opposite of q.

In practice, QALYs have been more frequently used in high-income country CEAs, while DALYs have been more frequently employed in global health analyses. Although there is no fundamental theoretical reason for this, we recommend continuing to use DALYs for CEAs in global surgery to maximize comparability.

Of note, nothing constrains what is placed in the denominator of Eq. (1). CEAs exist, for example, evaluating the cost per pregnancy averted [35]. The interpretability of non-traditional denominators is limited, however—how much *would* society be willing to pay, in this example, for an averted pregnancy?

Choosing appropriate disability weights

QALY weights [q in Eq. (4)] for many surgical conditions can be found in a searchable database maintained by Tufts University [36]. The most recent DALY DWs [d in Eq. (6)] can be found in Salomon et al.'s [37] *Lancet* article.

Unfortunately, surgery-specific DWs are sparse, making surgery CEAs difficult. This presents the analyst with an added challenge: Estimation of DWs is required, but overestimations of DW will, once again, introduce downward bias into the ICER.

A common response to this [38] is estimation based on a method proposed by McCord and Chowdhury [6] (Table 1), itself a slight simplification of a table in Murray's original 1994 paper [33]. This method is widespread, and although it has been suggested that it is validated [39], we could find no evidence of validation. Further, it is unambiguously subjective and routinely *over*estimates the disability weight of surgical conditions.

As an example of the overestimation, take cleft lip and palate. A reasonable argument may be made from Table 1 that these conditions should be assigned a disability weight of somewhere between 0.2 and 0.4—limited ability to perform activities in one or two of recreation, education, procreation, or occupation.

This would be an overestimate: formal evaluations have placed the DW for cleft lip around 0.1 [40], and the accepted DW for the Global Burden of Disease project is 0.122 (discussed next).

What if the disability weight does not exist?

The lack of DWs in surgery presents the field with vast research opportunity: direct DW elicitation is required. In the absence of directly elicited DWs, however, the Global Burden of Disease initiative recommends an estimation using DWs for "generalized illness" (Table 2). A patient with cleft lip and palate will have difficulty with speech (DW = 0.054) and will experience level 2 disfigurement ("a visible physical deformity that causes others to stare and comment. As a result, the person is worried and has trouble sleeping and concentrating"; DW = 0.072). Comorbid DWs combine multiplicatively:[41]

$$DW_{Total} = 1 - [(1 - DW_1) \cdot (1 - DW_2)] \tag{7}$$

giving a DW, in this case, of 0.122.

Finally, all DW calculation should be referenced against other DWs for (at least) face validity. Cleft palate has, in some papers, been estimated to carry a DW of 0.38 [42].

Table 2 GBD 2010 disability weights for general conditions (95% uncertainty ranges in parentheses) [37]

Abdominopelvic problem: mild	0.012 (0.005–0.023)
Abdominopelvic problem: moderate	0.123 (0.083–0.176)
Abdominopelvic problem: severe	0.326 (0.219–0.451)
Anemia: mild	0.005 (0.002–0.011)
Anemia: moderate	0.058 (0.038–0.086)
Anemia: severe	0.164 (0.112–0.228)
Disfigurement: level 1	0.013 (0.006–0.025)
Disfigurement: level 2	0.072 (0.048–0.103)
Disfigurement: level 3	0.398 (0.271–0.543)
Disfigurement: level 1 with itch or pain	0.029 (0.016–0.048)
Disfigurement: level 2, with itch or pain	0.187 (0.125–0.264)
Disfigurement: level 3, with itch or pain	0.562 (0.394–0.725)
Generic uncomplicated disease: worry and daily medication	0.031 (0.017–0.050)
Generic uncomplicated disease: anxiety about diagnosis	0.054 (0.033–0.082)
Kwashiorkor	0.055 (0.033–0.085)
Severe wasting	0.127 (0.081–0.183)
Speech problems	0.054 (0.034–0.081)
Motor impairment: mild	0.012 (0.005–0.022)
Motor impairment: moderate	0.076 (0.050–0.109)
Motor impairment: severe	0.377 (0.251–0.518)
Motor plus cognitive impairments: mild	0.054 (0.033–0.084)
Motor plus cognitive impairments: moderate	0.221 (0.141–0.314)
Motor plus cognitive impairments: severe	0.425 (0.286–0.587)

This implies that a year lived with a cleft palate is *worse* than a year lived without both arms (DW = 0.359), almost twice as bad as a year lived in blindness (DW = 0.195), and approximately equal to a year lived with both HIV and tuberculosis (DW = 0.399) [37].

Even with the most careful analysis, subjectivity is bound to remain. For this reason, we emphasize again that all assumptions must be clearly stated to allow the reader to draw the most accurate conclusions. In addition, a large opportunity for research exists in the elicitation and standardization of DWs for surgical disease.

Discounting and age-weighting

Although age-weighting and discounting are not without controversy, broad consensus exists that future benefits should be discounted at the same rate as future costs in at least the base-case scenario [15, 16]. Age-weighting was initially utilized by the WHO [15], but the updated Global Burden of Disease studies have done away with it [43]. The recommendation, then, is to present non-age-weighted, discounted DALYs in the base-case analysis, and to reserve age-weighting for scenario analyses, discussed below.

Probability estimation

The effectiveness of an intervention is based on the number of DALYs it averts. DALYs averted have previously been calculated as: [31]

$$DALY_a = YLL \times RD \times PST \tag{8}$$

for lethal conditions, and

$$DALY_a = YLD \times DW \times RPD \times PST \tag{9}$$

for non-lethal conditions, where RD represents the risk of death, PST represents the probability of successful treatment, and RPD represents the risk of permanent disability. Unfortunately, these are problematic, as will be discussed below.

Determining probabilities

CEAs in global surgery have used estimates for RD, RPD, and PST often been based on ill-defined ranges [38, 39]. For example, Gosselin et al. [39] assign the point estimates shown in Table 3 for RD.

This forces the analyst to guess probabilities. Even if the analyst guesses correctly, however, the table introduces systematic error unless the true RD is exactly 50%. If death risk is low (as is often the case), DALY estimates are inflated. If it is high, this procedure is an underestimate (Fig. 1).

Routinely overestimating a 25% probability as 31%, when multiplied over entire populations and multiple years, clearly results in an unacceptable error, especially when event probabilities are easily found in the literature or can be estimated from hospital records [44].

Avoiding oversimplification

A second problem with Eqs. (8) and (9) is their oversimplification. They assume that the only potential outcomes from treatment are cure, death, or an ill-defined "residual permanent disability." Equation (8) implies the decision tree in Fig. 2.

This ignores three very real issues:

1. the risk of complications from treatment
2. the fact that most conditions are not *only* lethal or *only* non-lethal, and
3. the change in the mortality risk after "unsuccessful" treatment.

A more complete tree is shown in Fig. 3.[3]
The true estimate of DALYs averted is:

---

[3] Note that YLD is included as a possibility after death with treatment. Although this is likely zero, including it in the calculation allows for accounting of any disability occurring between unsuccessful treatment and death.

Table 3 Probability estimation [39]

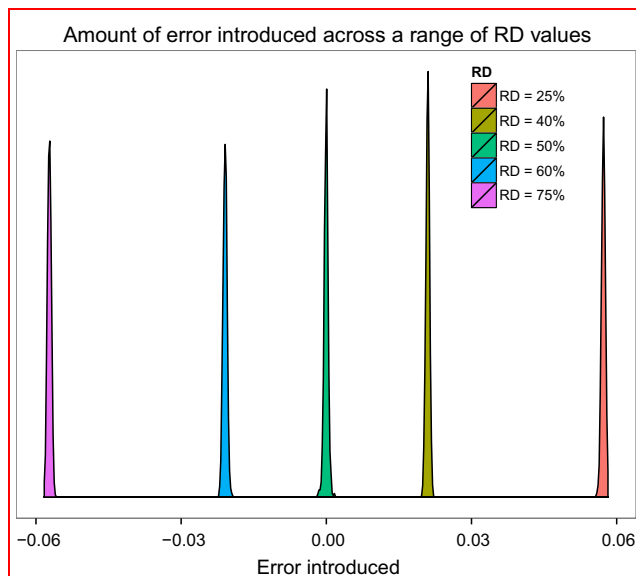| True probability of death	Estimated RD
>0.95	1.0
<0.95 and >0.50	0.7
<0.50 and >0.05	0.3
<0.05	0

Fig. 1 Error introduced by utilizing probability estimates found in Table 3. The distributions come from a microsimulation of 1000 cost-effectiveness analyses using the true probability compared with the estimates in Table 3. On average, very little error is introduced *if* the true probability is exactly 50%, with increasing error as the true probability rises or falls

$$\text{DALY}_a = \text{YLL}\big(\text{RD} - \text{RD}_{\text{postTx}}\big) + \text{PST}\big(\text{RD}_{\text{postTx}} \cdot \text{YLL}$$
$$+ \text{YLD}_{\text{dz}} - \text{pCompl} \cdot \text{YLD}_{\text{compl}}\big) \qquad (10)$$

Note that the value obtained from Eq. (10) is almost exclusively smaller than the value obtained from Eqs. (8) and (9), once again introducing downward bias into the ICER.

Importantly, even Eq. (10) is itself a simplification, since not all complications are created equal. As a result, we recommend against using simplified formulas and make the strong recommendation for the construction of decision trees, as in Fig. 3. Probability-based equations and/or decision trees must be applied to both the numerator (cost) and denominator (effectiveness) of Eq. (1).

Valuing the counterfactual

CEAs are used to answer two types of policy questions, and confusion between these two types has led to the propagation of errors. Published recommendations for CEA do not agree on which type of analysis is most appropriate [15, 16]. We make a recommendation in this paper, in line with the majority of the recommendations, but acknowledge that others disagree.

Which problem should be analyzed?

CEAs can answer either a "shopping spree" or a "competing choice" problem [45] (see "Appendix" for details).

The shopping spree problem assumes a health system does not exist, and asks which of a menu of *non-exclusive* options (e.g., surgery vs HIV treatment vs antimalarial care) should be in a newly constructed health system. On the other hand, the competing choice problem assumes a health system *does* exist and is already treating a condition, and asks, "Which of these (mutually exclusive) *methods* of treating this condition is most cost-effective?"

In the shopping spree problem, Eq. (1) may be simplified. Because the healthcare system is being designed *de novo*, there is no counterfactual to compare against, and the ICER simplifies to an *average* cost-effectiveness ratio:

$$\text{ICER} = \frac{c_a - 0}{e_a - 0} \rightarrow \frac{c_a}{e_a} \qquad (11)$$

This simplified cost-effectiveness ratio is almost exclusively smaller than an ICER.

Early global surgery CEAs had to advance an argument that surgery could be cost-effective and should be included in health systems. That is, the shopping spree problem needed to be solved. This argument has been made, and made strongly. Two systematic reviews [9, 10] as well as the *Lancet* Commission on Global Surgery [11] have all established that surgery is likely cost-effective.

As a result, however, the shopping spree problem is no longer relevant. This is doubly true if, in reality, a certain type of surgery is being performed. That procedure *has already been included* in the health system, making a simplified cost-effectiveness ratio misleadingly low.

In keeping with Gold [16] and the CHEERS consortium [13] (but in distinction with WHO [15]), we recommend that a true ICER be calculated, with all interventions measured against the status quo, *not* against the theoretical counterfactual of "nothing". It is much more relevant to evaluate platforms for surgical delivery than to spend research time answering a question that has already been answered.

The difference between an average CE ratio and an ICER is not small. Take, as an example, a hypothetical "surgical mission trip" to fix obstetric fistulas. Over 2 weeks, it repairs 20 fistulas, at a cost of $100,000 (similar to published valuations [46]). Making the (heroic) assumptions that every repair is successful, that no complications ensue, that no recurrence happens, that all patients are 18 years old, and that their life expectancy is an additional 40 years each, each repaired fistula nets
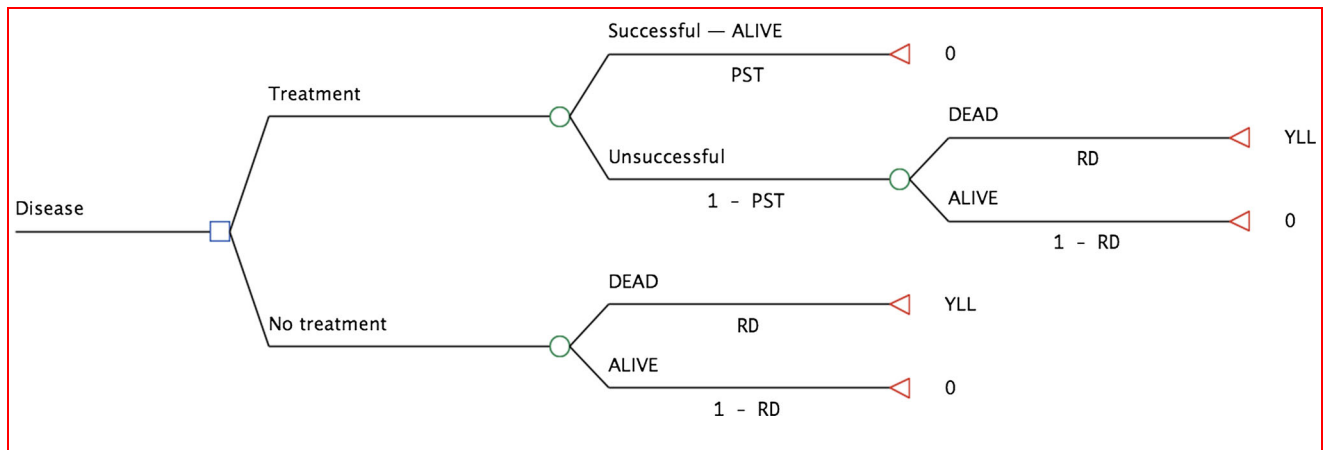
Fig. 2 Decision tree implied by usual estimates—in Eq. (8)—of DALYs averted
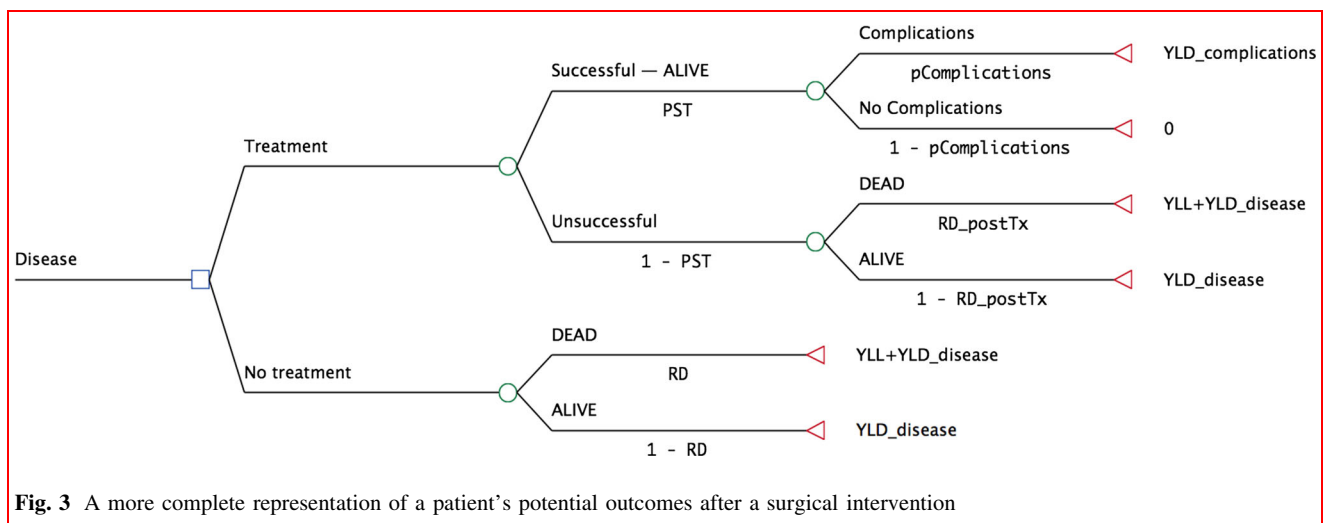

Fig. 3 A more complete representation of a patient's potential outcomes after a surgical intervention

$$\text{DALY}_a \sum_{t=0}^{39} \frac{0.338}{1.03^t} = 8.047 \tag{12}$$

DALYs averted (discounted at 3% per year, using the GBD 2010 DWs [37]).

The average CE ratio of the mission trip (Eq. 11) is

$$\text{CER} = \frac{\$100,000}{20 \times 8.047} = \$621.36 \tag{13}$$

However, a hypothetical district hospital repairs 5 fistulas a week, at a weekly cost of $5000 (at a much better average CE ratio of $124.27). The $100,000 spent on the mission trip could be used to scale up the repair of fistulas in this hospital. The question, then, is not whether the mission trip should be included in the health system, but whether the *additional* benefit gained from the mission trip is worth the *additional* cost:

$$\text{ICER} = \frac{\$100,000 - \$10,000}{(20 \times 8.047) - (10 \times 8.047)} = \$1118.40 \tag{14}$$

This implies that the additional 10 patients done by the mission trip actually cost $1118.40 per DALY averted—nearly twice the original estimate.

Addressing heterogeneity and uncertainty

Heterogeneity and uncertainty are not often addressed in global surgery CEAs.

Heterogeneity

Not all patients are identical, and this heterogeneity can substantively alter ICERs. For example, the following sets of life expectancies average to 40:

A : $\{40, 40, 40, 40, 40\}$
B : $\{10, 10, 50, 50, 80\}$

Using the same DWs and discount rate as in Eq. (12), an intervention in set A averts 40.23 DALYs, while an

Table 4 Checklist for CEAs in global surgery

Assumptions	
Assumptions are made explicit	[]
Assumptions which bias the ICER downward are avoided	[]
If the analysis is a benefit–cost analysis, value per statistical life assumptions are made explicit	[]
Analytic perspective and intervention definition	
The base-case analysis is from the societal perspective (other perspectives may be included as secondary results)	[]
Results are report for the intervention studied, including the platform and context for care delivery	[]
Results are not generalized beyond what is explicitly studied	[]
Measuring costs	
Which costs to include	
Costs to all levels of society are included	—
The health ministry	[]
The provider/hospital	[]
The patient's direct medical costs	[]
The patient's direct non-medical costs	[]
Indirect costs may be included, if available, in secondary analyses	[]
Fixed costs	
Capital costs are annualized across the lifetime of the capital, taking into account resale value and discounting	[]
Labor costs are explicitly detailed or are approximated by the salaries and benefits of the professionals in question	[]
Salaries and benefits of visiting surgeons are included, if they are involved	[]
Variable provider costs	
All variable costs are accounted for, including medications, supplies, and operating room time	[]
Patient costs	
Direct medical costs include anything for which a patient has to pay because of surgery	[]
Direct non-medical costs include transportation, food, lodging, and "informal payments" necessary to get care	[]
If caregivers commonly accompany patients, their direct costs are included	[]
Standardizing costs	
All costs are represented as inflation-adjusted US dollars, using GDP deflators for inflation, prevailing exchange rates for tradeable goods, and purchasing power parity conversion factors for nontradeable goods	[]
Discounting	
All future costs are discounted	[]
If a lifetime time horizon is used for discounting, age- and country-specific life tables are used to determine life expectancy	[]
Credibility	
The credibility of measured costs is checked against other available data	[]
Measuring effectiveness	
DALYs averted are the primary measure of effectiveness. QALYs, years of life lost, and deaths averted may also be used as the primary measure of effectiveness	[]
Other measures of health benefit may form secondary analyses	
Disability weights in the Global Burden of Disease studies are used if available. If the disability weight is unavailable, it is calculated from available data using a multiplicative formulation [see Eq. (7)]	[]
Subjective estimation of disability weights is avoided	[]
The credibility of disability weights estimates is confirmed by comparing against other disability weights of the same magnitude	[]
All future benefits are discounted at the same rate as future costs	[]
Non-age-weighting disability weights are used as the base-case (age-weighting may be treated in scenario analyses)	[]
Estimating probabilities	
Decision trees are used to represent all possible eventualities for patients in the analysis, and are applied to both costs and outcomes	[]
Probabilities are determined directly from data or from the literature	[]
Simplified and/or subjective probability estimates are avoided	[]
Valuing the counterfactual	
An incremental cost-effectiveness ratio, against the counterfactual of the status quo, is reported	[]

Table 4 continued

If a simplified, average cost/effectiveness ratio is reported—that is, if the counterfactual is "nothing"—a *strong* case has been made that the studied intervention is never performed in the region of interest	[]
Addressing heterogeneity and uncertainty	
Patient-level data are used to address heterogeneity. If patient-level data are not available, microsimulation methods may be used	[]
All parameters are subjected to one-way, two-way, or probabilistic sensitivity analyses	[]
Scenario analyses are included, as relevant	[]
ICERs are reported with appropriate uncertainty metrics	[]

intervention in set B averts only 34.37 DALYs. This difference alone would raise the ICER in Eq. (14) to $1309.

If individual patient data are available, they should be used instead of averages. In the absence of this level of granularity, heterogeneity can be dealt with in microsimulation models.

Uncertainty

Two layers of uncertainty exist: parameter uncertainty and scenario uncertainty. Mortality rates, risks of recurrence, probabilities of complications are all *estimates* and are therefore uncertain. Parameter uncertainty can be addressed in one-way, two-way, or probabilistic sensitivity analyses, which can be performed in both commercially available software and free statistical packages [14].

Scenario uncertainty allows for "what if?" questions to be asked. What if there were no complications? What if surgery was performed by task-shifted providers? What if DALYs were age-weighted but not discounted? These are less important but can be informative.

Conclusion

Cost-effectiveness analyses can be incredibly important for decision-making in global surgery, but methodologic discrepancies limit their utility. Early cost-effectiveness analyses in global surgery were incredibly important to demonstrate that surgery was a cost-effective intervention in the context of low- and middle-income countries. To that end, they were extremely successful. Future studies should, however, increase the methodologic rigor in cost-effectiveness analyses, and this review provides a discussion of the pitfalls in the field, the theoretical basis for rigorous cost-effectiveness analyses, and recommendations to standardize future analyses. These recommendations are summarized in the checklist in Table 4.

The goals of this checklist are fourfold: to encourage CEAs in global surgery and strengthen their quality, to maintain honesty and transparency, to avoid misleading results, and to maximize reproducibility and comparability—all in the service of making decision-making easier for policymakers in the field. Other checklists for CEAs have been developed; this checklist attempts to reconcile the differences among these divergent guidelines in a way that is relevant to global health broadly and to global surgery more specifically. Guidelines-based analyses can accomplish the goals above, and we encourage the use of the checklist in Table 4 in the design of future global surgery CEAs.

Appendix

The theoretical basis of the ICER

A more thorough discussion of the difference between the "shopping spree" and the "competing choice" problems follows.

The "shopping spree" problem [45] is answered by what Murray et al. [47] have called "generalized CEA." In this problem, the decision-maker *begins with a blank slate*. No health system exists—it will be created from scratch. The decision-maker is at a hypothetical health system "shopping mall", at which she gets to decide which of a number of non-exclusive *types* of interventions she should include in her shopping cart. Should she treat HIV? Should she include surgery? Should she provide vaccinations against human papilloma virus? Should she perform mammography? All sorts of options are available to her; her only constraint (but it is an important one) is that of her budget.

This sort of an analysis maximizes "allocative efficiency" [47]—that is, the money available to the decision-maker is spent in such a way as to maximize the benefit she is able to buy with it.

This may be best understood with a toy example. Assume the decision-maker is not creating a health system from scratch but shopping for her apartment. She has $10

to spend and has the following options at her disposal (each listed with its cost and the amount of "happiness"—in a unitless measure—it affords her. Assume for the purposes of this example that she knows both the cost and the happiness with certainty):

Item	Cost	Happiness	C/H
Toilet paper	$2	2	$1
Cereal	$4	3	$1.33
DVD	$10	5	$2
Wine	$15	3	$5
Ice cream	$4	10	$0.40

The final column of this table is the decision-maker's cost-per-happiness ratio. She should obviously spend her first $4 on ice cream, because this nets her the most happiness per dollar. Next most efficient is toilet paper, followed by cereal.

After purchasing these three items, she hits her budget constraint. Having spent all her money, she has gained 15 happiness points—the most she could gain with $10 (we leave it to the reader to convince themselves that there is no other combination of goods she can buy that will net her more than 15 happiness points for $10).

At this point, we can conclude that the decision-maker's "willingness-to-pay" ratio (how much she would pay for one unit of happiness) is $1.33, because that is the most she was able to pay for her last unit of happiness.

The translation into the allocatively efficient design of a healthcare system is obvious. However, we want to draw attention to the two settings in which this sort of decision-making occurs: It either occurs in the *de novo* design of a healthcare system or when a decision-maker has found herself with a sudden increase in her budget constraint. In settings in which neither of these two conditions holds, the shopping spree problem is irrelevant.

We also want to draw attention to the metric used in decision-making for the "shopping spree" problem. Instead of the ICER in Eq. (1), the shopping spree problem is answered by comparing simple cost-effectiveness ratios. This is because, with an empty shopping cart, the counterfactual in the ICER is "nothing". To be explicit:

$$\text{ICER} = \frac{c_a - 0}{e_a - 0} = \frac{c_a}{e_a}$$

The "competing choice problem" [47] is answered by what Murray et al. [47], have called "intervention-mix-constrained CEA". This sort of problem presents itself when *the health system is already up and running*, the budget constraint has not changed, and a new intervention that is *mutually exclusive* from an intervention already in

the health system presents itself for evaluation. In this case, an ICER is appropriate, as will be seen.

To return to the previous example. A competing choice problem exists when the decision-maker, having filled her shopping cart, encounters a second brand of toilet paper. This new toilet paper offers more happiness (2.5), but at a higher cost ($3). Should she swap out her old toilet paper for this new brand?

If the decision-maker uses the simple (non-incremental) cost-effectiveness ratio, she will make the wrong decision. The ratio for this new toilet paper appears favorable ($1.20, less than the $1.33 she is willing to pay), so, at first blush, it appears she should swap her old toilet paper.

However, with a budget constraint of $10, buying the new toilet paper *decreases* her overall happiness to 14.75.

What happened?

By using a simple cost-effectiveness ratio, she answered the wrong question. She already has toilet paper in her shopping cart—the question is not whether she should add this second brand to the already full cart (doing so would leave her very little money left for cereal), but whether she should swap the old toilet paper for the new. The two brands are competing choices.

It is more correct to ask, "Is the *added* happiness she gets worth the *added* cost she would have to pay?" That is, is the *incremental* benefit she gains from the substitution worth it to her. An ICER is appropriate:

$$\text{ICER} = \frac{\$3 - \$2}{2.5 - 2} = \$2$$

In this case, she would be effectively paying $2 for the additional happiness she gets from the new toilet paper. This is will above her willingness-to-pay ratio of $1.33; she clearly should keep her original basket of goods.

Again, the translation to global surgery is evident. As has been discussed in the main text of the paper, a CEA of a mission trip to perform hernia surgeries is not an analysis of hernia surgery in general, but an analysis of a mission trip to perform hernia surgeries. Unless the health system in the country of interest performs *no* hernia surgeries whatsoever, using simple cost-effectiveness ratios to evaluate this mission trip will vastly overstate its cost-effectiveness, when money may truly be better allocated improving the national surgical system instead.

Comparing multiple interventions

The ICER may be used to compare multiple mutually exclusive interventions. As a hypothetical example, we imagine three options for the treatment of thyroid diseases in a target country. The first is the status quo, which is hemithyroidectomy alone. The second couples subtotal thyroidectomy with short-term calcium supplementation,

while the third is total thyroidectomy with both calcium supplementation and thyroid hormone replacement.

We have ordered the interventions from least expensive to most expensive (and, incidentally, from least effective to most effective). The ICER for the subtotal thyroidectomy option, compared against the next most expensive option, is

$$\text{ICER} = \frac{\$1500 - \$1000}{0.9 - 0.8} = \frac{\$500}{0.1} = \$5000/\text{DALY averted}$$

Similarly, the ICER for the total thyroidectomy strategy is $24,000 per DALY averted. In a country willing to pay up to $10,000 per DALY averted, subtotal thyroidectomy is the correct option. For countries (as in the USA) willing to pay more than $24,000, total thyroidectomy is the correct option.

Of note, ordering by increasing cost does not necessarily guarantee an ordering by increased effectiveness:

Intervention	Cost	Incremental cost	Effectiveness (DALYs averted)	Incremental effectiveness	ICER ($/DALY averted)
Hemithyroidectomy	$1000	–	0.8	–	
Subtotal thyroidectomy	$1500	$500	0.9	0.1	$5000
Total thyroidectomy	$2700	$1200	0.95	0.05	$24,000

The numbers in this table are fabricated and are used only for the purposes of this example

Intervention	Cost	Incremental cost	Effectiveness (DALYs averted)	Incremental effectiveness	ICER ($/DALY averted)
Hemithyroidectomy	$1000	–	0.8	–	
Subtotal thyroidectomy	$1500	$500	0.75	–0.05	N/A
Total thyroidectomy	$2700	$1700	0.95	0.15	$11,333

In this hypothetical case, subtotal thyroidectomy is more expensive *and* less effective than hemithyroidectomy. An ICER is meaningless in this situation because subtotal thyroidectomy should not be chosen—for less money, the health system could get more benefit by simply doing hemithyroidectomies. Subtotal thyroidectomy is said to be "dominated" by hemithyroidectomy. As such, the ICER calculation is just between the two non-dominated options: hemithyroidectomy and total thyroidectomy.

References

1. Kiatpongsan S, Kim JJ (2014) Costs and cost-effectiveness of 9-valent human papillomavirus (HPV) vaccination in two East African countries. PLoS ONE 9(9):e106836
2. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG (2006) Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst 98(11):774–782
3. Lowson K, Jenks M, Filby A, Carr L, Campbell B, Powell J (2015) Examining the implementation of NICE guidance: cross-sectional survey of the use of NICE interventional procedures guidance by NHS trusts. Implement Sci 10:93
4. Marseille E (1996) Cost-effectiveness of cataract surgery in a public health eye care programme in Nepal. Bull World Health Organ 74:319–324
5. Evans TG, Ransom MK, Kyaw TA, Ko CK (1996) Cost effectiveness and cost utility of preventing trachomatous visual impairment: lessons from 30 years of trachoma control in Burma. Br J Ophthalmol 80:880–889
6. McCord C, Chowdhury Q (2003) A cost effective small hospital in Bangladesh: what it can mean for emergency obstetric care. Int J Gynaecol Obstet 81:83–92
7. Gosselin RA, Thind A, Bellardinelli A (2006) Cost/DALY averted in a small hospital in Sierra Leone: what is the relative contribution of different services? World J Surg 30:505–511
8. Kahn JG, Marseille E, Auvert B (2006) Cost-effectiveness of male circumcision for HIV prevention in a South African setting. PLoS Med 3:e517
9. Chao TE, Sharma K, Mandigo M et al (2014) Cost-eff ectiveness of surgery and its policy implications for global health: a systematic review and analysis. Lancet Glob Health 2:e334–e345
10. Grimes CE, Henry JA, Maraka J, Mkandawire NC, Cotton M (2014) Cost-effectiveness of surgery in low- and middle-income countries: a systematic review. World J Surg 38:252–263
11. Meara JG, Leather AJM, Hagander L et al (2015) Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet 386(9993):569–624
12. Mock C, Donkor P, Gawande AA et al (2015) Essential surgery: key messages from disease control priorities, 3rd edn. Lancet 385(9983):2209–2219
13. Husereau D, Drummond M, Petrou S et al (2013) Consolidated Health Economic Evaluation Reporting Standards (CHEERS)–explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health 16(2):231–250
14. Drummond MF (2005) Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford
15. Tan-Torres Edejer T, Baltussen R, Adam T et al (2003) Making choices in health: WHO guide to cost-effectiveness analysis. World Health Organization, Geneva
16. Gold MR, Siegel JE, Russell LR, Weinstein MC (1996) Cost effectiveness in health and medicine. Oxford University Press, New York
17. Weinstein MC, Toy EL, Sandberg EA et al (2001) Modeling for healthcare and other policy decisions: uses, roles, and validity. Value Health 4(5):348–361

18. Hammit JK, Robinson LA (2011) The income elasticity of the value per statistical life: transferring estimates between high and low income populations. J Benefit Cost Anal 2(1):1–29

19. Tadisina KK, Chopra K, Tangredi J, Thomson JG, Singh DP (2014) Helping hands: a cost-effectiveness study of a humanitarian hand surgery mission. Plast Surg Int 2014:921625

20. Egle JP, McKendrick A, Mittal VK, Sosa F (2014) Short-term surgical mission to the Dominican Republic: a cost-benefit analysis. Int J Surg 12(10):1045–1049

21. Xu K, Evans DB, Kadama P et al (2006) Understanding the impact of eliminating user fees: utilization and catastrophic health expenditures in Uganda. Soc Sci Med 62(4):866–876

22. Dasta JF, McLaughlin TP, Mody SH, Tak PiechC (2005) Daily cost of an intensive care unit day: the contribution of mechanical ventilation. Crit Care Med 33:1266–1271

23. WHO-CHOICE (2016) Country-specific unit costs. http://www.who.int/choice/country/country_specific/en/. Accessed 1 July 2016

24. Shrime MG, Dare AJ, Alkire BC, O'Neill K, Meara JG (2015) Catastrophic expenditure to pay for surgery worldwide: a modeling study. Lancet Global Health 3(Suppl 2):S38–S44

25. Ensor T (2004) Informal payments for health care in transition economies. Soc Sci Med 58(2):237–246

26. Morris SS, Carletto C, Hoddinott J, Christiaensen LJM (2000) Validity of rapid estimates of household wealth and income for health surveys in rural Africa. J Epidemiol Commun Health 54:381–387

27. World Bank. What is an "international dollar"? https://datahelpdesk.worldbank.org/knowledgebase/articles/114944-what-is-an-international-dollar. Accessed 10 Oct 2016

28. Bureau of Labor Statistics US (2015) Consumer price index databases. http://www.bls.gov/cpi/data. Accessed 3 July 2016

29. World Bank (2013) World development indicators. http://data.worldbank.org/. Accessed 4 July 2015

30. World Health Organization (2016) Global health observatory data repository: life tables by country. http://apps.who.int/gho/data/view.main.60000. Accessed 3 July 2016

31. Gosselin RA, Ozgediz D, Poenaru D (2013) A square peg in a round hole? Challenges with DALY-based "burden of disease" calculations in surgery and a call for alternative metrics. World J Surg 37(11):2507–2511

32. Pliskin JS, Shepard DS, Weinstein MC (1980) Utility functions for life years and health status. Oper Res 28(1):206–224

33. Murray CJL (1994) Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ 72(3):429–445

34. Schroeder SA (2012) Incidence, prevalence, and hybrid approaches to calculating disability-adjusted life years. Popul Health Metr 10:19

35. Babigumira JB, Stergachis A, Veenstra DL et al (2012) Potential cost-effectiveness of universal access to modern contraceptives in Uganda. PLoS ONE 7(2):e30735

36. Tufts University (2016) Cost-effectiveness analysis registry. https://research.tufts-nemc.org/cear4/SearchingtheCEARegistry/SearchtheCEARegistry.aspx. Accessed 5 July 2016

37. Salomon JA, Vos T, Hogan DR et al (2012) Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. Lancet 380(9859):2129–2143

38. Wu VK, Poenaru D (2013) Burden of surgically correctable disabilities among children in the Dadaab refugee camp. World J Surg 37(7):1536–1543

39. Gosselin RA, Maldonado A, Elder G (2010) Comparative cost-effectiveness analysis of two MSF surgical trauma centers. World J Surg 34(3):415–419

40. Bickler SW, Ozgediz D, Gosselin RA et al (2010) Key concepts for estimating the burden of surgical conditions and the unmet need for surgical care. World J Surg 34(3):374–380

41. Flanagan W, McIntosh C, Le Petit C, Berthelot JM (2006) Deriving utility scores for co-morbid conditions: a test of the multiplicative model for combining individual condition scores. Popul Health Metr 4:13

42. Poenaru D, Pemberton J, Frankfurter C, Cameron BH (2015) Quantifying the disability from congenital anomalies averted through pediatric surgery: a cross-sectional comparison of a pediatric surgical unit in Kenya and Canada. World J Surg 39(9):2198–2206

43. Institute for Health Metrics and Evaluation (2016) The global burden of disease: generating evidence, guiding policy 2013. http://www.healthdata.org/sites/default/files/files/policy_report/2013/GBD_GeneratingEvidence/IHME_GBD_GeneratingEvidence_FullReport.pdf. Accessed 5 July 2016

44. Shrime MG, Verguet S, Johansson KA, Jamison DT, Kruk ME (2015) Task-shifting, universal public finance, or both for the expansion of surgical access in rural Ethiopia: an extended cost-effectiveness analysis. Health Policy Plann. doi:10.1093/heapol/czv121

45. Hunink MGM, Weinstein MC, Wittenberg E et al (2014) Decision making in health and medicine: integrating evidence and values. Cambridge University Press, Cambridge

46. Gosselin RA, Gialamas G, Atkin DM (2011) Comparing the cost-effectiveness of short orthopedic missions in elective and relief situations in developing countries. World J Surg 35(5):951–955

47. Murray CJL, Evans DB, Acharya A, Baltussen R (2000) Development of WHO guidelines on generalized cost-effectiveness analysis. Health Econ 9:235–251