



Understanding cost-utility analysis studies in the trauma and orthopaedic surgery literature

Achi Kamaraj¹
Nikhil Agarwal²
K. T. Matthew Seah²
Wasim Khan²

- Cost-utility analysis (CUA) studies are becoming increasingly important due to the need to reduce healthcare spending, especially in the field of trauma and orthopaedics.
- There is an increasing need for trauma and orthopaedic surgeons to understand these economic evaluations to ensure informed cost-effective decisions can be made to benefit the patient and funding body.
- This review discusses the fundamental principles required to understand CUA studies in the literature, including a discussion of the different methods employed to assess the health outcomes associated with different management options and the various approaches used to calculate the costs involved.
- Different types of model design may be used to conduct a CUA which can be broadly categorized into real-life clinical studies and computer-simulated modelling. We discuss the main types of study designs used within each category. We also cover the different types of sensitivity analysis used to quantify uncertainty in these studies and the commonly employed instruments used to assess the quality of CUAs. Finally, we discuss some of the important limitations of CUAs that need to be considered.
- This review outlines the main concepts required to understand the CUA literature and provides a basic framework for their future conduct.

Keywords: cost-utility analysis; decision tree; incremental cost-utility ratio; quality-adjusted life years; quality of health economic studies; sensitivity analysis

Cite this article: *EFORT Open Rev* 2021;6:305-315.
DOI: 10.1302/2058-5241.6.200115

Introduction

With any musculoskeletal disease or injury, there are different potential management options available that have varying costs and benefits to the patient. A judgement can

be made as to which management option is the more cost-effective option compared to the others. This is usually achieved via calculating and then comparing the added health outcomes and associated costs over a specific time frame between two particular management options for a given condition. For example, one may consider which of resurfacing or non-resurfacing the patella in total knee arthroplasty (TKA) is the more cost-effective over a given duration. In trauma and orthopaedics, cost-effectiveness is becoming increasingly important due to rising healthcare spending, especially due to the ageing population, in the presence of many possible management options and a resource-limited environment.

A special case of cost-effectiveness analysis (CEA), the cost-utility analysis (CUA), is discussed in this review. A CUA is an economic analysis that compares the relative costs and health outcomes in quality-adjusted life-years (QALYs) of different management options and enables a judgement as to the more cost-effective option. Their aim is to minimize costs for the greatest possible justified increase in patient-reported health outcomes. The importance of cost-effectiveness analyses was initially emphasized to guide public health policies in developing countries. These CEAs were motivated partly by the belief that countries could achieve better health outcomes in quantity and quality simply by redirecting the limited resources available. In 1996, the First Panel on Cost-Effectiveness in Health and Medicine outlined the first consensus-based guidelines for the conduct of CEAs to improve the comparability and quality of these studies.¹ Therefore, CUAs offer important information that may guide decision-making by institutions in developing the most cost-effective clinical guidelines and practices.

However, in order to understand these studies, knowledge about several key areas is required. This includes an understanding of the methods used to quantify health outcomes and costs associated with a particular management option, the different CUA model types, the sensitivity

analyses used, and the methods employed for evaluating study quality. Below, we discuss these five areas as well as the limitations of CUAs.

Measuring health utility and incremental cost-utility ratio

In CUAs, healthcare outcomes are quantified using quality-adjusted life-years (QALYs). This value is a subjective self-reported measure of the perceived quality of life experienced by the patient over a given period of time. It is calculated by the utility (measure of worth or value in economics) of a given health state multiplied by the years lived in that state.² Thus, one QALY is equal to one year of perfect health. First, the QALY values following a management option and its comparator must be determined for a particular condition over a specific time frame. An incremental cost-utility ratio (ICUR) must then be calculated by dividing the difference in total costs between the two management options over the same specific time frame by the difference in the aforementioned reported QALYs for each. Thus, the ICUR gives the cost per additional QALY gained as a result of the management option over its comparator.³ It should be noted that many studies use the terms ICUR and ICER (incremental cost-effectiveness ratio) interchangeably.

The ICUR is then compared to the willingness-to-pay (WTP) threshold. In the European Union, the decision-makers' WTP is defined as the maximum a healthcare consumer is willing to pay per additional QALY gained by the more expensive management option. Furthermore, it is important to acknowledge that most countries do not set a threshold, and even in the few that do, it is not used as the sole criterion for decision-making. Other criteria may include the impact on budget, uncertainty around model estimates and the degree of unmet medical need. There is also significant variability in the thresholds set by the different countries that employ WTP thresholds, which may perhaps be attributed to the methodological differences used to establish them.⁴

Only the National Institute for Health and Care Excellence (NICE) in England and Wales explicitly uses a fixed threshold as its WTP, which is kept between £20,000 and £30,000.⁵ In contrast, other countries use specific figures or ranges as recommendations but have not formally adopted them. The commonly quoted WTPs for some of these countries are shown for comparison in Table 1 below. If the ICUR is below the WTP threshold, the management option is considered cost-effective relative to the comparator over the given time frame. Therefore, ICURs offer a valuable metric that may help inform policy-making in healthcare governing bodies.

The methods used to measure health utility values can be either direct or indirect.

Table 1. Examples of the ICUR thresholds used by CUA studies in various countries

Country	ICUR threshold (cost per QALY)
Ireland	€45,000 ⁶
The Netherlands	€10,000–80,000 ⁷
Spain	€30,000 ⁸
USA	USD50,000–150,000 ⁹
Australia	AUD69,000 ¹⁰

Notes. ICUR, incremental cost-utility ratio; CUA, cost-utility analysis; QALY, quality-adjusted life-years.

Direct measures of health utility

In time trade off (TTO), the patients choose between remaining in a given state of ill health for a given period of time or being restored to perfect health but trading in years off their life, e.g. living 10 years in a health state with severe knee pain or trading years off your life to live a shorter period in perfect health. If five of the 10 years are willing to be traded, then the knee pain health state has a utility value of 0.5. Rubén Mota used TTO in a Markov model to assess the cost-effectiveness of early primary total hip replacement (THR) for older adult patients with osteoarthritis against either non-surgical therapy followed by THR once the patient had progressed to a functionally dependent state, or non-surgical therapy alone from the Italian national health system perspective.¹¹

In Standard Gamble, the patient is asked to choose between two options. The first is a certain option whereby the person remains in a given health state A for 'X' years. The second is a gamble, an option with an element of risk whereby the person may either be returned to full health for the remaining 'X' years (with probability p) or dies immediately (with probability 1-p). The probability of death is altered until the respondent is indifferent to the two available options (point of indifference). If this point of indifference is found when the probability of death is, for example 80%, this would suggest that the individual values health state A to be 80% of full health.¹² This means the utility of health state A would be 0.8. Marsh et al used Standard Gamble to measure health utility for the calculation of the QALY in patients who received both arthroscopic debridement of degenerative articular cartilage and resection of degenerative meniscal tears in addition to non-operative management for knee osteoarthritis and in patients who received non-operative management only.¹³

The visual analogue scale (VAS) method offers the simplest and most subjective measure of health utility. Here the patient rates their own state of health, often in terms of pain, on a scale of 0 (no pain) to 100 (severe pain).¹⁴ Wang et al sought to evaluate the cost-utility of percutaneous endoscopic transforaminal discectomy (PETD) and percutaneous endoscopic interlaminar discectomy (PEID) for the treatment of L5-S1 lumbar disc herniation in a CUA using VAS.¹⁵

Indirect measures of health utility

The EQ-5D is a questionnaire designed by the EuroQol group for self-completion by the patient as an indirect measure of healthcare utility using a scoring system for each response.¹⁶ In this version, five dimensions are used as measures including mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The patients rate their degree of severity for each dimension using either the three-level (EQ-5D-3L) or five-level (EQ-5D-5L) scale. Following an additional EQ-VAS, the score from this section is then converted into a preference weight as a measure of health utility.¹⁶

The Short Form-36 (SF-36) Health Survey is a 36-item questionnaire of healthcare utility designed for self-completion by the patient.¹⁷ The SF-36 measures eight scales: vitality, physical functioning, bodily pain, general health, physical role functioning, emotional role functioning, social role functioning and mental health. It has been shown through component analysis that two distinct concepts are measured: a physical dimension and a mental dimension, represented by the physical component summary (PCS) and mental component summary (MCS), respectively.¹⁷ Each of the component scales contribute in different proportions to the calculation of these two summary measures. They offer valuable insight into a patient's current health state. However, it is worth noting that although many researchers continue to use and to extrapolate from these measures an overall single health utility, this is not supported by the SF-36 scoring manual.¹⁸ In osteoarthritis studies, the SF-36 is often combined with the Western Ontario and McMaster's Universities Osteoarthritis Index (WOMAC), or the Lequesne index, which measure function and pain.^{19,20}

Developed in 2005, the Osteoarthritis Knee and Hip Quality of Life questionnaire (OAKHQOL) was the first questionnaire developed to measure quality of life in patients specifically with knee and hip osteoarthritis.²¹ This instrument consists of 43 items in five dimensions (physical activity, mental health, social functioning, social support and pain) and three independent items. Each item is measured using a numerical self-reported rating from 0 to 10. If the individual fails to answer at least half the items for a particular dimension, the score for that dimension is dropped. Raw scores are obtained by first computing the mean of the item scores for each dimension. These are then normalized to a scale from 0 (worst) to 100 (best possible health utility). This instrument has been shown to cover the highest number of osteoarthritis core set categories and capture specific aspects that are very important for knee and hip osteoarthritis patients.²² These core set categories refer to the International Classification of Functioning, Disability and Health (ICF) core sets, which are short lists of ICF categories developed by experts that are important for patients with a specific disease. The ICF core set for osteoarthritis contains 55 categories.²³

Perspective used for calculating costs

The healthcare costs incurred in the management of medical conditions used in ICUR calculations can be considered from either the direct (healthcare) or indirect (societal) perspective. The direct costs comprise healthcare costs such as the primary and subsequent treatments, medication and hospital bed space.²⁴ In contrast, the indirect costs take into account the societal costs incurred as a result of the management option such as loss of productivity, absenteeism and informal healthcare costs.²⁵

The choice of the perspective used is dependent on the decision problem (e.g. an investment decision for a hospital or to inform a reimbursement request to the government). However, the European Network for Health Technology Assessment (EUnetHTA) recommends that all economic evaluations to be conducted from a healthcare perspective at a minimum. However, several countries recommend a societal perspective including Norway, Denmark, and Sweden.²⁶

Weeks et al evaluated only the direct healthcare costs when comparing the cost-utility of patellar resurfacing in TKA and non-resurfacing.²⁴ The costs were obtained from the finance department at the Schulich School of Medicine and Dentistry in Ontario, Canada. The procedure costs included the cost of the equipment, implant, theatres, time spent in theatres, length of stay in the hospital and the various medical and laboratory tests during the initial hospitalization period.

Buvik et al evaluated the cost-utility of telemedicine in remote orthopaedic consultations from the societal perspective.²⁷ Three types of costs were included: the costs of implementing and running the telemedicine service in clinical practice (e.g. screen and camera at the remote centre and hospital for videoconferencing), travel costs of the patient to the remote centre (considering time, distance and mode of transport) and production losses due to patients having to take time off work to attend the orthopaedic consultation.

In order to improve the accuracy of these calculated costs, they must be first corrected via inflation adjustment and discounting. These are outlined below.

Inflation adjustment

Healthcare costs are measured in various currencies depending on the country of origin of the data and may be accrued over many different years. This is important because, for example, in the UK, the value of a pound in one year may not hold the same value in another due to inflation or deflation. Thus, inflation adjustment must be carried out for these costs. The general recommendations would be to express the costs in values of the current (or most recent) year and that if older values are used, these should be adjusted for inflation using the appropriate

price index figure. There are several indices that could be used. For European countries, indices of consumer prices may be found on the Eurostat webpage.²⁸ Unadjusted values are ‘nominal’, whilst those that are adjusted are considered ‘real’. Conversion into ‘real’ values allows the comparison of costs incurred in different years within and between studies and is important for ICUR calculations.²⁹

Discounting

Another important consideration when calculating both costs and health outcomes of healthcare management options whose benefit materializes in the future is the use of discounting.³⁰ This is the adjustment of both costs and health outcomes to the ‘present value’ for the time at which they occur. Discounting is essential because people generally value future costs and health outcomes less than current costs and health outcomes. Thus, their values diminish the more distantly in the future that they occur. To perform this adjustment, the value of the costs or health outcome measures (e.g. QALY), for each year in the future (n), is multiplied by $(1/(1 + \text{discount rate})^n)$.³¹ Hence, greater discount rates or longer delays between treatment and benefit precipitate lower net present values for both costs and health outcomes. For example, discounting may be significant in the case of autologous chondrocyte implantation (ACI) to treat isolated full-thickness articular cartilage defects of the knee (upfront costs and delayed health benefits) or negligible in intra-articular steroid injections (where costs and benefits occur upfront simultaneously).

The discount rates stated in the national guidelines for economic evaluations should be applied. The EUnetHTA recommends the use of a discount rate between 3% and 5% for both costs and effects, as do most recommendations in European guidelines.²⁶ It is recommended that both costs and effects are discounted in the base case analysis with the same rate. It is also recommended that sensitivity analyses which explore the effect of varying the discount rate and the use of differential discount rates (e.g. a lower discount rate for benefits than costs) should be incorporated into the CUA.

Time horizon length

The time horizon length refers to the duration over which health outcomes (e.g. QALYs) and costs are measured and calculated. The same time horizon length should be used for both. An appropriate time horizon depends on the nature of the disease, the management option under consideration and the purpose of the analysis. As recommended by the EUnetHTA, it should be long enough to reflect all the important differences in costs and health outcomes between the management options being compared in the CUA.²⁶ Importantly, guidelines from certain

countries (e.g. those from the UK, Finland, Ireland, Norway and Spain) explicitly state that this may mean a lifetime if the management option leads to differences in health utility that persist for the remainder of a patient’s life. This may require extrapolation of available data, which is certainly relevant to many orthopaedic surgery procedures.³²

Different types of CUA model design

Different CUA model designs may be used to calculate the different costs and utility values of different management options over a given time horizon. These may be divided into either real-life clinical studies or computer simulations. According to the EU guidelines, the decision on whether to use a real-life clinical trial or a computer-simulation model depends on the research question. Model-based studies should be used when the management option effects are expected in the long term, when the incidence of the clinical endpoint is low or when it is not possible for all relevant alternative management options to be included in one study. A trial-based study is appropriate in other cases.³³ The most commonly used real-life clinical study types are randomized controlled studies and prospective cohort studies.

Randomized controlled trials

Randomized controlled trials are considered to be the gold standard of scientific research. They are longitudinal studies whereby a number of people are randomly assigned to two or more groups to test the effect of a particular management option. Each group will have different management options being tested on them. The outcome measures (i.e. costs and health utility values) are measured at specific times within the particular time frame being used. The aforementioned study by Buvik et al evaluated the cost-utility of remote telemedicine orthopaedic consultations compared to standard outpatient consultations at a hospital via a randomized controlled trial of 389 patients from 2007 to 2012.²⁷

Prospective cohort studies

Prospective cohort studies are similar to randomized control studies. However, the individuals assigned to each study group are not randomly selected but have a particular factor (e.g. treatment via a particular management option) in common, whose impact on health outcomes and costs is under study. Therefore, this type of study allows the impact of, in the case of CUAs, the management option used on these aforementioned outcome measures. Thus, the more cost-effective management option may be identified for a particular condition. Miyazaki et al, for example, conducted a prospective cohort study model of

47 patients with spinal metastasis who had a surgical indication from 2010 to 2014.³⁴ This cohort was divided into a group of 31 patients who underwent spinal surgery and another group of 16 patients who did not. Therefore, this study assessed the cost-effectiveness of surgical treatment compared to non-surgical management in patients with spinal metastasis.

Computer-simulation models allow the evaluation of the effect of a proposed management option used for patients without the need for a real-life clinical study. These models should use the most rigorous and up-to-date information for clinical parameters from the literature. The European Network for Health Technology Assessment recommends that if computer modelling is used, it should always be justified and it is imperative that it is presented with transparency so that it may be reproduced.²⁶

The three most frequently used types are discussed below.

Decision trees

Decision trees offer a diagrammatic illustration of a decision problem regarding which management option to use out of multiple possibilities, as faced by clinicians for a particular condition. They are the simplest of the commonly used decision modelling techniques. They are used to model problems that involve a series of choices which are in turn constrained by previous decisions. They consist of two key components: branches and nodes. Nodes represent the key elements of a decision problem in a computer model and the branches stemming from them represent their possible outcomes.³⁵

Nodes are divided into three types: decision nodes, probability nodes and terminal nodes. Decision nodes represent a possible decision that can be made by the clinician (e.g. which management option to use). Depending on the decision made, the patient's health state moves along one or more branches. Probability nodes represent probability-based events that follow a decision node or another probability node. For example, at these nodes, the disease may develop a certain complication with a given probability or progress without it. Terminal nodes represent the final outcome of a decision analysis that follows a certain combination of decisions and probability-based events. These outcomes include: cured, death, or a health state somewhere in between.³⁶ The branches extending out from these nodes are mutually exclusive and collectively exhaustive (must add up to one). In addition, for a certain final outcome, by multiplying along the nodes and branches taken to reach it, one may find the overall probability for it to occur following a clinical decision or a series of decisions. The health outcome measures of interest (e.g. health utilities) are attached to the distinct final outcomes of the tree. Costs may be attached to both events within the tree and to the final outcomes.

Therefore, for a particular pathway that follows a certain decision on which management option to use, the summation of the component costs involved until the final outcome is reached as a result of that pathway gives the respective total costs involved.³⁷ Zindel et al used a decision tree model to illustrate the decision problem of choosing rivaroxaban or enoxaparin sodium for thromboprophylaxis after total hip and knee replacement in the German healthcare setting.³⁸

However, decision trees (see Figure 1) can become very complex when used to model chronic disease because this will inevitably involve many lengthy pathways representing recurring events. This is very time-consuming both to interpret and analyse

Markov models

Markov models are the most common type of model used in economic evaluations of healthcare management options. These models assume the patient is always in one of a finite number of discrete 'Markov' states. Within a Markov model, events are modelled as transitions from one health state to another one. The time horizon used for the model is split into clinically meaningful time intervals, or 'cycles', of equal length. At the end of each cycle, a patient may either remain in the same health state or move to a consequent one. These transitions between health states continue until a patient enters an absorbing state. This refers to a state where, once entered, it cannot be left (e.g. death). The occurrence of these events is dictated by conditional probabilities. These probabilities, as with decision trees, are conditional upon the last health state the patient was in. However, it is important to acknowledge that transition probabilities may change with time.³⁷

For each of the health states modelled, health utility values can be attached which reflect the quality of life for the state. Likewise, costs are also attached to each health state reflecting the costs incurred of remaining in that particular state for the duration of one cycle. To calculate the QALYs for each patient for particular management options, the health utilities associated with each health state the patient experiences are multiplied by the associated time spent in them. These values are then summed across all health states experienced in the model. Similarly, the total costs involved for particular management options are calculated for each patient by multiplying the costs incurred by each health state the patient experiences by the time the patient spends in that state. These costs are then summed across all the experienced health states in the model.³⁷

Markov models (see Figure 2) are used when there are many potential health states with the possibility of bidirectional transition between them. This usually applies to the modelling of chronic disease. As mentioned previously,

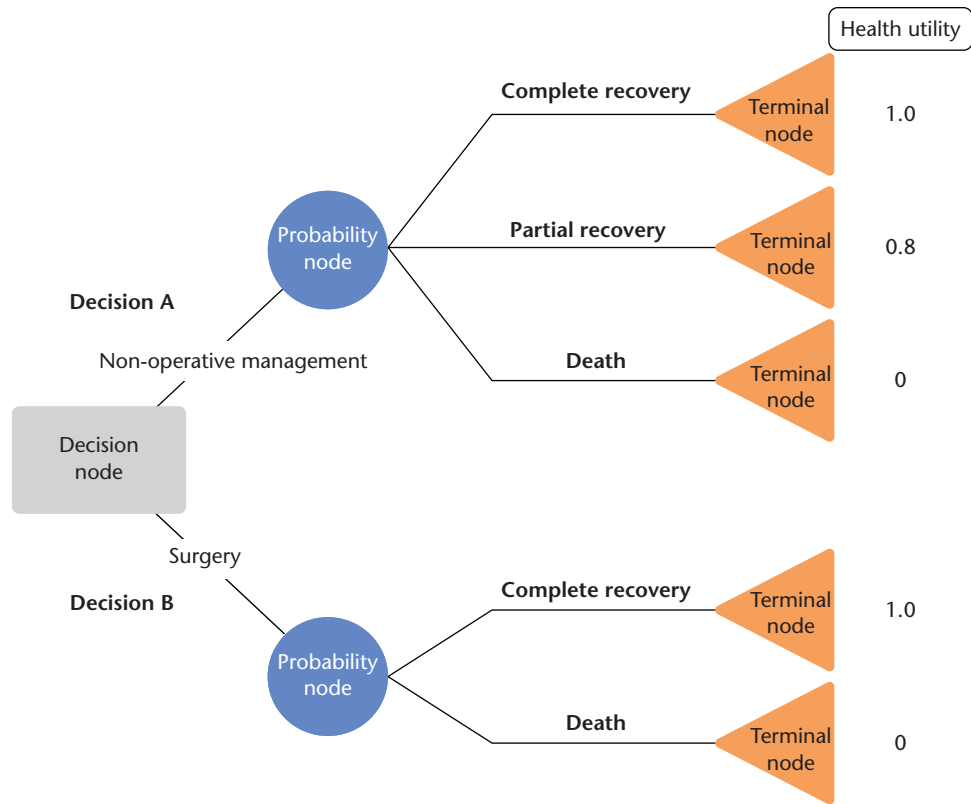


Fig. 1 An example of basic decision tree modelling.

decision trees will become far too complex to model such problems efficiently.

Health model microsimulations

Health model microsimulations are very detailed analyses that use highly realistic computer-simulated individuals that differ in various characteristics.⁴⁰ These include factors such as age, gender, ethnicity and educational attainment. The populations used are to reflect that of the

desired country of study and so are the most difficult to create. For example, Si et al used a microsimulation model for the cost-utility evaluations of various pharmaceutical and primary care management options to treat osteoporosis in the Australian population.⁴¹ These studies evaluate the effect of the particular management option on the entire population of interest. The model tracks individuals throughout their lifetime to see whether or how long the disease persists, what happens to it and the medical costs incurred throughout this time period.

These models are particularly useful when individuals have a mixture of interrelated (and often dynamic) risk factors that influence the experience of disease over time. This is because the model cycles throughout the lifespan of each individual one by one, twice, with aggregation occurring at the end. The first cycle just includes the base case. The second cycle includes the base case with the management option. Following this, the two cycles are compared to show the impact of each management option on health outcomes and differences in costs. These simulations are very similar to randomized controlled trials. However, microsimulations allow the understanding of the effects and costs of a particular management option over a much longer time horizon than a real-life study with regard to future costs, potential savings and improved health quality.⁴⁰

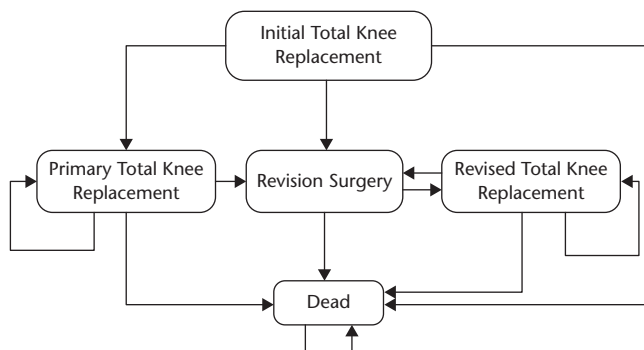


Fig. 2 An example of a Markov model.

Source: Adapted from Pennington et al, who used a Markov model to compare the cost-utility of five commonly used cemented brands of unconstrained prostheses with fixed bearings in total knee arthroplasty.³⁹

Downloaded from Bioscientifica.com at 02/09/2025 11:44:14PM via Open Access. This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed. <https://creativecommons.org/licenses/by-nc/4.0/>

Sensitivity analysis

A sensitivity analysis can be used to assess the level of confidence that a researcher may have in the conclusions of a CUA study. Within these studies, uncertainty may derive from the parameter values (e.g. due to sampling bias), assumptions made for unknown or parameter values within a given range, or the structure of the model itself. The objective of the sensitivity analysis is to appreciate uncertainty inherent in the study and allows the estimation of a confidence interval for the ICUR reported.

There are various sensitivity analysis methods. They all function by varying the parameter values used for which substantial uncertainty exists and then recording the results of the economic evaluation to see how they are affected. If the results change significantly in response, then these variables are likely to heavily influence the result of the economic evaluation. These potential sources of uncertainty may include the associated costs of the management option, the complication rates, and the quality-of-life estimates considered. Such values need to be stringently varied across a range of clinically plausible values to test the fidelity of the original results of the CUA where mean values were used. This allows the reviewer to determine which of these parameters are the key drivers of the results of the CUA and so they are very informative.⁴² To meet the preferences of most EU countries, the EUnetHTA guidelines recommend performing both deterministic (one-way, multi-way and threshold) and probabilistic sensitivity analysis.²⁶ The different types of sensitivity analysis are discussed below.

One-way analysis

In deterministic one-way sensitivity analysis, the uncertain input parameters are varied one by one by setting each to its upper and lower bounds. The impact of these changes in one parameter on the main outcomes of the primary analysis is then assessed with all other parameter values set at their mean values. This analysis gives an indication of the impact of individual parameter uncertainty on the main outcome measures. As a result of this input variation, the ICUR may increase, decrease or remain essentially the same.

Multi-way analysis

Deterministic multi-way analysis is very similar in principle to one-way analysis except that multiple parameters are varied simultaneously, and the outputs of the analysis are then measured. For example, Brauer et al used both one-way and multi-way sensitivity analysis in a cost-utility analysis of operative versus non-operative management of displaced intra-articular calcaneal fractures.⁴² Both revealed that the cost per additional QALY gained with operative management over non-operative management

was sensitive to the inclusion of estimates of costs due to time lost from work.

Threshold analysis

Deterministic threshold analysis attempts to identify the value of a given parameter where the output of the analysis changes sign regarding whether or not a particular management option is of a greater cost-utility compared to a comparator. The parameter is varied until this tipping-point is found for it with the remaining inputs kept at their mean values. Parameters that have tipping-points closer to their mean value indicate that they hold a strong influence over the outcome of the model.

Probabilistic analysis

In most computer-simulation models, each of the uncertain parameters (such as the health state transition probabilities) are assigned point estimate values. These are often based on meta-analysis data. Thus, they will inevitably have a range of possible values around them. For example, if the 95% confidence interval of a particular probability value (e.g. the probability for a management option to be successful) published in a meta-analysis is within a very close range, then greater certainty is indicated.⁴³

In probabilistic sensitivity analysis, instead of assigning a single value to each parameter, computer software is used to assign a distribution to all parameters in the model, which reflects the uncertainty in the true value. All parameters must remain practical (e.g. probabilities must remain between zero and one and costs cannot be negative). From each distribution, samples are then repeatedly drawn and are used as model inputs. Each unique combination of inputs (a single 'simulation vector') results in a unique combination of model outputs. As a result, by considering the results of many simulations, an estimation of the expected (mean) model outputs and the uncertainty associated around them may be derived.⁴⁴

Ponnusamy et al conducted a Markov study to assess the cost-utility of TKA against non-operative management in patients across six body mass index (BMI) cohorts. A probabilistic analysis was performed with an ICER of \$30,000/QALY which revealed that TKA would be cost-effective in 100% of simulations of patients with a BMI below 50 and in 99.16% of simulations of super-obese (BMI of 50 and over) patients.⁴⁵

Grading the quality of CUA studies

Many CUA studies are produced using vastly different methodologies, reporting and data sources. Therefore, it is important for an objective grading system to exist to assess the quality of each of these studies with proven high construct and content validity. These assessment instruments will assist the identification of CUA studies of superior merit.

In addition, a greater standard of reporting will inevitably be encouraged, leading to greater quality and rigour. Four commonly used instruments are discussed below.

The Quality of Health Economic Studies (QHES) instrument is one such method. The QHES highlights the importance of suitable methods, valid, transparent results and their comprehensive reporting in each CUA study. There are 16 independent criteria used, with each having a weighted point value. The more important criteria are considered to be of greater relative weight. The perfect quality score for a study is 100, which is calculated by the sum of all the points of all the questions answered ‘yes’.⁴⁶

The Consensus Health Economic Criteria (CHEC) checklist focuses only on the methodological quality of economic evaluations. It is only suitable for systematic reviews based on full economic evaluations based on clinical trials (cohort studies, randomized controlled trials and case-control studies) that compare costs and health outcomes of two or more alternative options (e.g. management options). Both costs and health outcomes of each of the alternatives must be examined. The CHEC cannot be used for model-based study designs due to other methodological criteria being relevant which are absent in this checklist. The checklist consists of 19 binary yes-or-no questions. Each question pertains to a single category. A, ‘no’, response should be given if insufficient information was available in the article or in other relevant published material.²⁷

The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist was developed to

facilitate consistent and transparent reporting of economic evaluation research. Importantly, it reflects the most up-to-date and widely accepted standards in this field. It was initially designed as a guide to best practice reporting for such research types. However, several researchers have used it as an appraisal tool for risk of bias in these studies, but it should be acknowledged that it is not intended to assess methodological quality. It consists of a 24-item checklist with accompanying recommendations on the minimum amount of information required when reporting such economic evaluations.⁴⁷

It should be noted, that although these instruments were developed to promote a greater standard of reporting in economic evaluations over time, this has not always been observed within orthopaedics. For example, as demonstrated by Rajan et al, there has been a growing number of lower-quality orthopaedic-based CUA studies (assessed using the QHES) relating to the management options of the upper limb in recent years.⁴⁸

Common limitations in cost-utility analysis studies

Although CUA studies offer valuable information for healthcare decision-makers, they are vulnerable to a variety of potential limitations that may harm their overall credibility. The main limitations, as outlined by the European Network for Health Technology Assessment, are described in Table 2.⁴⁹

Table 2. The main sources of limitations associated with CUA studies (adapted from the European Network for Health Technology Assessment)⁴⁹

Source of limitation	Description
Efficacy/effectiveness and safety of the management option	<ul style="list-style-type: none"> All evidence should be taken into account, both published and unpublished, to avoid publication and reporting bias of these metrics. A balanced assessment of all clinical evidence must be performed as this provides the input for the economic evaluation. The impact of adverse events on costs and health outcomes should be taken into account.
Comparator	<ul style="list-style-type: none"> Ideally, the comparator should be the reference treatment according to the most up-to-date high-quality clinical practice guidelines at European or international level with strong literature evidence on efficacy and safety, and with recognized regulatory approval for the respective clinical indication. Therefore, readers must be aware of the inclusion of inappropriate, non-cost-effective comparators in place of more relevant (and possibly more cost-effective) ones to calculate the ICER values.
Subgroup analysis	<ul style="list-style-type: none"> Measurements of cost-effectiveness for an overall study population may lead to incorrect management option recommendations as cost-effectiveness may differ between subgroups. For example, the absolute treatment efficacy of a particular management option may differ based on sociodemographic (e.g. age, sex, socioeconomic class etc.) or clinical (e.g. baseline risk, disease severity etc.) characteristics. Often, subgroup analyses are simply exploratory and should be interpreted with caution as the subgroup sizes are often too small to detect moderate differences.
Baseline risk of the target population	<ul style="list-style-type: none"> There may be differences in the baseline risk for certain events in a specific population (e.g. that selected for an RCT) versus the general population to which the decisions of policy-makers apply. A failure to adjust for these differences may have a significant impact on the modelled absolute health outcomes and costs.
Compliance	<ul style="list-style-type: none"> In most cost-effectiveness evaluations, compliance to the particular management option is not explicitly considered but is likely to have an impact on cost-effectiveness. If compliance is not the same in the underlying trial population compared to the target population, then the health outcomes and costs may be biased. Ideally, the impact of poor adherence and the overall adherence in both study populations should be quantified using the appropriate evidence.
Quality of life	<ul style="list-style-type: none"> Readers should be cautious when non-evidence-based utility weights (e.g. based on expert opinion) are used because a generic utility instrument was not used in the underlying trials. If different direct or indirect methods were used to evaluate the health utilities of the management options under comparison, there is increased uncertainty as different values may be achieved with each.

Downloaded from Bioscientifica.com (continued) 1:44:14PM via Open Access. This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed. <https://creativecommons.org/licenses/by-nc/4.0/>

Table 2 (continued)

Source of limitation	Description
Time horizon and extrapolation	<ul style="list-style-type: none"> As a modelled time horizon increases and/or there is more extrapolation, there is a greater associated inherent uncertainty. When extrapolating data beyond the duration of the clinical trial, it is important for the underlying assumptions and data sources to be stated. The impact of different extrapolation scenarios should be assessed via sensitivity analysis to adequately capture the uncertainty present.
Discount rate	<ul style="list-style-type: none"> Discount rates may have a significant impact on the primary outcome measures of health utility and costs, especially in long-term models. Therefore, the different discount rates used by economic evaluations may hinder comparisons between them.
Perspective	<ul style="list-style-type: none"> Omitting relevant costs or incorrectly including irrelevant ones may introduce bias of unknown direction. The choice of perspective used may significantly influence the calculated ICERs and thus the cost-effectiveness judgement made. Therefore, it is advisable to present results separately for different perspectives.
Context-specific costs	<ul style="list-style-type: none"> Ideally, prices and resource used for specific cost items should be summarized in a prices * quantities table to provide information for critical assessment of results. The healthcare financing system of a country in which an economic evaluation is performed needs to be considered when gathering cost information as it may be different from other countries. Patient costs are often skewed towards those carrying lower costs. The assumed parametric statistical distributions of common applied statistical tests may not reflect this and so leading to incorrect confidence intervals and <i>p</i>-values.
Sensitivity analysis	<ul style="list-style-type: none"> Ideally, confidence intervals should be presented for key parameters and with the upper and lower bounds being linked to the best available evidence, be plausible and adequately reflect uncertainty. The sensitivity analyses should comprehensively identify all parameters and assumptions that contribute to uncertainty in the model outputs. The statistical distributions for parameters should not enable implausible values (e.g. negative costs). The model outputs reported should adequately reflect their uncertainty (e.g. via confidence intervals) and key deficiencies in available data and assumptions discussed.
Model verification and validation	<ul style="list-style-type: none"> Model verification asks whether the model has implemented the assumptions correctly and model validation asks if they are reasonable and reflect reality. The question of whether the results are consistent with those from other studies and whether differences can be explained is also important.
Transferability of economic evaluation results	<ul style="list-style-type: none"> It is important to consider the transferability of study findings from an economic evaluation performed in one specific decision-making context to another. There are now checklists that help investigators to identify the parameters that are more vulnerable to differ between settings which may have implications for the ICER values and so cost-effectiveness.
ICER threshold	<ul style="list-style-type: none"> Authors may potentially use WTP thresholds that are relatively high and are not accepted in their country at that given time. Therefore, it is crucial to justify the reasons for the specific WTP chosen. Conclusions about cost-effectiveness in countries where there is no explicit WTP cannot be made as they may lead to incorrect conclusions.
Publication bias of economic evaluations and conflicts of interest	<ul style="list-style-type: none"> Industry-sponsored cost-effectiveness studies are more likely to report favourable conclusions than those with other funding sources, which may imply a source of publication bias.⁵⁰ There may be a tendency for favourable input variables to be chosen for the sponsor's product.

Notes. CUA, cost-utility analysis; ICER, incremental cost-effectiveness ratio; RCT, randomized controlled trial; WTP, willingness-to-pay.

Concluding remarks

In summary, cost-utility analysis studies are becoming increasingly important within trauma and orthopaedic surgery. They offer a powerful tool to elucidate valuable information that will help inform clinical decision-making which will ultimately lead to the optimization of routine clinical practice via the incorporation of the most cost-effective methods. This review has outlined the main concepts required to understand the cost-utility literature and provides a basic framework for their future conduct.

AUTHOR INFORMATION

¹School of Clinical Medicine, University of Cambridge, Cambridge, UK.

²Department of Surgery, University of Cambridge, Cambridge, UK.

Correspondence should be sent to: K. T. Matthew Seah, Department of Surgery, University of Cambridge, Box 180, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, UK.
Email: ms2340@cam.ac.uk

ICMJE CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest relevant to this work.

FUNDING STATEMENT

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

OPEN ACCESS

© 2021 The author(s)

This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

REFERENCES

- Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996;276:1253–1258.
- Sassi F. Calculating QALYs, comparing QALY and DALY calculations. *Health Policy Plan* 2006;21:402–408.

3. **Rai M, Goyal R.** Chapter 33: pharmacoeconomics in healthcare. In: Vohora D, Singh GBT-PM and TCR, eds. Boston, MA: DAcademic Press, 2018:465–472. <http://www.sciencedirect.com/science/article/pii/B9780128021033000341> (Date last accessed 10 November 2020).
4. **McDougall JA, Furnback WE, Wang BCM, Mahlich J.** Understanding the global measurement of willingness to pay in health. *J Mark Access Health Policy* 2020;8:1717030.
5. **National Institute for Health and Care Excellence.** *Guide to the methods of technology appraisal 2013* <http://nice.org.uk/process/pmg9> (Date last accessed 15 November 2020). Process and Methods Guides 9. London: National Institute for Health and Care Excellence (NICE), 2013.
6. **O'Mahony JF, Coughlan D.** The Irish cost-effectiveness threshold: does it support rational rationing or might it lead to unintended harm to Ireland's health system? *Pharmacoeconomics* 2016;34:5–11.
7. **Franken M, Koopmanschap M, Steenhoek A.** Health economic evaluations in reimbursement decision making in the Netherlands: time to take it seriously? *Z Evid Fortbild Qual Gesundheitswes* 2014;108:383–389.
8. **Vallejo-Torres L, Garcia B, Serrano-Aguilar P.** Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Econ* 2018 Apr;27(4):746–761
9. **Padula W, Chen HB, Phelps C.** PRM39: is the choice of willingness-to-pay threshold in cost-utility analysis endogenous to the resulting value of the technology? *Value Health* 2018;21:S362.
10. **Cleemput I, Neyt M, Thiry N, De Laet C, Leys M.** Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care* 2011;27:71–76.
11. **Mota REM.** Cost-effectiveness analysis of early versus late total hip replacement in Italy. *Value Health* 2013;16:267–279.
12. **York Health Economics Consortium.** Standard Gamble, 2016. <https://yh.ec.co.uk/glossary/standard-gamble/> (Date last accessed 15 November 2020).
13. **Marsh JD, Birmingham TB, Giffin JR, et al.** Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. *BMJ Open* 2016;6:e009949.
14. **Bird SB, Dickson EW.** Clinically significant changes in pain along the visual analog scale. *Ann Emerg Med* 2001;38:639–643.
15. **Wang D, Xie W, Cao W, He S, Fan G, Zhang H.** A cost-utility analysis of percutaneous endoscopic lumbar discectomy for L5-S1 lumbar disc herniation: transforaminal versus Interlaminar. *Spine* 2019;44:563–570.
16. **Whynes DK, TOMBOLA Group.** Correspondence between EQ-5D health state classifications and EQ VAS scores. *Health Qual Life Outcomes* 2008;6:94.
17. **Lins L, Carvalho FM.** SF-36 total score as a single measure of health-related quality of life: scoping review. *SAGE Open Med* 2016;4:2050312116671725.
18. **Saris-Baglana RN, Dewey CJ, Chisholm GB, et al.** *QualityMetric health outcomes™ scoring software 4.0*. Lincoln, RI: QualityMetric Incorporated, 2010:138.
19. **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;15:1833–1840.
20. **Lequesne MG, Mery C, Samson M, Gerard P.** Indexes of severity for osteoarthritis of the hip and knee. Validation—value in comparison with other assessment tests. *Scand J Rheumatol Suppl* 1987;65:85–89.
21. **Rat AC, Coste J, Pouchot J, et al.** OAKHQOL: a new instrument to measure quality of life in knee and hip osteoarthritis. *J Clin Epidemiol* 2005;58:47–55.
22. **Rat A-C, Guillemin F, Pouchot J.** Mapping the osteoarthritis knee and hip quality of life (OAKHQOL) instrument to the international classification of functioning, disability and health and comparison to five health status instruments used in osteoarthritis. *Rheumatology (Oxford)* 2008;47:1719–1725.
23. **Cieza A, Ewert T, Ustün TB, Chatterji S, Kostanjsek N, Stucki G.** Development of ICF Core Sets for patients with chronic conditions. *J Rehabil Med* 2004;44:9–11.
24. **Weeks CA, Marsh JD, MacDonald SJ, Graves S, Vasarhelyi EM.** Patellar resurfacing in total knee arthroplasty: a cost-effectiveness analysis. *J Arthroplasty* 2018;33:3412–3415.
25. **Nagata T, Mori K, Ohtani M, et al.** Total health-related costs due to absenteeism, presenteeism, and medical and pharmaceutical expenses in Japanese employers. *J Occup Environ Med* 2018;60:e273–e280.
26. **European Network for Health Technology Assessment (EUNEHTA).** *Guidance document. Methods for health economic evaluations: a guideline based on current practices in Europe, May 2015.* <https://eunetha.eu/methodology-guidelines/> (Date last accessed 10 November 2020) EUNEHTA.
27. **Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA.** Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. *J Med Internet Res* 2019;21:e11330.
28. **Eurostat Database.** Your key to European statistics. <https://ec.europa.eu/eurostat/web/hicp/data/database> (Date last accessed 10 November 2020).
29. **Turner HC, Lauer JA, Tran BX, Teerawattananon Y, Jit M.** Adjusting for inflation and currency changes within health economic studies. *Value Health* 2019;22:1026–1032.
30. **Jit M, Mibei W.** Discounting in the evaluation of the cost-effectiveness of a vaccination programme: a critical review. *Vaccine* 2015;33:3788–3794.
31. **Severens JL, Milne RJ.** Discounting health outcomes in economic evaluation: the ongoing debate. *Value Health* 2004;7:397–401.
32. **Nice.org.uk.** The reference case guide to the methods of technology appraisal 2013: guidance. <https://www.nice.org.uk/process/pmg9/chapter/the-reference-case> (date last accessed 3 July 2020).
33. **van Lier LI, Bosmans JE, van Hout HPJ, Mookink LB, van den, Hout WB, de Wit GA, et al.** Consensus-based cross-European recommendations for the identification, measurement and valuation of costs in health economic evaluations: a European Delphi study. *Eur J Health Econ* 2018;19:993–1008.
34. **Miyazaki S, Kakutani K, Sakai Y, et al.** Quality of life and cost-utility of surgical treatment for patients with spinal metastases: prospective cohort study. *Int Orthop* 2017;41:1265–1271.
35. **York Health Economics Consortium.** Time horizon, 2016. <https://www.yhec.co.uk/glossary/time-horizon/> (date last accessed 1 July 2019).
36. **Kamiński B, Jakubczyk M, Szufel P.** A framework for sensitivity analysis of decision trees. *Cent Eur J Oper Res* 2018;26:135–159.
37. **Karnon J, Brown J.** Selecting a decision model for economic evaluation: a case study and review. *Health Care Manage Sci* 1998;1:133–140.
38. **Zindel S, Stock S, Müller D, Stollenwerk B.** A multi-perspective cost-effectiveness analysis comparing rivaroxaban with enoxaparin sodium for thromboprophylaxis after total hip and knee replacement in the German healthcare setting. *BMC Health Serv Res* 2012;12:192.

- 39. Pennington M, Grieve R, Black N, van der Meulen JH.** Cost-effectiveness of five commonly used prosthesis brands for total knee replacement in the UK: a study using the NJR Dataset. *PLoS One* 2016;11:e0150074.
- 40. Zuchelli E, Jones A, Rice N.** The evaluation of health policies through microsimulation methods. *Heal Econom Data Gr Work Pap* 2010. (HEDG) Working Papers 10/03, HEDG, c/o Department of Economics, University of York.
- 41. Si L, Eisman JA, Winzenberg T, et al.** Microsimulation model for the health economic evaluation of osteoporosis interventions: study protocol. *BMJ Open* 2019;9:e028365.
- 42. Brauer CA, Manns BJ, Ko M, Donaldson C, Buckley R.** An economic evaluation of operative compared with nonoperative management of displaced intra-articular calcaneal fractures. *J Bone Joint Surg Am* 2005;87:2741–2749.
- 43. Taylor M.** What is sensitivity analysis? *York Health Economics*, 2009. http://www.bandolier.org.uk/painres/download/whatis/What_is_sens_analy.pdf (date last accessed 1 July 2019).
- 44. Briggs AH, Gray AM.** Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;3:1–134.
- 45. Ponnusamy KE, Vasarhelyi EM, Somerville L, McCalden RW, Marsh JD.** Cost-effectiveness of total knee arthroplasty vs nonoperative management in normal, overweight, obese, severely obese, morbidly obese, and super-obese patients: a Markov model. *J Arthroplasty* 2018;33:S32–S38.
- 46. Chiou C-F, Hay JW, Wallace JF, et al.** Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003;41:32–44.
- 47. Husereau D, Drummond M, Petrou S, et al; CHEERS Task Force.** Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ* 2013;346:f1049.
- 48. Rajan PV, Qudsi RA, Dyer GSM, Losina E.** Cost-utility studies in upper limb orthopaedic surgery: a systematic review of published literature. *Bone Joint J* 2018;100-B:1416–1423.
- 49. European Network for Health Technology Assessment (EUNEHTA).** *Guidance document. Practical considerations when critically assessing economic evaluations.* Version 1.0, 09 March 2020. <https://eunetha.eu/methodology-guidelines/> (Date last accessed 15 October 2020) EUNEHTA.
- 50. Garattini L, Koleva D, Casadei G.** Modeling in pharmacoeconomic studies: funding sources and outcomes. *Int J Technol Assess Health Care* 2010;26:330–333.