

University of Southern Denmark · Syddansk Universitet

Methods for cost-effectiveness evaluation alongside trials in spine surgery

Rikke Søgaard, Mirja Elisabeth Kløjgaard and Jens Olsen

Health Economics Papers

2010:5



UNIVERSITY OF SOUTHERN DENMARK

Editors preface

Back disorders represent a substantial economic and social burden to society. New technologies are steadily evolving and for some of these clinical outcome and cost-effectiveness remain uncertain. Research based on clinical evidence and taking into account the economic constraints will help create a basis for decision making on a strategic, political level.

This working paper enables clinical researchers to perform cost-effectiveness analysis, on a suitable methodological basis, alongside clinical trials in the field of spine surgery. Aiming at guiding researchers in their choice of methods, the paper describes how to collect costs and effects while performing a controlled clinical trial and how to combine the data into a cost-effectiveness evaluation. The paper concludes with some recommendations for the reporting of economic evaluations, thereby contributing to uniformity in reporting standards if these are followed. This would be of great value to decision-makers that are to use the results.

Odense, December 2010,
Mickael Bech

Authors preface

The rationale for this paper developed with the launch of a larger research project: “Cost-effectiveness of New Treatment Strategies in Spine Surgery: Evaluation of Patient-based Outcome, Health and Social Economic Consequences in Denmark”, financed by the Danish Strategic Research Council (grant 2142-08-0017). The project includes a series of planned cost-effectiveness evaluations to be conducted by clinical research fellows without formal backgrounds in economics or a related field. Thus the objective of the paper is to guide methodological choices for conducting cost-effectiveness evaluation of technologies related to spine surgery, thereby facilitating a methodological consistency of project outputs.

It should be underlined that this paper is neither exhaustive nor can it stand alone as guidance of how to conduct economic evaluation; rather should it be seen as a paper addressing special issues related to the context of spine surgery, which supplements common textbooks of e.g. Drummond et al.’s *Methods for the Economic Evaluation of Health Care Programmes* or Gold et al.’s *Cost-effectiveness in Health and Medicine*.

Although targeted a specific context, this paper might interest a broader group of researchers in chronic pain or researchers evaluating surgical regimens in general. Its applicability is however based on the availability of register data at the level of the individual patient and, it is restricted in so called stochastic evaluation (as opposed to modelling studies) not necessarily because is the most appropriate approach but, as modelling studies were not scheduled in this specific context.

We are grateful to Professor Terkel Christiansen, Health Economics Unit, Institute of Public Health, who has reviewed the paper in an earlier version and provided constructive comments.

Odense, December 2010

The authors

Contents

1	Introduction.....	5
1.1	Epidemiology of low back pain	5
1.2	Costs of back pain	5
1.3	The rationale for prioritisation.....	6
2	The evaluation problem	7
2.1	Market failures.....	7
2.2	Key tenets of economic theory	8
2.3	Economic evaluation	8
3	Costing.....	11
3.1	Identification, measurement and valuation	11
3.1.1	Intervention costs	14
3.1.2	Costs in the hospital sector	16
3.1.3	Costs in the primary health care sector	16
3.1.4	Costs in other sectors.....	17
3.1.5	Costs for patient and family.....	17
3.1.6	Production loss/gain	18
3.1.7	Future costs.....	19
3.2	Discounting and price level.....	20
3.3	Budget impact or fiscal analysis	20
4	Measuring and valuing outcomes.....	22
4.1	Generic, preference-based instruments	22
4.2	The choice between the SF-6D and the EQ-5D.....	24
5	Reporting cost-effectiveness evaluations	27
5.1	Base-case analysis	27
5.2	Parameter uncertainties: bootstrapping	28
5.3	Structural uncertainties: sensitivity analysis.....	31
5.4	Recapitulation	31
6	References.....	33

1 Introduction

1.1 Epidemiology of low back pain

Almost every citizen in the Western part of the world experiences low back pain (LBP) at some point in their life but 80 to 90% recover within 3 months and without residual functional loss (1). In contrast, recovery after this timeframe is slow and uncertain; less than 50% of patients with a history of more than 6 months of LBP return to work and almost none with a history of more than 12 months of LBP return to work (2). There is no agreement on a consistent clinical definition of LBP yet most would probably accept a broad definition of localized, diffuse, or referred pain in the region below the costal margin and above the inferior gluteal folds, with or without radicular pain.

The epidemiology of chronic LBP is less informed, possibly because of a complex aetiology and a lack of consensus about the definition. Often, LBP is said to be chronic when persisting for more than three months but some authors prefer a dichotomy of disabling versus non-disabling pain. Using the former, a Danish study estimated the prevalence of chronic back pain at 10% in the adult population (3). A North American study estimated the prevalence at a similar level and while commenting that the current prevalence represents the result of a factor 2.5 increase since 1992 (4).

The consequences of chronic pain to the individual patient are extensive and often manifest in significantly reduced quality of life. Absolute values of health-related quality of life (measured using the EuroQol group's EQ-5D) have been estimated between 0.23 in surgical candidates with degenerative disorders (5) through 0.41 in patients with LBP for at least one year (6). Such levels are in the same region as e.g. those of patients with life-threatening cancer (7).

1.2 Costs of back pain

Several studies have reported the costs of back pain to society (1;8-11). Although most studies came about in the late 1990s and many structural and technological advances have developed since, it should be noted that the societal impact of back pain has been estimated at as much as 1.5% of gross domestic products(12). Such impact corresponds to that of other major diseases as depression, heart disease or diabetes and is to a wide extent attributable to the small proportion of patients who manifest chronic pain. Patients with chronic pain have been found to account for 65 to 85% of the total costs associated with back pain (13).

The reason for a majority of costs being attributable to chronic disease is multifaceted. Not only does production loss of chronic patients accrue over a long time, their service utilisation of health care also accumulates and often leads to an endpoint event of spinal surgery. This is currently a somewhat controversial issue as it has been argued that the rise in costs of spinal surgery over the past two decades cannot be associated solely with a rising need. Deyo and colleagues have argued that the introduction of the cage in 1996 can be shown to coincide with a major upwards shift in the frequencies and total costs of surgical procedures (14). In addition to the factual increase in the rate of surgery of 113% from 1996 through 2001, the authors noted that such trend was much greater than seen in other major orthopaedic procedures.

1.3 The rationale for prioritisation

Altogether, the complexity of the disease, the life-time consequences to the individual and to society if chronic pain manifests, as well as the rise in the number of spine surgeries, automatically attract political focus. This is even further incentivised by the broader health political challenge of costs exceeding budgets in most publicly funded health care systems, thus making prioritisation a necessity. Economic evaluation is a rational approach for informing such decisions about what technologies or interventions should be prioritised over others to maximize the value for money in health care (and elsewhere).

This paper will proceed with a brief introduction to the evaluation problem to illustrate core principles of economic evaluation and to pinpoint a few caveats that might require attention in the specific context of spine surgery. The following two sections will then address in more detail some methodological issues relating to the measurement and valuation of resource use and consequences, respectively. A final section will touch on the issue of reporting results of cost-effectiveness evaluations.

2 The evaluation problem

2.1 Market failures

Classical economic theory is based on the idea that a free (competitive) market will automatically lead to efficient allocation of resources given some key conditions. The economist Adam Smith who used the metaphor of “the invisible hand” of the market introduced this already in 1759.

In health care however there is a number of premises that undermine the conditions for which the market will automatically lead to an optimal solution. These can be overall referred to under the headings of three market failures: uncertainty, information asymmetry, and externalities (15). Health and consumption of health care is associated with much uncertainty in that individuals generally do not know whether they become ill and when ill, they generally do not know the exact outcome. This often leads risk-averse individuals to take up insurance, given the incalculable consequences if they become ill. However, as individuals often have private information about their health, they might seek to opt in for insurance programmes that are based on people in lower risk and therefore have lower premiums. This is a market failure referred to as adverse selection. A related failure arises when individuals after they have taken up insurance change their behaviour in a way that increases their risk for falling ill. This is referred to as moral hazard.

Asymmetric information is another aspect, referring to either the buyer or the seller being more informed than the counterpart. This is a characteristic describing many levels of a health care market, e.g. patient versus doctor, payer versus provider, or insurance provider versus insurance taker. Information asymmetry most often leads the under-informed to employ a third party (an agent) to handle the interests of the under-informed party (now becoming a principal) in a belief that such constellation will maximise the principal’s utility more than had he or she handled own interests him or herself. If the agent is acting in concordance with the preferences of the patient the principal-agent relationship is perfect. If however this is not the case, e.g. due to the agent having conflicting interests, the agent-principal relationship is imperfect and considered a market failure. Needless to say, professionalism of the health professionals in most cases prevent such scenario.

The third issue of externalities refers to demand curves of individuals not necessarily summing to a societal demand curve. Externalities can be defined by the fact that a purchase decision affects other individuals than the buyer (think about passive smoking, vaccine for contagious disease etc.). Given that the individual will go for a solution that maximises his or her utility and not any other parties’ utility (overall or societal utility), the market solution will not be optimal. A common type of externalities arises from individuals’ altruistic preferences, i.e. that people care for other peoples’ health. This means that the value of an intervention, say, may be higher than that of the individual receiving it. If this is not taken into account the market will fail to lead to optimum (and instead lead to an under-supply relative to what is optimal).

The presence of just one of these market failures will obstruct market forces to lead to an optimal solution and thus the market will no longer ensure that new or running technologies, for example various types of surgical management for chronic low back pain, are worthwhile. To support decisions about what

technologies should be recommended from a rational economic standpoint, information about whether expected benefits outweigh expected costs becomes essential. This brings in the approach of economic evaluation, which represents a bunch of tools for informing rational decisions about allocation of scarce resources.

2.2 Key tenets of economic theory

There are three tenets of economic theory, forming the rationale for economic evaluation, which will be outlined in this section. First, resources are scarce and can only be used once (at least simultaneously). For example, there is a certain limit to how many surgeons can be educated and trained to serve a health service system, both due to restrictions on the uptake of medical students as well as individual preferences in that not all individuals prefer a medical career. Second, resources have alternative uses and thus allocating a resource for one purpose hinders (simultaneous) allocation for other purposes. This fact is in economic theory referred to as the opportunity cost, which is more explicitly defined as the value of a resource in its best alternative use. This means that while the opportunity cost and the financial cost is often the same, this is not always the case. For example, if a surgeon chooses to work one hour for free, the financial cost would be zero whereas the opportunity cost would be the value of his or her time in the best alternative use, that being spending leisure time, working in another clinic or whatever would be the second-best choice for the surgeon. The important understanding from the two tenets is that since resources are finite and since they cannot be used twice, any decision to allocate a resource for a certain use has a cost, which should be understood as an opportunity cost. When costs are referred to in the remainder of this paper, the interpretations should always be opportunity costs.

Estimating the cost of different choices does not provide a foundation for a decision, as alternative choices are typically associated with different benefits. Therefore to inform which choice represents the optimal allocation of resources we need a measure of benefit. Who to ask and how to operationalise benefit has been subject to much debate over time, but it is well established that what we want to capture is the individual's utility from acquiring or consuming a good. Accordingly, the notion of utility is a key concept that can be defined as the benefit, satisfaction or pleasure an individual gets from acquiring and/or consuming a good. The exact amount of utility acquired by one individual may differ from the amount acquired by another individual, even from the consumption of identical goods, because utility is determined by the degree to which individual preferences are fulfilled. For example, it is not uncommon that two patients with identical pathology prefer different treatment alternatives due to different family- or job situations, impact of functional restrictions on leisure activities that arise from the surgery etc. The only judge of the extent to which individual preferences are fulfilled is the individual him- or herself, i.e. economic theory is based on the sovereignty of the individual in judging utility. The introduction of the notion of utility facilitates the third tenet of economic theory, namely that the optimal allocation is represented by the choice where utility is maximised (subject to scarce resources).

2.3 Economic evaluation

Economic evaluation in health care has evolved over the past 30 years. It can be defined as a systematic assessment of both costs and benefits of two (or more) alternatives and thus it provides a rational foundation for prioritisation of scarce resources. Evaluations can take many forms; from quick assessments, within a narrow perspective of for example a hospital department, through more sophisticated models of

lifetime consequences for the society as a whole. This section introduces some basic principles whereas a more comprehensive representation is referred to textbooks (see e.g. Fox-Rushby et al. (16) for an introductory text or Drummond et al. (17) and (18) for intermediate-level texts).

It should be realised already at this point that the framework for economic evaluation can inform decision-makers about what choices maximises utility (given the resources available) but, since far from all decisions are made under the objective of rationality, it has important limitations. The objective of most national health services is multifaceted and includes issues of maximising population health, ensuring equity in access to services, ensuring high standards of quality, limiting the problem of waiting lists, cost containment etc. Some of these facets might even conflict, although this is an area outside the scope of the present paper. In practice therefore, economic evaluation can inform decisions, which are then made upon all information available and not just the guidance of rational behaviour.

Planning economic evaluation alongside a randomised controlled trial can be seen as just another way of outcome measurement. It is most validly done prospectively, using validated instruments and following a detailed protocol ensuring consistency throughout the course of the trial. Figure 1 illustrates a conventional flow diagram of a randomised controlled trial with the grey boxes indicating what information should be collected to facilitate estimation of the so called incremental cost-effectiveness ratio (ICER – further explained in last chapter). The quantification of costs on one side and effects on the other side is addressed in detail in the two following chapters of this paper.

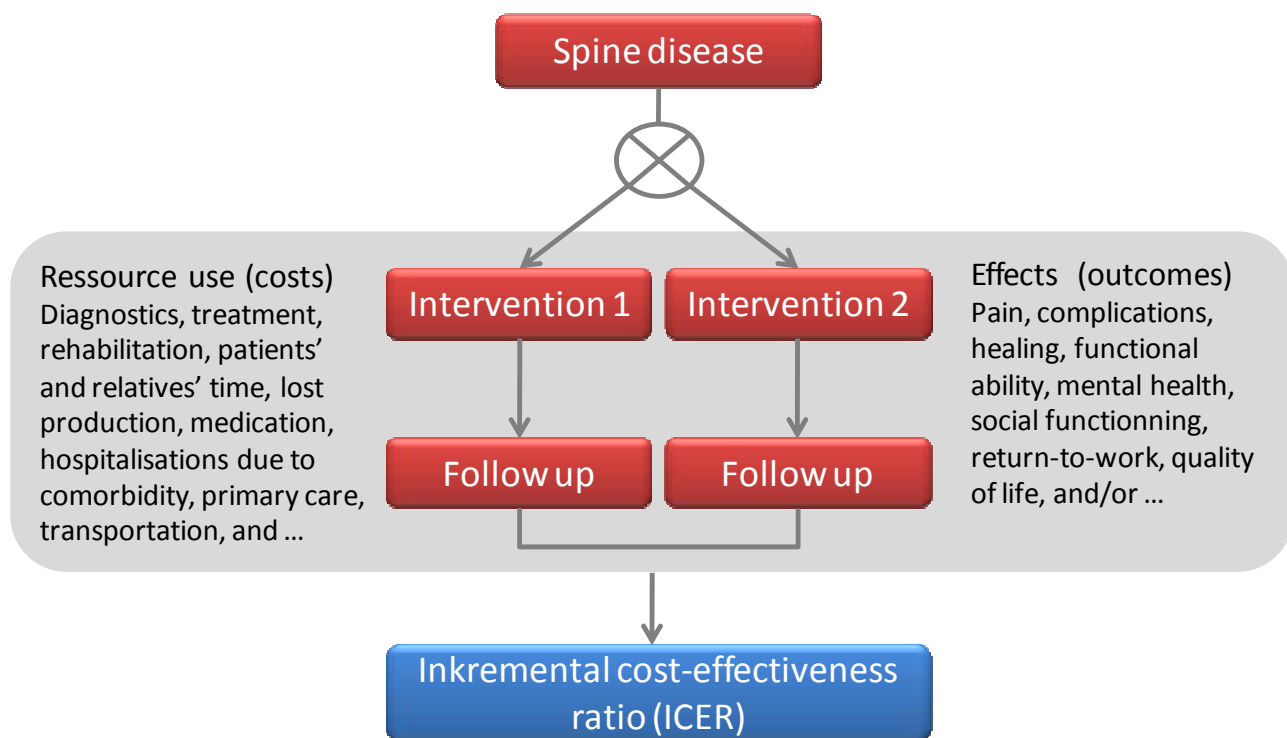


Figure 1 Flow diagram of a conventional clinical trial with the appendix of parameters for economic evaluation: resource use and effects.

The ICER summarises the result of an economic evaluation in one parameter and is defined by the ratio of additional costs per additional unit of effect – often simply referred to as a cost per effect unit gained or a cost per quality-adjusted life-year (QALY) gained.

For the ICER to guide decisions, a threshold value for the maximum willingness-to-pay per unit of effect is required i.e. a decision rule indicating when to accept or reject interventions. However, few authorities have made their thresholds explicit (if they have any) and there is no general consensus on how to elicit them (19). One of the most proactive parties on this issue is the National Institute for Health and Clinical Excellence (NICE) in the UK where a threshold value for a QALY gain between £20,000 and £30,000 has been used more or less explicitly for the past decade or so with some alterations concerning e.g. expensive cancer drugs or highly specialised treatment for rare life-threatening diseases (20;21). In general, the maximum willingness-to-pay per unit of effect varies depending on different factors (e.g. disease severity, number of patients affected, budget impact).

For the results of an economic evaluation to hold relevance to national decision-making it is imperative to employ a national perspective i.e. to include all sectors of society that might be affected by a decision. It has often been referred that only 25 % of costs of low back pain is due to direct treatment costs whereas the remainder costs are production losses or patients' time and transportation costs (22). Limiting an analysis to the health care sector thus could lead to severely biased guidance for national decision-making. On the other hand it might be justified to apply a more narrow perspective if the decision problem is regional or at the hospital level. The important message is that the perspective of analysis should always be judged according to the use of results.

The analytical time horizon is another critical issue for the validity of guidance. Unless an economic evaluation captures all differences between comparators, i.e. not only from a sufficiently broad perspective but also during a sufficiently long period of follow up, it will possibly be biased. This sometimes means that patients should be followed for lifetime. Specific to spine surgery it has been suggested that there could be a natural order of when different dimensions of outcomes manifest with a first category being the biological outcome (for example improvement on pain scales), a next category being the psychological outcome (for example improvement on anxiety and depression scales), and a last category being a social outcome (for example return to work or daily activities) (23). When economic evaluation piggybacks on clinical trials, as is often the case in practice, the length of follow up must often coincide with the length of follow up of the clinical trial. This is a pragmatic solution, which is not necessarily optimal but generally regarded as better than not evaluating at all.

3 Costing

This section describes the different types of costs and how to identify, measure and value them in the frame of a cost-effectiveness analysis. For a more extensive description of costing in health economic analyses see for example Drummond et al. (17) and the Danish handbook for conducting health technology assessments (24).

A cost is defined as the value of a resource use with the most basic resources being human resources and materials/utensils. When resources are used in the treatment of a given patient at a given time, the same resources are not available for other patients or other purposes and therefore all resource use should be valued by its opportunity cost (this notion was introduced in the section about key tenets of economic theory).

When estimating the costs of a treatment alternative, one has to consider the whole pathway of a patient through the health care system as well as consequences in other sectors of society. For a patient in need of spine surgery the costs of the treatment actually starts from the moment the disease emerges and lead the patient to begin a path in the health care sector. From there on the patient may undergo diagnostics, hospital admission including a surgical procedure, rehabilitation and follow-up. Given that hospital services are typically denoted secondary health care, there will often be derived effects in the primary health care sector, for example, fewer or more visits to the general practitioner, chiropractors or practising physiotherapists. In a broader societal perspective costs might occur due to production losses arising from patients' absenteeism or presenteeism (on-the-job productivity loss). To inform the full costing picture one has to gather all these types of costs for every treatment alternative included in the analysis.

3.1 Identification, measurement and valuation

The first step in assessing the cost is identification of the resources used in the alternatives (typically an experimental and a control trial arm) that are compared. Strictly speaking, it is only resource use expected to vary between the alternatives that needs to be collected. But in order not to restrain comparison of results to results of future evaluations, the optimal approach is to include all costs for the provision of both a total intervention cost (per patient) as well as the extra cost (per patient) associated with the experimental arm.

Obviously the perspective of the analysis determines its complexity but also the usability of results. Alternative perspectives range from the societal (the most extensive perspective) to more narrow hospital sector perspectives, including only the costs of in- and outpatient services. Guidelines recommend that analyses should be performed from a societal perspective (17;25) and in any way it should be realized that the choice of perspective is crucially important since an intervention that turns out cost-effective from a hospital perspective may be less or even not cost-effective from a societal viewpoint and vice versa. The perspective should always be stated explicit. Table 1 gives more information about different perspectives.

Table 1 Perspective of analysis and types of costs/resource use.

Perspective			Type of costs	Resource use
Society	Health care sector	Hospital	Costs in the hospital	Health personnel, medicine, utensils, tests, capital equipment (plant & buildings), in-patient stay (hotel), outpatient visits, overheads (food, lighting, heat, etc.), research & training
			Costs in the primary health care sector	Consultation with general practitioner, practicing specialist, physiotherapist, etc., prescription medicine (the Danish National Health Insurance Service's share), public surveys
			Costs in other sectors	Home care & home nursing, social events, including support for medicine (municipal grants), aids
		Costs for patient & family	User payment (medicine, dentist), transport, time spent on investigation/treatment, (unpaid) time spent by family or friends in caring for patients	
		Production loss/gain in society	Changes in patients' temporary absence through sickness, reduced ability to work due to sickness and disability, or lost production in the case of premature death	
		Future costs	Future unrelated costs including health costs generated as a result of a patient's lifetime being extended or shortened	

Source: Own modification of Table 9.2 in (24).

Apart from considerations about the perspective of analysis, a choice about the analytical time period also has to be made, that is, for how long should resource use be tracked? In principle, patients should be followed for as long as resource use vary between the randomization groups and this may involve a lifetime perspective (17). However in cost-effectiveness analyses conducted alongside single trials the time horizon will often be the same (for pragmatic reasons) as that of the clinical trial which is often limited to one or two years. A two-year time horizon involves a risk that relevant costs are ignored and could lead to a biased

conclusion. Costs that arise beyond the two years could be additional treatment of e.g. removal of implants, reoperations, long-term complications or even productivity changes. Such effects should always be discussed in a limitations section when reporting a cost-effectiveness analysis from a limited time horizon.

The second step in costing is measurement of the identified resource use. There are several methods and sources that can be used to collect resource use, depending on how much time the researcher has and/or the richness of secondary data already available. In principle, a distinction between prospective and retrospective collection of data is made. This is addressed in more detail under specific cost categories in the following sections.

The final step refers to choosing an appropriate unit cost to multiply with the measured quantity to provide a total cost estimate. The market price from a competitive market is a first-best estimate of the opportunity cost of a resource use but as most services in the Danish health care sector are not exchanged on a competitive market a second-best estimate has to be come from alternative sources. In practice this often leads to tariffs of the Diagnosis-Related-Grouping (DRG) system being used as unit cost estimates. The availability of the DRG-system is a convenient premise for conducting cost analysis in the Danish secondary health care sector. In principle, the system includes tariffs for all inpatient (DRG; www.drg.dk) and outpatient (Danish Ambulant Grouping System, DAGS) treatments in public hospitals. The National Board of Health manages the systems and updates the tariffs every year upon dialogue with the clinical communities. As the use of tariffs as unit cost estimates may be problematic a sensitivity analysis where different unit cost estimates are applied should always be considered.

One problem of using DRG- and DAGS-tariffs has been much debated, namely that these do not reflect the true opportunity costs of resource use (24). Although this debate has not prevented the widespread use of these tariffs it should be noted they do not include interest and depreciation of buildings and equipment (part of fixed costs) and, that they are average cost estimates not necessarily valid for a specific procedure type or a specific patient type. In the specific application of cost-effectiveness evaluation the interest is in incremental costs (between alternatives) and thus fixed costs may cancel out, in turn relieving the problem of depreciation not being included. However, fixed cost may not always be cancelled out e.g. if the two alternative interventions compared induce a subsequent difference in complications and co-morbidity and thus leading to a difference in resource use (and also a difference in the utilisation of buildings and equipment) in the health care sector.

To be able to choose a valid unit cost estimate a brief introduction to different cost concepts is appropriate. For example one needs to know that the difference between average costs and marginal costs is that fixed costs such as overhead are included in the average costs but not in the marginal costs. When analysing the consequences of introducing a novel technology, which is assumed to produce extra clinical effect, the objective is to assess what extra costs are related to acquiring such extra clinical effect. It is generally recommended that total costs of each of the alternatives being analysed are assessed then subsequently, one can subtract costs of alternative A from the costs of alternative B, say, to provide an incremental cost. Table 2 explains the definitions of key cost concepts. For cost-effectiveness analysis the ultimate interest is in the incremental cost, which typically denotes the average difference in total costs between alternatives.

Table 2 Definition of cost concepts.

Cost concept	Explanation
Fixed costs (FC)	Costs incurred by production regardless of its scale, e.g. investments
Variable costs (VC)	Costs that vary with the scale of production, e.g. materials
Total costs (TC)	All costs relating to the production of a quantity (q); $TC = FC + VC$
Average costs (AC)	The costs per unit produced; $AC = TC / q$
Marginal cost (MC)	The extra cost associated with producing one extra unit; $\Delta TC / \Delta q$
Incremental cost (ΔC)	Difference in costs between technology A and B; $\Delta C = C_A - C_B$

3.1.1 Intervention costs

In a trial assessing the cost-effectiveness of a novel technology for spine surgery, say, a micro-costing approach is often needed to estimate intervention costs because there are no DRG-tariffs available for novel procedures. Even when a technology is not new the DRG-system may not be an appropriate source of unit costs as it holds less than ten procedure tariffs specific to spine surgery.

Micro-costing denotes a branch of strategies where the idea, in principle, is to identify all activities and utensils related to a resource use, measure them individually and then add them up to form a cost estimate. For example if no unit costs exists for a surgical procedure the idea is to break the activity down into components that are measurable and associated with a valid item cost. This has been done for the procedure of lumbar spinal fusion as illustrated in Figure 2. From the figure it appears that 4 activity centres were defined, each holding a bunch of components that may or may not vary between patients; those not varying between patients were included as a fixed quantities and those varying were measured stochastically for every patient by reviewing patient files. Having broken down an activity into homogenous standard components generally means that cost estimates (either market prices or internal cost calculations) can be obtained from the hospital's accounting department. Time of doctors, nurses, secretaries etc. are typically valued using their gross wage, which can be obtained from The National Municipal Wage Data Office (www.fldnet.dk) or the hospital's wage administration office. Alternatively Statistics Denmark (www.statistikbanken.dk) offers wage statistics on a more general level. The gross wage is often adjusted for the fact that not all paid work hours are productive; non-productive time could be breaks, administrative tasks, meetings, conferences, courses and sick leave. This adjustment is often referred to as the application of a load-factor (for an example see Sogaard et al. 2005 (12)).

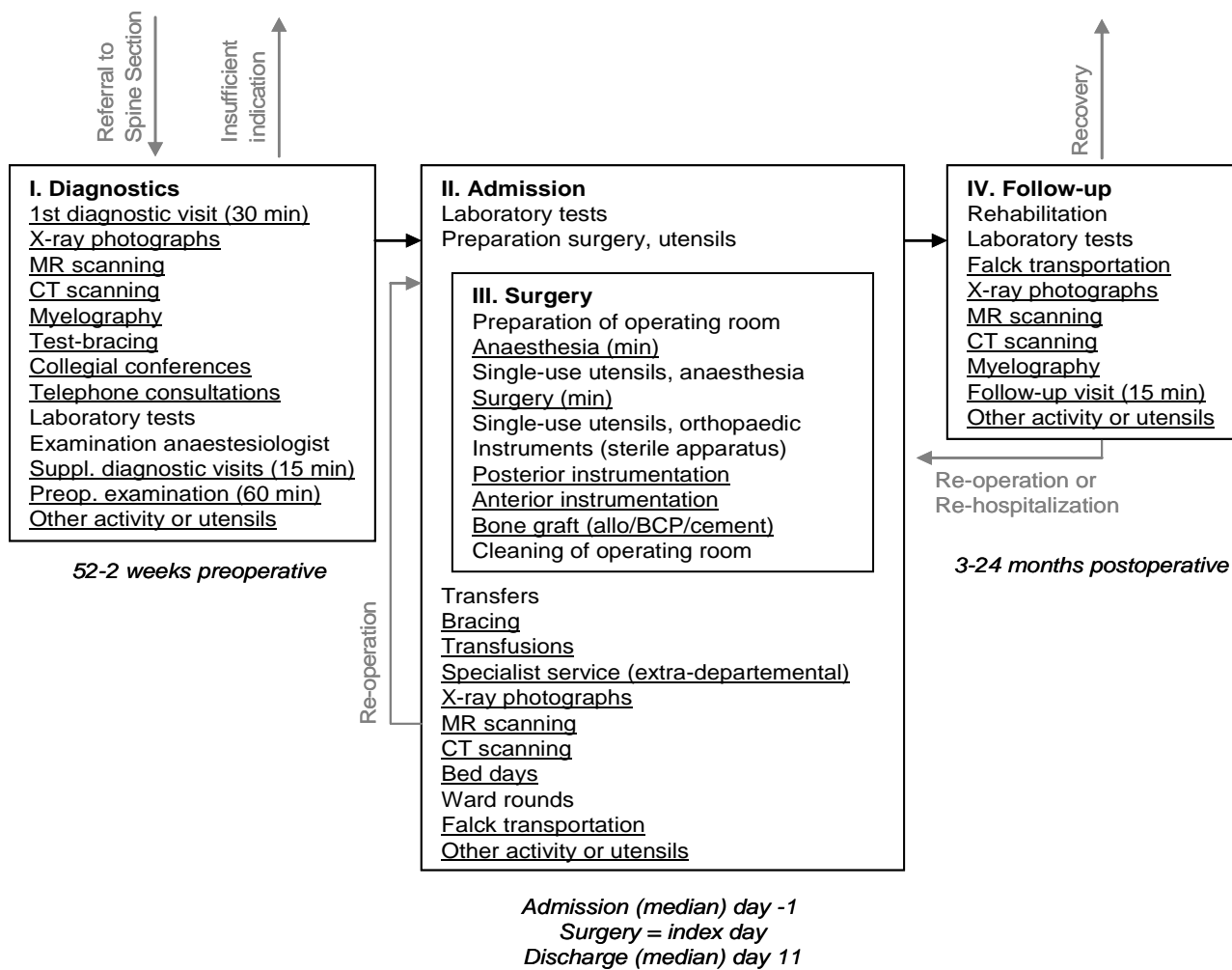


Figure 2 Activity-Based-Costing in lumbar spinal fusion (Søgaard et al. 2005(12))

An intervention may involve more sectors – the health care and the municipal care sectors, say – and thus assessing its costs requires data from different sources. If for example a new intervention is home-based rehabilitation, which is performed by municipal-employed physiotherapists, as compared to hospital-based rehabilitation it becomes relevant to estimate the costs in the municipality (experimental group) as well as at the hospital (control group). The cost in the experimental group will typically be the time spent by the municipal-employed physiotherapist, including the time spent on transportation to the patients’ homes and, if applied, the costs of using assistive technologies. The costs in the control group may have a relevant DRG- or a hospital tariff but then it should be given consideration whether patients’ costs of transportation should be included (as municipal-employed physiotherapists’ transportation were for the alternative).

3.1.2 Costs in the hospital sector

In addition to estimating intervention costs it is sometimes relevant to include also costs of co-morbidity, readmissions, reoperations etc. The following paragraphs will therefore concern counting of costs that are related to, but not included in the intervention costs.

While the assessment of the intervention costs often calls for a micro-costing approach it is usually feasible to use register data and DRG- or DAGS-tariffs to estimate the costs of co-morbidity and readmissions. Patient-level information about every somatic admission and outpatient visit can be found in The National Patient Register (declaration of the register etc. can be found via www.sst.dk). It should be noted that data on psychiatric admissions and outpatient visits are held in a separate register – The Psychiatric Central Register – that is also administered by the National Board of Health. The registers contain information about civil registration number, primary diagnosis, possible secondary diagnoses, possible procedure or treatment codes, date for contact and length of stay among others.

To access data from the registers of the National Board of Health a first step is to apply The Danish Data Protection Agency (www.datatilsynet.dk) for a permission to establish a research register. Having this permission an application for register data can be filled in directly from the homepage of The National Board of Health (Research Service department) where prices and delivery times etc. are also described: <http://www.sst.dk/Indberetning%20og%20statistik/Forskerservice.aspx>.

3.1.3 Costs in the primary health care sector

Resource use in the primary health care sector relates to services provided by general practitioners, privately practicing specialists, dentists, physiotherapists, chiropractors etc. This information is too monitored at the patient-level in the Primary Care Sector Register from which data can be acquired via the National Board of Health (same procedure as for secondary care data).

For every contact to the general practitioner, privately practicing specialists, dentists, physiotherapists, chiropractors etc. the Primary Care Sector Register contains information about civil registration number, date (week and year) for contact, type of services/examinations performed, number of services/examinations, speciality of the physician/healthcare personnel and the fee paid (each service provided is associated with a fee being paid to the provider by the public health insurance system). The consultation/examination is the measurement of the resource use and the fee is the valuation of the consultation/examination that is, the fees are applied as the unit cost estimates.

It should be noticed that the Primary Care Sector Register do not include information about the patient's diagnosis i.e. the register do not include information about the cause of the consultation. And again, it should be noted that the fee paid by the public health insurance system does not necessarily reflect the true opportunity costs; in particular, the remuneration of general practitioners is based a mixed payment system of approximately 1/3 capitation and 2/3 fee-for-service.

Further to service provision primary care includes prescribed medicine, which can be found in another patient-level register: the Danish Register of Prescription Medicine administered by The Danish Medicines Agency. The register holds information of all sale of prescribed medicine; for every record there is information about civil registration number, date of purchase, ATC-code, product name, package size,

dispensing form, sales price, the patient's copayment etc. Based on the ATC-code it is possible to extract records specific to a disease area. The sales price (incl. VAT) is typically used to represent the market price as an approximation of the opportunity cost estimate.

While it is possible to merge data of prescription medicine with other register data, the procedure for acquisition and analysis of data from this register is different and has to go via Research Service at Statistics Denmark (<http://www.dst.dk/TilSalg/Forskningsservice.aspx>). An authorization granted by Statistics Denmark is needed (on top of the approval by the Data Protection Agency previously introduced).

3.1.4 Costs in other sectors

Sometimes when intervening in one sector (secondary health care) costs occurs in another sector. A typical example would be that spine surgery affects the patient's ability to perform tasks of daily life, which then means that for example aids and assistive technology and/or home care is needed after discharge. Such costs are typically held at the municipal level and are not currently systematically registered in national databases (although individual municipalities might have registers). If this information is essential for the cost-effectiveness evaluation prospective data collection usually has to be initiated. There are validated questionnaires (in English) for this purpose; for example the cost-diary (26).

Many of these products and services are traded on competitive markets and thus market prices exist for their valuation. It should be noted that the cost of some aids or assistive technologies are shared between the municipality and the patient.

3.1.5 Costs for patient and family

There can be substantial costs for the patient and family during and after a spine surgery due to the patients' restrictions in functional ability, use of medicine etc. Such costs primarily fall into two categories: patient or family time and so called out-of-pocket expenses. "Costs for patients and their families are rarely monitored in registers and therefore has to be recorded ad hoc – either prospectively by the use of validated questionnaires or retrospectively by interviewing the patient, for example, when attending follow up visits at the hospital.

In addition to the possible need of formal care provided in the primary and secondary health care sectors and aids described above, patients' might also need informal care (unpaid care provided by for example relatives) or formal care (domestic assistance paid by the patient) in a period of time after surgery. This type of care enables the patient to perform conventional activities of daily living and the postoperative rehabilitation regimen. Often patients are discharged with recommendations not to drive a car and thus relatives' time is required to assist the patient in attending training sessions for rehabilitation for example.

The economic consequences of extra transportation required to undergo spine surgery and subsequent rehabilitation may be the transportation expense itself (taximeter or ticket cost), the time cost of a relative's time (if the spouse needs to drive the patient to the hospital) and, in principle, also the time cost of the patients. The ticket or taximeter cost is sometimes estimated by the distance (in kilometres) to the hospital or rehabilitation centre multiplied by the national standard tariff for transportation (in 2010 DKK 3.56 /km). Methods to valuate time costs will be described in productivity gains/losses paragraph.

As introduced previously, the cost of prescription medicine is shared between the patient and the primary health care sector and registered in the Danish Register of Prescription Medicine. The copayment scheme is regularly adjusted and the current version defines that patients are reimbursed according to their need; from DKK 0-850 per year the patients hold all costs themselves, from DKK 850-1,385 the patients hold 50% of the costs, from DKK 1,385-2,990 the patients hold 25% of the costs while the need for prescription that exceeds DKK 2,990 per year are reimbursed by 85% (more information on www.laegemiddelstyrelsen.dk).

A brief comment on the availability of a validated (specific to low back pain) cost diary is warranted. In order to collect patient-specific ad hoc data Goossens et al. developed a so called cost diary for the patient to fill in concurrently (26). The diary provides the possibility of collecting the full amount of out-of-pocket expenses on both medicine, transportation and other resource use as well as estimations on time spent on care and hence the productivity losses/gains.

3.1.6 Production loss/gain

Production losses or gains can be caused by changes in both morbidity and mortality. Morbidity costs are costs associated with permanent or temporary sick leave or impaired on-the-job performance. Mortality costs are costs due to positive or negative changes in life expectancy from an intervention.

The quality-adjusted life year (QALY), which will be introduced in the next section, is a common measure of the outcome in cost-effectiveness evaluations. It is usually said to include part of the consequences to patients' ability to work of an intervention. Relating to the patients' ability to work (and production loss/gain) in particular, consideration should be given not to double-count. This would occur if a regained ability to work, say, is included as part of the outcome (by improved quality of life) and as a saved cost (by production gain). Similarly production loss/gain due to changes in life expectancy are included in the QALY-measure (27). Among others due to the double counting issue productivity changes should be reported separately and transparent.

Despite the debate about double counting it is common practice to include production loss due to morbidity. This refers to costs occurring because patients or their relatives are disabled to work due to the disease, due to attending the treatment regimen or due to caring or being cared for in the home. These costs usually vary depending on the patient's employment and family situation. Patients who are already on pension or early retired are usually considered in an irreversible category and thus incur no production gain or loss. Patients who are working, unemployed (but available to the labour market) or on sick leave incur production gain or loss if their status changes, i.e. the patient baseline situation is important in order to estimate changes in productivity affected by an intervention. If for example a patient is on a sick leave due to their illness and can return to work after surgery and recovery, the surgical intervention has created a productivity gain that should be subtracted from the costs. The opposite example could also be likely. If for example a patient has been able to work prior to surgery but needs sick-leave or maybe even (early) retirement after a surgical intervention, the lost years of possible production should be counted.

The quantification of time lost (or gained) at work for both patient and family is the first step in estimating production losses/gains. It should be commented that cost-effectiveness studies increasingly considers also presenteeism (on-the-job productivity loss) as a source of productivity losses.

Next, a value has to be placed on that amount of time. The most common ways to apply costs to the time spent on care or off work is the Human Capital (HC) or the Friction Cost (FC) approach.

The HC method accumulates the patients' and possible informal caregivers' hours of production that are lost due to the illness and its treatment and estimates productivity costs as the product of that quantification multiplied by gross salaries for respective individuals. Accordingly, this approach considers every hour not worked as a production loss (i.e. workers cannot be replaced by unemployed from the working force, say). As a result, the HC method has been criticized for overestimating productivity losses and estimating potential costs, rather than actual costs (28).

The FC method was introduced partly as a result of criticism and thus takes into account the context of the labour market where a worker may be replaced with another worker from the unemployed work pool. Accordingly, it counts production loss attributable to hours occurring before a replacement worker takes over the patient's work (the friction time).

While the HC and the FC approaches are the most commonly used, it should be noted that there are other methods to value production that can be seen as a mixture of the two. All the methods require use of general wages and employment data. These are offered by Statistics Denmark (www.statistikbanken.dk), where wage statistics and work force statistics (unemployment rates, data on full or part-time employment etc.) can be found. For public employees detailed data on wages can be found in The National Municipal Wage Data Office (www.fldnet.dk).

For further details on the HC and FC methods and their differences in estimations of the value of productivity losses see (17;28).

In the specific context of evaluations alongside clinical trials in spine surgery the effect of including productivity costs is limited by the time horizon of analysis. With a two-year follow-up period, say, productivity effects might not be fully captured and that is an important issue for discussion when reporting such cost-effectiveness.

3.1.7 Future costs

Having included productivity effects because, for example, that a patient has regained the ability to work one should, in principle, also include other future consequences. Future costs include both unrelated costs and health costs generated as a result of a patient's functional ability and/or lifetime being affected. The most evident example is that when an intervention increases patients' life expectancy, the patients will *ceteris paribus* consume extra health care services in the gained life years. As these extra life years comes on top of the baseline life expectancy, patients should be followed for lifetime to account for such costs. Furthermore, as future costs can happen in all sectors, the identification of all future costs can be difficult.

In practice, however, spine surgery affects morbidity rather than mortality. This along with the fact that most trial-based cost-effectiveness evaluations are pragmatic and do not extend the length of the clinical follow up time means that the issue of whether or not to include future costs becomes irrelevant. But, the issue is indeed relevant to include in relation to the discussion about time horizon, which should appear in a limitations section of a cost-effectiveness report given that a limited time frame is adapted from the clinical trial.

3.1.7.1 Omission of costs

Costs that are omitted due to lack of data in registers and no prospective data collection most likely have an effect on the cost-effectiveness ratio. Literature has shown that for spine disease these types of costs vary based on diagnosis and treatment. There seem to be a tendency towards a greater use of informal care of surgical versus non-surgical patients (29). Others have shown that the value of production losses can be even as high as all other costs (30) while the production losses are a minor expenditure in yet other studies (31).

It is important to notice that if the costs can be regarded as equal for both arms in the study, they will not affect the incremental costs. Suspected differences in costs due to for example differences in rehabilitation hours or content, number of hospital visits or inpatient days etc. should thus be given consideration. It is important to describe if and in which direction omission of costs affects the incremental cost-effectiveness ratio – this should be reported in any cost-effectiveness evaluation where relevant.

3.2 Discounting and price level

Comparison of the costs and the consequences in an economic evaluation must be made at one point in time, which is usually the present reflecting that evaluations are conducted to support present decision-making about the acceptance a novel technology. However not all costs and consequences occur in the present and therefore adjustment for different time profiles must be made. The rationale for adjustment is that individuals generally prefer to have a positive consequence today rather than tomorrow and vice versa for a negative consequence (or a cost) – this is referred to as positive time preference in economics.

The method for adjustment is to discount future costs and consequences using the discount factor $1/(1+r)^t$, where r indicates the chosen discount rate and t the time in years. The current value of a cost of DKK 5,000 that occur in 4 years, say, will be $5,000/(1+0.03)^4 =$ DKK 4,442 at a discount rate of 3%. The exact discount rate is somewhat arbitrary but often a rate of 3 or 5% is chosen for a base-case analysis, which is then subject to sensitivity analysis using alternative rates.

Further to adjusting all costs and consequences to the present time, it is important to make sure that the unit costs applied are of the same price year. If for instance some unit costs are in 2009-DKK and some others in 2010-DKK one of the years should be selected after which all estimates not in the specific year should be converted. This is usually done using a price index, which can be found at the homepage of Statistics Denmark (www.statistikbanken.dk).

3.3 Budget impact or fiscal analysis

Combining clinical research with health economic evaluation provides a thorough insight to both clinical effects and costs of a new technology. This combination is very relevant to decision makers, who are to decide whether or not to spend resources on a given intervention or technology when working in a public health care system. Conventional economic evaluation does not, however, provide decision-makers with an impact analysis in relation to the budget constraint they face. Budget Impact analysis is a general approach to demonstrate how the acceptance of a new technology will impact the national, regional, or local health care budgets. A Budget Impact analysis hence addresses the financial flow of resources related to the

implementation of a technology whereas CEA evaluates the costs and consequences of alternatives to estimate their efficiency and not affordability for the budget holders and decision makers.

Costs for the individual budget holders will often differ, and savings on one budget may result in expenditure for another. For example an investment at one hospital department may lead to expenditures for another at the same hospital or a saving for a region may lead to expenditures for a municipality. Therefore the costs for each involved budget holder are important

Many of the cost data used for the economic evaluation can be re-used in a Budget Impact analysis and the methodological requirements are similar although the intention in use is substantially different. Hence only costs and consequences relevant to the budget holder are counted and e.g. family and patient costs or productivity costs are rarely included.

For more information look at Mauskopf et al. (32) and Kristensen et al. (24).

4 Measuring and valuing outcomes

Few people can disagree that 'health' is more than being alive, yet its definition is inherently subjective and many views have been presented. One widely accepted definition was suggested in 1946 by the World Health Organization (WHO): "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (33). Its critics however argue that health cannot be defined as a state at all, but must be seen as a process of continuous adjustment to the changing demands of living and of the changing meanings we give to life. In particular, it has been argued that the ability to lead a socially and economically productive life should be part of the definition. Today, this view has led to most instruments for outcome assessment including dimensions about the respondents social functioning as well as ability to perform paid and unpaid work.

The remainder of this section examines consequences as a synonym for health consequences in its broadest definition. The section is divided in two: an overview of alternative instruments for measuring and valuing health consequences and a few more specific comments on the choice between two of the most widely used instruments, the SF-6D and the EQ-5D.

4.1 *Generic, preference-based instruments*

Assessing health consequences inevitably requires the two steps of identification and measurement. Measurement that leads to a quantification of physical units is basically sufficient to inform cost-effectiveness evaluation whereas for cost-utility evaluation, a third step of assigning a value of utility to the measured consequences is required. The notion of utility is a key term in economics that expresses the satisfaction an individual gets from acquiring or consuming a good. Intuitively, one can think of a hierarchy of outcome measures with the least complex being the one-dimensional, disease-specific measure and the more complex being the multidimensional, generic measure that had been assigned a utility-weight – thereof the notion of preference-based – as listed in Table 3.

A preference-based measure is thus defined by its scope of condensing the multidimensional construct of health into a single-index score, principally ranging from zero (dead) to one (full health). This process requires at least two components: an instrument that classifies health status and a scoring algorithm that assigns a preference-value to health states. While the classification component usually takes form of a questionnaire aimed at study participants, the valuation component is usually a scoring model derived from a survey of the general population's preference-values.

Table 3 Types of outcome measures for cost-effectiveness and cost-utility evaluation in health care

Measure	Output	Applicability for prioritisation
Disease-specific	Quantity	Specific disease (CEA)
Generic	Descriptive classification	Health care sector (CEA)
Preference-based generic	Single-index	Health care sector (CUA)
Willingness to pay	Monetary unit	Whole society (CBA)

Note: CEA = Cost-effectiveness evaluation, CUA = Cost-utility evaluation, CBA = Cost-benefit evaluation.

In the 1970s, some early preference-based measures of the Index of Well-Being (later termed the Quality of Well-Being index, QWB) and the Rosser classification (RC) were proposed in the United States and in the United Kingdom, respectively(34;35). The RC classified individuals into as few as 29 health states, which were valued using a sample of only 70 respondents. For these reasons the sensitivity to identify differences between diseases as well as sensitivity to change over time was somehow restricted. During the 1980s, the early measures were refined with respect to classification as well as valuation components while new instruments emerged: the 12D (later termed the 15D) and the Health Utility Index (HUI), among others (36;37). The most recent measures came about in the 1990s with the EQ-5D and the Australian Assessment of Quality of Life (AQoL) being two of the most influential developments (38;39).

In 1999, Brazier et al. reported a systematic review of the use of health status measures for economic evaluation (40). They identified five preference-based, generic measures: RC, QWB, HUI, 15D, and EQ-5D. Claiming conventional psychometric tests of validity inappropriate for the evaluation of economic validity, the authors instead assessed the instruments' ability to describe health as well as valuation components' theoretical and empirical validity. On grounds of limited ability to describe health, the RC was found inferior to the others. Further two measures, the QWB and the 15D, were judged inferior to others because their valuation components were not established using choice-based techniques. In terms of practicality all of the measures, except for the QWB, were found to be brief and easy to use. In terms of test-retest reliability, three measures were found to be supported from evidence: HUI, 15D, and EQ-5D, leading to an overall recommendation of HUI and EQ-5D being the best choices for outcome assessment in economic evaluation.

Since the review of Brazier et al. another preference-based, generic measure has evolved, namely the SF-6D (41) which was derived from the, perhaps, most widely used instrument clinically, the SF-36(42). If that extensive amount of validation studies carried out for the SF-36 is directly transferable to the SF-6D, it holds a major potential for the discipline of economic evaluation; the enormous amount of historic and concurrent (non-preference-based) SF-36 data has now become usable for economic evaluation as preference-values have been assigned.

In accordance with the recommendations of Brazier et al. (40) plus the addition of the most recent development, the SF-6D, the choice of measure for generic outcome assessment in economic evaluation stands between the HUI, the EQ-5D, and the SF-6D. The current recommendations of the UK National Institute for Health and Clinical Excellence (NICE) are to choose between EQ-5D or SF-6D (43) whereas the HUI is not explicitly mentioned, probably, because it offers no version aimed at adults which is valued by the UK general population.

The SF-6D came about in its first version in 1998 as a result of Brazier and colleagues' conceptual restructuring of the SF-36 into some ranked levels of selected dimensions: physical functioning, role limitations, social functioning, pain, mental health, and vitality(44). In 2002, a revised version, which was valued using the standard gamble (SG) technique in a representative sample of the UK general population was reported (41). The revised version included the same dimensions as the first (although the item-mix was different) and presented with four to six levels of function in each dimension, producing a total of 18,000 health states. As the original model contained some logical inconsistencies a consistent model that has not been formally published later on replaced it. The SF-6D is not yet valued in a Danish setting.

The EQ-5D was developed in a large multidisciplinary group, the EuroQol group, which proposed the first version of the instrument in 1990 (38). The original dimensions, selected after review of other generic health status measures, included mobility, self-care, main activity, social relationships, pain, and mood (45). These were shortly modified into the current version including the dimensions mobility, self-care, usual activity, pain/discomfort, and anxiety/depression, each with three levels of function and thus producing a total of 243 health states (245 when added, for completeness, *unconscious* and *immediate death*). The most influential valuation study of the EQ-5D was conducted by the Measurement and Valuation of Health (MVH) group at York, using the time-trade-off (TTO) technique in a representative sample of the UK general population (46;47). Danish weights are available for the EQ-5D (48) as well as population norms (49).

4.2 The choice between the SF-6D and the EQ-5D

The choice between instruments can be informed from several empirical studies examining the comparative performance of the two measures in low back pain. For example, Sogaard et al. found the SF-6D to produce a mean value of on average 0.085 higher than that of the EQ-5D (50). Such discrepancy may seem moderate but it was concluded that it masked considerable bidirectional variation: for values below the mean average of the two, the SF-6D produced significantly higher values and vice versa for values above the mean average of the two. The expected variation for any true average of future observations was estimated at 0.55 (at the scale where 0 corresponds to dead and 1 to perfect health), obviously violating the idea that the two can be used interchangeably. Other studies have come to similar conclusions (51;52).

Conventional psychometric properties of construct validity, reliability and practicality of the EQ-5D and the SF-6D have been established in low back pain (5;53;54). Responsiveness has also been examined, however, for the EQ-5D only. Yet, a simple comparison of the dimensions of the two instruments as illustrated in Table 4 reveals that they are not measuring identical constructs. First, the SF-6D includes a dimension, vitality, which is not covered in the EQ-5D. Second, the weighting of included dimensions is different between instruments; physical functioning for example is addressed by two of five dimensions in the EQ-5D in comparison with only one of six dimensions in the SF-6D. The inverse is the case for psychosocial

functioning, which is addressed by three of six dimensions in the SF-6D and only one of five in the EQ-5D. The domain of pain is covered in a resembling manner in the two instruments although the SF-6D focuses on the disability associated with pain whereas the EQ-5D focus on pain and discomfort per se.

Another evident issue is that measures address different levels of severity. For example, the worst function in like dimensions relating to physical functioning is referred by “limits you” in the SF-6D whereas the wording in the EQ-5D is “unable” or “confined to bed”. This could enable a floor effect of the SF-6D whereas for the EQ-5D, the limited number of levels (only three) could lead to a poor ability to discriminate between health states and possibly a ceiling effect. These theoretical predictions have moderate support from the literature specific to low back pain.

Having argued, that the SF-6D and the EQ-5D cannot be used interchangeably the question is what measure is the optimal for coming trials in spine surgery? There is no gold standard from a theoretical viewpoint of economic theory or psychometrics, and both measures have demonstrated their practicality. That being said it is important to remember that spine surgery addresses an extremely heterogeneous population with large variation in levels of quality of life. The difference in mean values between SF-6D and EQ-5D is most significant in populations suffering poor health whereas they seem almost interchangeable in good health. However, the average patient of a clinical trial typically improves over time and thus trialists have to consider both the expected baseline and endpoint values.

Table 4 Dimensions of SF-6D and EQ-5D (levels are indicated in parentheses)

SF-6D		EQ-5D	
Physical Functioning		Mobility	
Your health does not limit you in vigorous activities	(1)	No problems walking about	(1)
Your health limits you a little in vigorous activities	(2)	Some problems walking about	(2)
Your health limits you a little in moderate activities	(3)	Confined to bed	(3)
Your health limits you a lot in moderate activities	(4)	Self care	
Your health limits you a little in bathing and dressing	(5)	No problems with self-care	(1)
Your health limits you a lot in bathing and dressing	(6)	Some problems washing or dressing myself	(2)
		Unable to wash or dress self	(3)
Role limitations		Usual activities	
You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems	(1)	No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)	(1)
You are limited in the kind of work or other activities as a result of your physical health	(2)	Some problems with performing usual activities	(2)
You accomplish less than you would like as a result of emotional problems	(3)	Unable to perform usual activities	(3)
You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems	(4)		
Social functioning			
Your health limits your social activities none of the time	(1)		
Your health limits your social activities a little of the time	(2)		
Your health limits your social activities some of the time	(3)		
Your health limits your social activities most of the time	(4)		
Your health limits your social activities all of the time	(5)		
Pain		Pain/discomfort	
You have no pain	(1)	No pain or discomfort	(1)
You have pain but it does not interfere with your normal work (both outside the home and housework)	(2)	Moderate pain or discomfort	(2)
You have pain that interferes with your normal work (both outside the home and housework) a little bit	(3)	Extreme pain or discomfort	(3)
You have pain that interferes with your normal work (both outside the home and housework) moderately	(4)		
You have pain that interferes with your normal work (both outside the home and housework) quite a bit	(5)		
You have pain that interferes with your normal work (both outside the home and housework) extremely	(6)		
Mental health		Emotions	
You feel tense or downhearted and low none of the time	(1)	Not anxious or depressed	(1)
You feel tense or downhearted and low a little of the time	(2)	Moderately anxious or depressed	(2)
You feel tense or downhearted and low some of the time	(3)	Extremely anxious or depressed	(3)
You feel tense or downhearted and low most of the time	(4)		
You feel tense or downhearted and low all of the time	(5)		
Vitality			
You have a lot of energy all of the time	(1)		
You have a lot of energy most of the time	(2)		
You have a lot of energy some of the time	(3)		
You have a lot of energy a little of the time	(4)		
You have a lot of energy none of the time	(5)		

5 Reporting cost-effectiveness evaluations

This section will deal with the final part of a cost-effectiveness analysis, namely how to analyse the collected data and report the results. The section will introduce both the reporting of base-case and sensitivity analyses and touch upon statistical analysis and decision scenarios. Again, a more detailed guidance is referred to textbooks of for example (17).

5.1 Base-case analysis

Base-case analysis is a term used for the most likely scenario as any evaluation is based on certain assumptions due to various uncertainties. It is thus the full-scale analysis generating the summary measures for the primary results section of a cost-effectiveness report.

If possible the reader of a cost-effectiveness evaluation should be able to follow the measured quantities of resource use, the unit costs used, and the resulting total costs (both as a raw mean and discounted) for all alternatives included in the analysis. This is important since unit costs are rarely generalisable and thus enabling the recipient to recalculate the cost difference between alternatives supports the usability of results. Needless to say the same applies for the effect measure which, if QALYs, should be reported both as the observed utility index values and as QALYs (both as a raw mean and discounted) for all alternatives in the analysis.

To conclude which of the technologies that are the most efficient seen from an economic perspective the total costs must be compared to the total effects.

When comparing the costs and effects there are several possible outcomes, which lead to different decision scenarios. In two somewhat rare cases in practice a cost-effectiveness analysis can lead to obvious decision scenarios, where further analysis is redundant. These scenarios are 1) when the new treatment leads to a greater effect (more QALYs) at lower costs and 2) the opposite scenario where the new technology is both more costly and less effective. In the two remainder possible scenarios, cost-effectiveness is less straightforward to determine and a threshold value for decision-maker's willingness to pay is required to operationalise the recommendations of an evaluation. These scenarios are 3) where the new technology provides greater effect (more QALYs) at a higher cost (this is the most common outcome of an evaluation) and 4) where the new technology provides less effect but also cost savings (this is more controversial and rarely seen in a Danish context as new technologies that are clinically inferior are usually not launched at all). The following will therefore concern the scenario 3) which in practice also covers scenario 4).

When a new technology is found to be more effective and more costly the decision-analytic question concerns whether the extra effect gained from the technology is worth its extra costs. Decision-makers are usually given an incremental cost-effectiveness ratio (ICER), which expresses the cost of one extra unit of effect produced with the new technology – for example, the cost per additional QALY if investing in the new technology. The decision rule for accepting a new technology is:

$$ICER = \frac{C_A - C_B}{E_A - E_B} \leq \lambda$$

Where C is total average costs and E is total average effects for the alternatives A and B. λ is the decision-makers' willingness to pay per unit of effect.

The cost-effectiveness of a new technology that provides extra effect at extra costs as compared to usual practice, say, depends upon the maximum price that one is willing to pay for that extra effect. It does not make sense to claim that something is cost-effective or not without mentioning a threshold value for willingness to pay, although often seen in the literature. It should also be noted that no true value for willingness to pay exists although attempts have been made to elicit one using preference-studies as well as some have sought to estimate it using the revealed (minimum) willingness to pay from what is implemented in the health care sector today.

5.2 Parameter uncertainties: bootstrapping

The distributions of costs and effects are most often skewed (many patients express moderate costs and an expected outcome while a very few patients express extreme costs and/or a zero or even negative effect) and therefore classical summary statistics are not applicable. Sometimes data can be transformed using for example log-transformation or non-parametric statistics can be employed but most often this is not sufficient for a valid analysis. Furthermore, the ICER is a two-dimensional summary measure of four stochastic variables as opposed to conventional one-dimensional measures where for example a t-test is applicable. For these reasons the technique of bootstrapping has emerged and is common practice in cost-effectiveness evaluation.

Bootstrapping is less restrictive than conventional statistics and has several advantages; in particular it provides precision estimates that can be interpreted as the usual confidence intervals. Another advantage is the intuitive interpretation when bootstrapped replicates are presented graphically. Figure 3 illustrates such representation in the cost-effectiveness plane with average costs on the one axis and average effects on the other. The plane here is not comparative; each dot represents a best bet on the combination of average costs and average effects for a single intervention – and importantly, each dot is a simulation of the average cost and the average effect in a target population, not observed values in the sample (an often heard misunderstanding). It can be seen that the intervention is associated with significantly higher costs, as all replicates are located above zero. Nothing can be concluded about the effect as replicates are spread equally on both sides of the y-axis i.e. some recipients can expect a negative impact on survival from the intervention whereas others can expect to benefit.

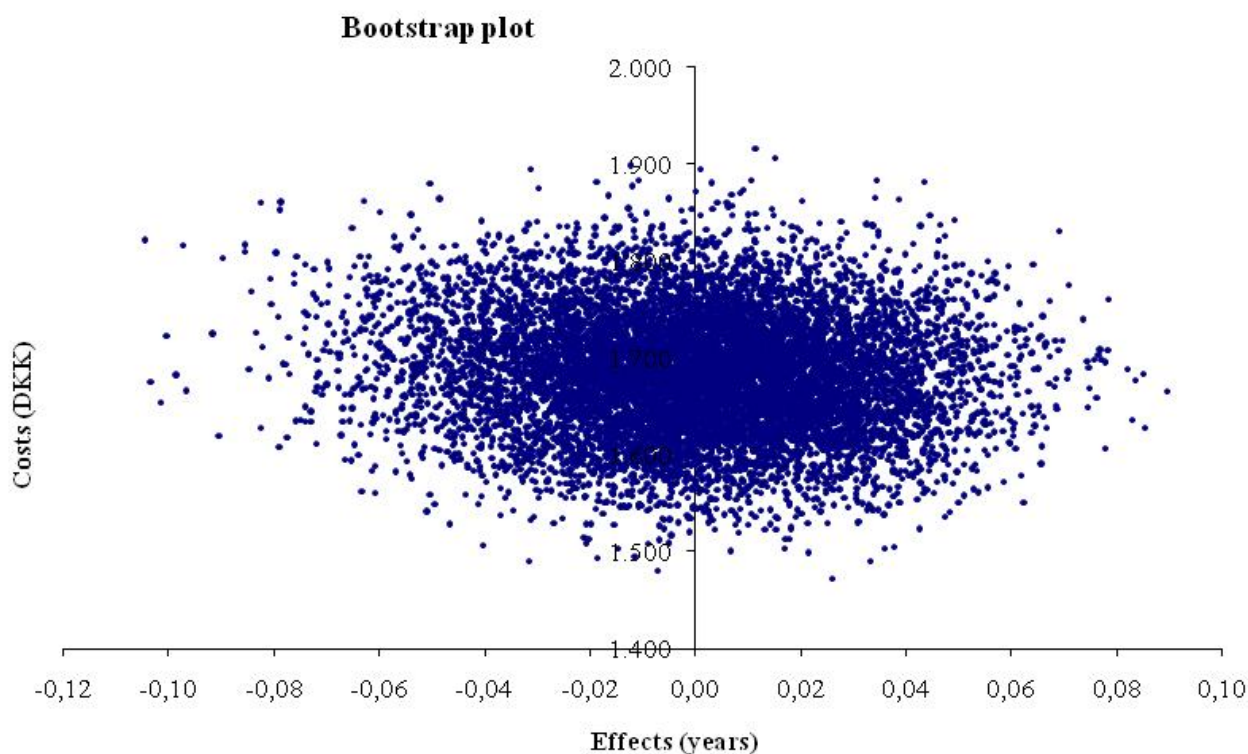


Figure 3 Cost-effectiveness plane with bootstrapped estimates of pairs of average costs and average effects.

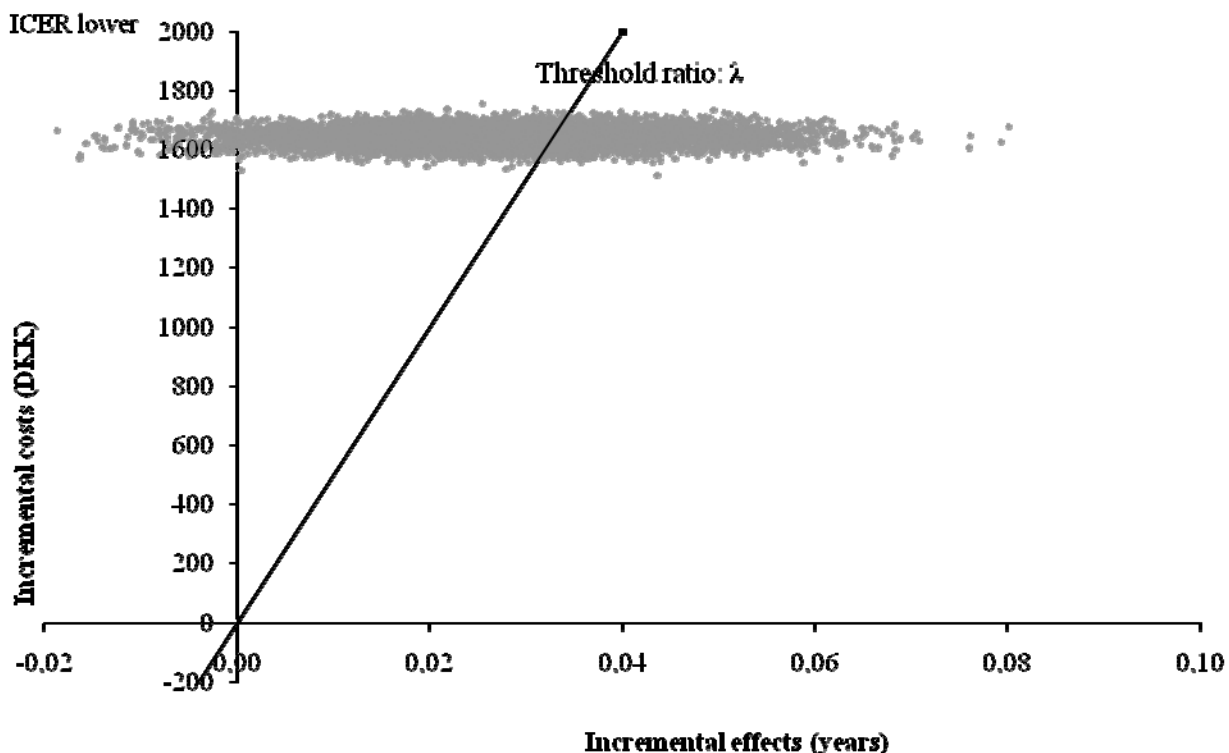


Figure 4 Cost-effectiveness plane illustrating bootstrapped values of the Incremental Cost-Effectiveness Ratio (ICER) of an intervention as compared with its control.

The plot of average costs and average effects for a single intervention should be considered as a piece of nice-to-know information when reporting a cost-effectiveness analysis in a journal; it is important for the researcher to fully understand the aetiology behind the ICER – and here understanding individual intervention’s performance is a necessity – but given the limited space in scientific journals, non-comparative illustrations can be omitted. A recommended illustration though is the same cost-effectiveness plane but with each dot representing a bootstrapped replication of the ICER. This is shown in Figure 4, demonstrating a hypothetical scenario with a (new) intervention that is significantly more costly and on average more effective. The question therefore is whether decision-makers are willing to pay the extra cost for extra health gains. Decision-makers maximum willingness to pay can be illustrated by means of the slope of a line from the origin of the cost-effectiveness plane. The exact slope is often unknown but for all slopes it is possible to count the proportion of replicates that are located below this line, i.e. for any hypothetical value of decision-makers maximum willingness to pay a probability for the intervention being cost-effective can be derived. A counter clockwise rotation of this line around the origin implies an increasing willingness to pay, and the proportion of replicates below the line would increase. This is basically what is already done in Figure 5, which is the key result of a cost-effectiveness report: the probability that an intervention is cost-effective as a function of the threshold value of willingness to pay – often referred to as cost-effectiveness acceptability curves (CEAC).

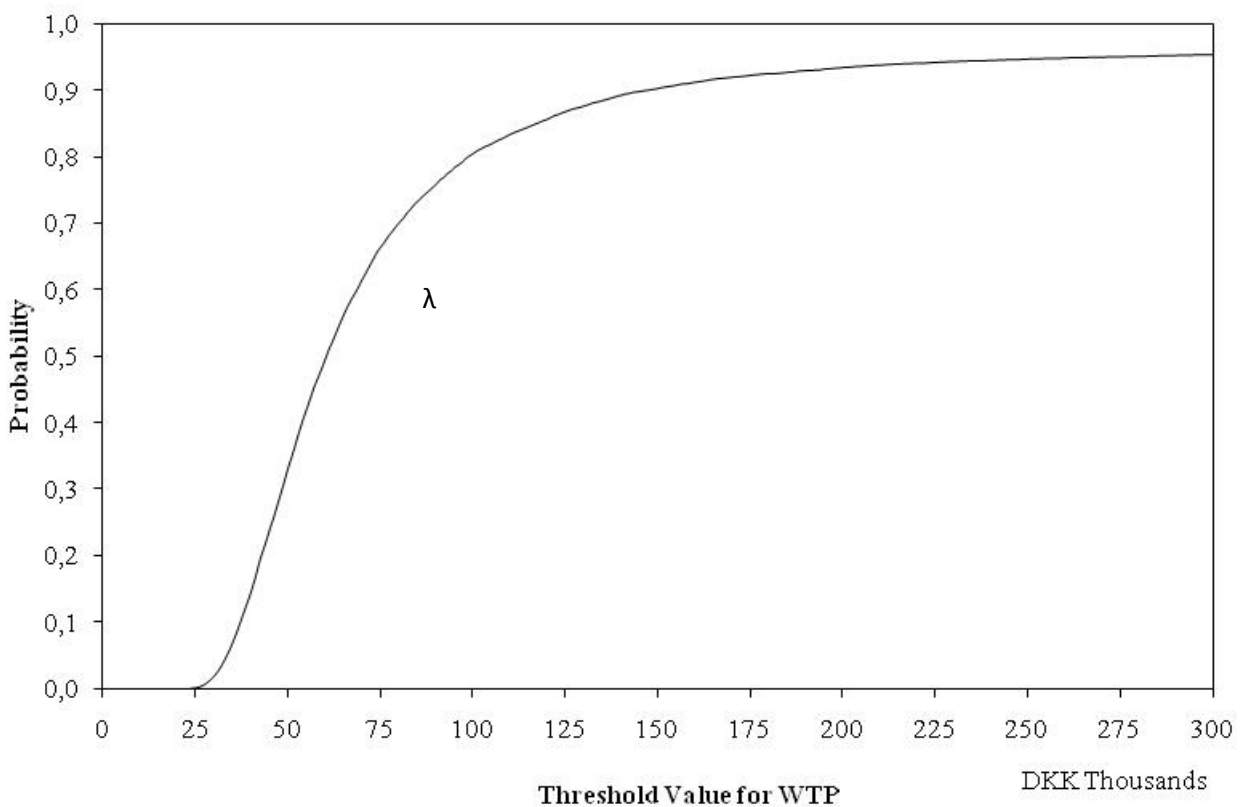


Figure 5. Cost-effectiveness acceptability curve

A CEAC thus presents the probability that the ICER falls below the maximum willingness to pay. In the given example in Figure 5, a decision maker with a willingness to pay per QALY of 100.000 DKK could assume a probability of 80% for the intervention being cost-effective.

As the cost-effectiveness acceptability curve is a key part of reporting you might want to consult some of the original literature (55-58). For the more pragmatic interpretation of cost-effectiveness curves there are as well several papers (59;60).

5.3 Structural uncertainties: sensitivity analysis

There are at least two types of uncertainty relating to the results of a cost-effectiveness evaluation. While the former paragraph commented on statistical (sampling) uncertainty this paragraph comments on structural uncertainty, i.e. uncertainty relating to assumptions made for the final design of the analysis. This particular type of uncertainty is independent from the sample size and therefore the whole system relies on whether assumptions are reasonable. The sensitivity of results to alternative assumptions should therefore always be tested; if one parameter is varied at a time the analysis is denoted one-way sensitivity analysis, if two-parameters are varied at a time the analysis is denoted two-way sensitivity analysis, etc. If conclusions doesn't change when the assumptions and estimates does, the conclusion is said to be robust to the tested alternative assumptions.

One-way sensitivity analysis is the most common in analysis alongside single trials. By that every single parameter included in the cost-effectiveness analysis is varied at a time and the effect on the ICER is observed. Typical parameters include unit costs, discount rate, compliance, included cost types, effect measure (e.g. QALY weights) etc.

When performing the sensitivity analysis the varying of variables will most likely result in a change in ICER and hence a change in the distribution of data. This can lead to a different CEACs and decision scenarios due to changed probability of acceptance of the procedure based on the hypothetical willingness to pay for the gained effect. It is important to state the changes in analysis based on the sensitivity analysis, and to discuss the robustness of data.

5.4 Recapitulation

This paper has focused on selected issues of how to prepare a cost-effectiveness evaluation of an intervention in spine surgery. Following a general introduction to the methodology two sections focussed on the costing side and the effect side, respectively, and this final section has given some comments on what is recommended for the analysis and reporting. There may be other views on how this is best conducted but the present paper represents an agreed upon consensus in the CESpine project, which meets general state-of-the-art.

It was recommended to report the following in a cost-effectiveness report for publication in a scientific clinical journal: a table of resource use (a column for each of the comparators and a column with differences, preferably with relevant significance tests), a table (or a detailed paragraph in the methods section) of unit costs, a table of costs (a column for each of the comparators and a column with differences, preferably with relevant significance tests), a table or a figure of utility index for each point in time of measurement (for individual groups, their difference and a significance test), and a CEAC. The cost-

effectiveness plane with bootstrapped replicates of the ICER is a nice gesture if the journal allows several figures but this should not replace the CEAC, which is the key diagram for decision-makers. Finally, the results of the sensitivity analysis should always be reported - in a table or as an extra CEAC with a curve for every alternative scenario. Needless to say, relevant description of trial design and sample characteristics are assumed to begin with.

The uniformity of evaluations in the CESpine project will be unique if individual researchers choose to adhere to this recommendation. Reporting several evaluations using identical methodology opens for a true comparability, which is rarely seen in most of the applied literature but nevertheless is assumed for results to guide resource allocation decisions.

6 References

- (1) Andersen CK, Søgaard J, Gyrd-Hansen D. Hvad koster ryg- og bevægeapparatslidelser samfundet og "kasserne". Odense: Odense Universitet; 1996.
- (2) Hagen KB, Thune O. Work incapacity from low back pain in the general population. *Spine* 1998 Oct 1;23(19):2091-5.
- (3) Eriksen J, Jensen MK, Sjogren P, Ekholm O, Rasmussen NK. Epidemiology of chronic non-malignant pain in Denmark. *Pain* 2003 Dec;106(3):221-8.
- (4) Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. *Arch Intern Med* 2009 Feb 9;169(3):251-8.
- (5) Solberg TK, Olsen JA, Ingebrigtsen T, Hofoss D, Nygaard OP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. *Eur Spine J* 2005 Dec;14(10):1000-7.
- (6) Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-MacDonald J. Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial. *BMJ* 2005 May 28;330(7502):1239.
- (7) Pickard AS, Wilke CT, Lin HW, Lloyd A. Health utilities using the EQ-5D in studies of cancer. *Pharmacoeconomics* 2007;25(5):365-84.
- (8) Frymoyer J, Duret C. The economics of spinal disorders. In: Frymoyer JW, editor. *The adult spine: Principles and Practice*. 2 ed. Philadelphia: Lippincott-Raven; 1997. p. 143-50.
- (9) Luo X, Pietrobon R, Sun SX, Liu GG, Hey L. Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. *Spine* 2004 Jan 1;29(1):79-86.
- (10) Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain* 2000 Jan;84(1):95-103.
- (11) The Swedish Council on Technology Assessment in Health Care. Back pain. *Int J Technol Assess Health Care* 2000;16(3):929-43.
- (12) Søgaard R, Christensen FB. The economics of spondylolysis. In: Szpalski M, Gunzburg R, editors. *Spondylolysis, spondylolisthesis and degenerative spondylolisthesis. Controversies in ethiopathogenesis, management and costs*. 1 ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
- (13) Maetzel A, Li L. The economic burden of low back pain: a review of studies published between 1996 and 2001. *Best Pract Res Clin Rheumatol* 2002 Jan;16(1):23-30.
- (14) Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005 Jun 15;30(12):1441-5.

- (15) Wonderling D, Gruen R, Black N. Introduction to Health Economics. Oxford: Open University Press; 2005.
- (16) Fox-Rushby J, Cairns J. Economic Evaluation. Oxford: Open University Press; 2005.
- (17) Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddard G. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford: Oxford University Press; 2005.
- (18) Drummond M, McGuire Ae. Economic Evaluation in Helath Care. Merging theory with practice. Oxford: OUP; 2001.
- (19) Gyrd-Hansen D. Willingness to pay for a QALY: theoretical and methodological issues. *Pharmacoeconomics* 2005;23(5):423-32.
- (20) Wagstaff A. Health rationing in Europe: can cancer get a fair hearing? *Cancer World* 2009 May 1;30.
- (21) McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008;26(9):733-44.
- (22) van TM, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol* 2002 Dec;16(5):761-75.
- (23) Sogaard R, Bungler CE, Laurberg I, Christensen FB. Cost-effectiveness evaluation of an RCT in rehabilitation after lumbar spinal fusion: a low-cost, behavioural approach is cost-effective over individual exercise therapy. *Eur Spine J* 2008 Feb;17(2):262-71.
- (24) Kristensen FB & Sigmund H (ed.). Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health; 2007.
- (25) Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 3rd ed. Ottawa: 2006.
- (26) Goossens M, Rutten-van Mólken M, Vlaeyen J, van der Linden S. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *Journal of Clinical Epidemiology* 2000;53(7).
- (27) Luce B, Manning W, Siegel J, Libscomb J. Estimating Costs in Cost-Effectiveness Analysis. In: Gold M, editor. *Cost-Effectiveness in Health and Medicine*. 3 ed. Oxford Medical Publications; 2005.
- (28) van den Hout W. The value of productivity: human-capital versus friction-cost method. *Ann Rheum Dis* 2010;69:89.
- (29) Fritzell P, Haag O, JOnsson D, Nordwall A. Cost-Effectiveness of Lumaber Fusion and Nonsurgical Treatment for Chronic Low Back Pain in the Swedish Lumbar Spine Study: A Multicenter, Randomized, Controlled Trial From the Swedish Lumbar Spine Study Group. *Spine* 2004;29(4):421.
- (30) Johnell O. Economic implication of osteroporotic spine - disease cost to society. *Eur Spine J* 2003;12(Supl. 2).

- (31) Tosteson AN, Skinner J, Tosteson TD, Lurie JD, Andersson GB, Berven S, et al. The Cost Effectiveness of Surgical *Versus* Nonoperative Treatment for Lumbar Disc Herniation Over Two Years. Evidence From the Spine Patient Outcomes Research Trial (SPORT). *Spine* 2008;33(19):2108.
- (32) Mayskopf JA, Sullivan SD, Annemans L. Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices – Budget Impact Analysis. *Value in Health* 2007;10:336.
- (33) World Health Organisation (WHO). Constitution of the World Health Organization. Geneva: WHO; 1946.
- (34) Fanshel S, Bush J. A health status index and its application to health service outcomes. *Operations Research* 1970;18:1021-66.
- (35) Rosser RM, Watts VC. The measurement of hospital output. *Int J Epidemiol* 1972;1(4):361-8.
- (36) Sintonen H. An approach to measuring and valuing health states. *Soc Sci Med [Med Econ]* 1981 Jun;15(2):55-65.
- (37) Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res* 1982 Nov;30(6):1043-69.
- (38) EuroQol group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990 Dec;16(3):199-208.
- (39) Hawthorne G, Richardson J, Osborne R, McNeil H. The Australian Quality of Life (AQoL) instrument. Monash University Working Paper 1997.
- (40) Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *J Health Serv Res Policy* 1999 Jul;4(3):174-84.
- (41) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002 Mar;21(2):271-92.
- (42) Ware JE. SF-36 health survey: manual and interpretation guide. Nimrod Press: Boston; 1993.
- (43) National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal. London: NICE; 2004.
- (44) Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 1998 Nov;51(11):1115-28.
- (45) Kind P. The EuroQol instrument: an index of health-related quality of life. In: Spiker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2 ed. Philadelphia: Lippincott-Raven Press; 1996. p. 191-201.
- (46) Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. *Health Econ* 1996 Mar;5(2):141-54.
- (47) The MVH group. The measurement and valuation of health: Final report on the modelling of valuation tariffs. York: University of York; 1995.

- (48) Wittrup-Jensen KU, Lauridsen J, Gudex C, Pedersen KM. Generation of a Danish TTO value set for EQ-5D health states. *Scand J Public Health* 2009 Jul;37(5):459-66.
- (49) Sorensen J, Davidsen M, Gudex C, Pedersen KM, Bronnum-Hansen H. Danish EQ-5D population norms. *Scand J Public Health* 2009 Jul;37(5):467-74.
- (50) Sjøgaard R, Christensen FB, Videbæk T, Bunger CE, Christiansen T. Interchangeability of the EQ-5D and the SF-6D in long-lasting low back pain. *Value in Health* 2009;12(4):606-12.
- (51) Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004 Sep;13(9):873-84.
- (52) McDonough CM, Grove MR, Tosteson TD, Lurie JD, Hilibrand AS, Tosteson AN. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. *Qual Life Res* 2005 Jun;14(5):1321-32.
- (53) Hollingworth W, Deyo RA, Sullivan SD, Emerson SS, Gray DT, Jarvik JG. The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain. *Health Econ* 2002 Jan;11(1):71-85.
- (54) Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology (Oxford)* 2000 Jul;39(7):783-90.
- (55) van Hout B, Al M, Gordon G, Rutten F. Cost, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994;3(5):309-19.
- (56) Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ* 1998;7(8):723-40.
- (57) Briggs A, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technology Assessment* 1999;3(2).
- (58) O'Brien B, Briggs A. Analysis of uncertainties in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Statistical Methods in Medical Research* 2002;11(6):455-68.
- (59) Fenwick E, O'Brien B, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. *Health Econ* 2004;13(5):405-15.
- (60) Löthgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Econ* 2000;9(7):623-30.

Studies in Health Economics present the results of health economics research at Institute for Public Health, Health Economics, University of Southern Denmark. Professor Mickael Bech is editor of the series. He is professor of health economics and head of the department of Health Economics University of Southern Denmark.

Further information
Institute of Public Health
Department of Health Economics
University of Southern Denmark
J.B. Winsløvsvej 9, 1
DK-5000 Odense C
Telephone: +45 6550 3081
Fax: +45 6550 3880
email: hmj@sam.sdu.dk
ISBN nr.:978-87-89021-77-5