HEALTH ECONOMIC EVALUATION:

AN OPTION FOR OSTEOPATHY?

Master Thesis to obtain the degree of

Master of Science (Osteopathy)

at the **Donau Universität Krems**

presented by **Monika Perndl**

Vorchdorf, December 2007

Translated by Mag. Barbara Schnürch

Preface

When I was thinking about the topic of my master thesis I came again and again across the question of how osteopathy could create a place in the Austrian health care system. The effectiveness of osteopathic treatment has been proven time and again but nevertheless osteopathy is still not recognized in Austria.

My first thought was to compare a series of osteopathic treatment with a series of physical therapy and to evaluate and contrast the costs and the results of the treatments. However when I discussed this idea with my colleagues I realized that this project sounded simpler than it was in reality to carry out.

That said I could not let go of the thought of having osteopathy to be officially recognized in Austria. Thus I searched for another possibility to maybe help to promote this development.

I kept searching and some would say “accidentally”, I stumbled across a short article in the magazine “Physiopraxis”, which dealt with cost-utility analysis.

Finally, I had a key-word and was ready to immerse myself in the search of a concrete topic for my thesis.

I hope that my paper will provide a little glimpse into a (as I discovered) vast subject area and perhaps inspire some or the other not only to provide proof that that our work is effective but also to show that it is “worth” the money.

1 Introduction

In August 2007 the new Austrian Minister for Health, Family and Youth, Andrea Kdolsky, caught the public's interest with a spectacular announcement: the Austrian health care system had to reduce costs by almost 3 billion Euros! At that time she could not explain how this could be realized, but she insisted that it was the amount to be saved. It was tough talk and it was not enthusiastically received. The consequence reignited the heated discussion about costs in the health care system. While everybody is talking about reducing costs in the health care system, osteopathy strives for official recognition in Austria. The question is how can a policy that is focused on "reducing costs" be persuaded to accept an additional cost factor? The only way is if osteopathy proves that it works "economically". Meaning we must prove an osteopathic treatment incurs costs but when compared with other already recognized therapy methods it is cheaper or at least not more expensive than the forms of therapy which have been financed by the statutory health insurance providers so far. Health economic evaluation studies could show whether this is really the fact. To provide the proof of osteopathy's cost-effectiveness is definitely a major task, but is it also a manageable task?

For more than three decades so-called health economists have brooded over possible ways to make this proof easier. The results of their research (which are not uncontroversial) are among other things: health-technology-assessment, cost-effectiveness-analyses (CEA), cost-utility-analyses (CUA), cost-benefit-analyses (CBA), etc…

These different forms of analysis should help to evaluate whether a medical intervention is "worth" the money and whether it really "pays off" to cover a patient's costs.

The question is how can the above mentioned forms of analysis be applied to osteopathy?

In some European countries there are institutes which are responsible for health economic evaluations. In Great Britain, for instance, it is the "National Institute for Clinical Excellence" (NICE), in Germany the "Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen" (IQWiG). In the Netherlands there is the Euroqol Group (European Organization for "Health Related Quality of Life"), which thoroughly deals with the various aspects of Health-Technology-Assessments. Since July 3, 2007 Austria has the "Bundesinstitut für Qualität im Gesundheitswesen" (Federal Institute for Quality in the Health Care System) which has the

task to integrate health economic studies in the Austrian health care system. In Austria there exists furthermore a Ludwig Boltzmann institute for Health Technology Assessment (http://hta.lbg.ac.at/de/index.php).
Below you will find an excerpt on the integration of HTA (Heath Technology Assessment cf. 2.3) into the Austrian health care system from the institute's website (www.biqg.org):

"*The progressive development of new procedures and technologies in the health care system also demand decisions about how the catalogue of publicly financed services can be organized. In many countries regional or national institutions commissioned with HTA have been established to help the political decision makers.*

In Austria the institutionalized implementation of HTA is still in the set-up phase. But experiences of other countries' institutions can be of use in the integration of HTA in the Austrian health care system.

In the first project phase the organizational structures and procedures of already established HTA institutions of other countries (Canada, Great Britain, the Netherlands, Germany, Denmark) will be evaluated to draw conclusions for the Austrian health care system.

Within the framework of an analysis of the current situation the implemented and planned activities in the field of HTA of the federal authorities, provinces, funds, hospital organizations, the Main Association of Austrian Social Security Institution, statutory health insurance providers etc. are recorded and described."

This means that in the future these forms of evaluation of medical interventions could be incorporated also in the Austrian health care system.

The cost-effectiveness analysis (CEA) and the cost-utility analysis (CUA) could be of importance for osteopathy because these are the most commonly used and also recommended forms of analysis (cf. chapter 6). With regard to the discussion about the official recognition of osteopathy as independent profession clinical studies accompanied by CEA/CUA could show that osteopathy is not only effective but also cost-effective. In other words, this would show if osteopathy is worth the money.

This may all sound quite simplified and it also does not comply with the original idea of osteopathy, but if you want to come to an arrangement with the health care system in your country, you cannot escape this development.

My work will focus mainly on cost-utility and cost-effectiveness analyses, provide some examples from the field of non-invasive treatment measures and hopefully stimulate a discussion.

2 Methodology

As mentioned in the introduction that an article from the magazine “Physiopraxis” (Mehrholz J; Studiendesign: Kosten-Nutzwert-Analyse; Physiopraxis 7-8/06) represented the starting point for my work.
With the first key word “cost-utility analysis” I searched several databases and websites for other references on the topic:

- Medline and PubMed
- OsteopathicResearch
- Ostmed
- Cochrane Libary
- EuroQol Database
- BMJ
- website of NICE
- website of IQWiG
- website of BIQG
- Google
- Wikipedia
- Healtheconomics.com and Healtheconomics.nl

At first, I searched with the following key words or combinations of key words:

Cost-Utility Analysis		Osteopathy
Cost-Effectiveness Analysis		Physiotherapy
Costs	AND	Rehabilitation
QALY		Spinal Manipulation
Quality of Life		Low back pain

This first search produced a number of references which I studied to expand my search with other key words and authors. Above all, the database searches helped me to find renowned textbooks on health economics. I chose to use the books that were most frequently quoted as references in the various articles to serve as basis for my work.

The search produced a huge amount of data thus I had to shift through hundreds of abstracts to filter out the articles that were relevant for the general section of my paper as well as those interesting for the osteopathic and therapeutic section.

In addition, I contacted Dr. Markus Narath (from the medical management of the KAGES (styrian hospital company)), who made some scripts and working papers available to me and helped me in dealing with some tricky questions, for which I am very grateful.

Another important basis for my paper is the 2001 diploma thesis of Axel Schrauder:
“Erstellen eines Methodenhandbuches: Die Evaluation des Faktors “Kosten“ bei vergleichenden Studien in der Osteopathie“. (“Development of a methodological compendium: The evaluation of the factor “cost” in comparative studies in osteopathy“)
The thesis was presented in 2001 at the “Osteopathie Colleg München” (Osteopathic College Munich).

At the very beginning I realized “health economic evaluations” is a specific, very comprehensive field of scientific research. A lot more time is necessary to read all the articles, papers and books available on this topic. Thus I would like to point out that the information in this paper was filtered by me.
It would go beyond the scope of this paper to analyse every aspect in detail. In addition, this paper is a master thesis in osteopathy and not in health economics.
Thus I have tried to present the information on health economics in a way and in an extent that it is also understandable and palatable for laypersons.

3 Fundamentals/definitions

Since many articles and studies use health economic terms as a matter of course, some of the most important and frequently used of these terms are presented and briefly explained:

3.1 Evidence-based medicine (EBM)

Many definitions with different wordings can be found for evidence-based medicine.
The term was first coined in the early 90s by Gordon Guyatt who belonged to a group with David Sackett at the McMaster University, Hamilton, Canada (www.wikipedia.org).
After reading through a number of definitions from different sources I decided to present the definition of the website of the German network for evidence-based medicine (Deutschen Netzwerkes für evidenzbasierte Medizin) because it is clear and comprehensive.

Evidence (lat. evidentia = obviousness) colloquially means something that is obvious, apparent, totally clear. "That is evident" thus means that something is no longer to be questioned or challenged.

In the context of evidence-based medicine the term evidence has a different meaning. The English term "evidence" (= testimony, proof, confirmation, result, demonstration) refers to information obtained through scientific studies and systematically gathered clinical experiences, which confirm or deny certain facts.

The term evidence-based medicine thus designates the conscientious, specific and reasonable use of the currently best external scientific evidence to make and justify decisions in medical care of individual patients.

The term evidence-based practice (EBP) is used as synonym for evidence-based medicine (EBM). Both terms mean the patient receives the best possible most proven medical treatment.
Evidence-based health care (EbHC) is a related term which includes other fields and aspects of the health care system, e.g. logistic and organisational aspects concerning processes within the health care system.

To practice of EBM means to integrate individual clinical expertise with the best possible external evidence from systematic research (adapted from Sackett at al. 1999). (DNEbM 2006 - http://www.ebm-netzwerk.de/grundlagen/definitionen/)

3.2 Quality of life

The term quality of life is widely used and commonly known. Everybody has his/her own image of what quality of life means to him/her. And that is exactly it: everybody has THEIR OWN definition.

In order to use the term quality of life in a scientific paper or in a health economic evaluation study (cf. 3.5) the term has to be clearly defined and its meaning explained. It is important for the reader to know which values were incorporated in the assessment of the quality of life and which were not considered and why.

In general, quality of life comprises several factors like degree of wealth, education, professional satisfaction, social status and also health (www.wikipedia.org).

The WHO (1993) defines quality of life as follows:

"Quality of life is the perception of individuals of their position in life, in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns." (www.drnawrocki.de/empfehlung/lebensqualität%20.html) (English source: http://whqlibdoc.who.int/bulletin/2001/issue11/bul-11-2001/79(11)1047-1055.pdf)

The term quality of life has its origin in welfare. Welfare is the "old term" for quality of life. But welfare did not only mean the accumulation of material goods but also the subjective well-being.

(www.irs-net.de/anzeigen.php?choice=monitoring&choice2=monitoring2)

This means that there are in fact two aspects of quality of life to be evaluated: one is objectively verifiable and deals with the amount of material goods someone disposes of and the other one is a person's subjective perception of well-being or how he/she feels.

Again I could continue to try to precisely define the terms well-being or feeling. But I will let them stand for themselves because again, this would exceed the scope of this paper.

The quality of how someone is feeling results from the person's own individual ideals, values and moral concepts. If someone attaches importance to a high social status, the person will feel an extreme decrease of well-being if he/she experiences a social downfall. Someone else might consider health as his/her "most precious" asset and thus feel a disturbance of his/her well-being and thus decrease in his/her quality of life in case of disease.

In a presentation on the topic Heinrich Schipperges draws the following conclusion:

"All experts agree that quality of life can hardly be defined comprehensively and it is even harder to quantify it. The new attitude that arises in this context is more important than any cataloguing, any measurement, it is a habit which refers to old values and is out for new valuations. Nevertheless, you cannot ignore that „quality of life" threatens to degenerate into a mere buzzword, a term that has become so fashionable that you have to first of all fill it with life before it starts to have a meaning. But who could fill such an actual anaemic term (and

this is my theory), who could fill it better with vibrant life than doctors and their art of medicine! Quality of life has less to do with wealth than with being well, feeling good, being healthy.

The sciences of health basically are always sciences of quality." (www.klinik-angermühle.de/buch/buchband/a-band5a.htm; line 36-44)

It has already been mentioned above that it is important to define the term quality of life as precisely as possible with regard to the context in which it is used and to justify why certain aspects are included or not.

In the context of health economics (cf. 3.4) another term derived from the umbrella term quality of life is used: **H**ealth **R**elated **Q**uality **o**f **L**ife (= HRQoL).

3.2.1 Health Related Quality of Life = HRQoL

According to the WHO definition health is not only the absence of disease but *"a state of complete physical, mental and social well-being"*. Thus we realize that health actually consists of three components: a physical, mental and social component. (Lauterbach et al. 2006)

As the term implies the health related quality of life concerns mainly the physical health aspect of quality of life. The aim is to make health related quality of life measurable or quantifiable. Chapter 6 will look at the evaluation methods and the advantages of this assessment in more detail.

3.3 Health Technology Assessment

Health Technology Assessment (HTA) or Medical Technology Assessment (MTA) is a form of Technology Assessments (TA). The term is sometimes used as buzzword by the media and thus it will briefly be explained in the context of this paper.

TA analyses technologies with regard to their effect (physical, biological, social, financial, medical). (Schwartz et Dörning 1992)

In the context of health care HTA specifically looks at the effects of the application of medical technologies and compares them with regard to their effectiveness, efficacy, costs, ethical harm_(lessens) and legality (www.healtheconomics.nl). The difference between effectiveness and efficacy will be discussed in more detail in chapter 3.7.

3.4 Health economics

Health economics is a sub-discipline of economics. It links medicine with economics. With the aid of comparative studies, it analyses various interventions with regard to their costs and effectiveness and afterwards evaluates which intervention is more cost-effective. It focuses directly on actual interventions and due to its complexity it is a field of interdisciplinary research.

The science of health economics has significantly progressed in recent years. Although a lot of disagreement among health economists existed in the early nineties, over the recent years more consensus on methodology has been reached. This development resulted in a standardized approach, allowing valid comparisons of studies in different fields of health care and across countries. (Lauterbach et al 2006, www.healtheconomics.nl/Health_Economics)

3.5 Health economic evaluation

A health economic evaluation is a comparative analysis of alternative health care interventions with regard to their costs and benefits (Drummond et al, 1997).

The various forms of health economic evaluations will be explained in chapter 6 in more detail.

The only thing that should be pointed out in this context is that there are different forms of health economic evaluations. The two most commonly applied forms are the cost-effectiveness analysis (CEA) and the cost-utility analysis (CUA).

The main difference between them is that in a CEA the costs and consequences of alternative interventions are expressed as cost per unit of health outcome, while in a CUA the consequences of interventions are expressed as 'utilities'. Such measurable units are e.g. the number of lives saved, the number of years gained, successfully treated diseases, the number of working days gained, etc. A ‘utility’ is a figure between 0 and 1, with 0 meaning death and 1 perfect health.

3.6 Utility

In the habitual use of language everybody has a different concept of the term “utility“. To use this term in a scientific paper it has to be defined as precisely and unambiguously as possible. It should always be clear for the reader how “utility” is to be understood in the specific case and also how it was evaluated. Chapter 5 on the evaluation of utility and effectiveness will discuss this in further detail.

It is important to point out hat in health economic evaluations the terms utility and utility weights are only used within the framework of cost-utility analyses (CUA), while in cost-effectiveness analyses (CEA) the terms effectiveness and efficacy are used (cf. 3.7 on effectiveness/efficacy). CUAs and CEAs are the most common forms of health economic evaluations. They will be described in more detail in chapter 6.

3.7 Effectiveness – Efficacy – Efficiency

The terms 'effectiveness', 'efficacy' and 'efficiency' are often used as synonyms. Occasionally this is also the case in health economic evaluations.
Many authors, however, do differentiate between these terms even though their meanings only differ marginally from each other.
Above all it has to be mentioned that there is a difference to the term 'utility' in the context of health economic analyses. While the terms ‚effectiveness' and 'efficacy' are used in cost-effectiveness analyses (CEA), the terms 'utility' and 'utility weights' are applied in cost-utility analyses (CUA).
When I researched the differences of these terms on the World Wide Web I found the following distinguishing features for 'effectiveness', 'efficacy' and 'efficiency':

Encarta (http://encarta.msn.com/dictionary/effectiveness.html) defines the three terms as follows:
"***effective****: causing result, especially the desired or intended result*
efficacious*: having the power to achieve a desired result, especially an improvement*
efficient*: capable of achieving the desired result with the minimum use of resources, time and effort"*

Other sources use the following definitions in the context of clinical results:
The **'effectiveness'** of a clinical intervention or a medication is the positive effect the intervention or the medication has in a real world setting.
The **'efficacy'** of a clinical intervention or a medication designates the effects that can be observed under controlled circumstances, e.g. in a clinical study. (http://dict.leo.org/forum, www.healthecnomics.nl)
In contrast to this **'efficiency'** is understood in terms of profitability. It describes how many resources have to be used to achieve the set goal. Thus 'efficiency' should rather be used as synonym for 'cost-effectiveness'. (http://dict.leo.org/forum)
In summary it can be said that within the framework of health economic analyses the terms 'effectiveness' and 'efficacy' relate to the true effects of an intervention while the term

‘efficiency’ looks rather at the economic aspect and thus can be used along the lines of ‘cost-effectiveness’.

3.7.1 Cost-effectiveness

The term cost-effectiveness is one of the most frequently used terms in health economic studies. Thus it is indispensable to explain its meaning. It has already been indicated in chapter 3.7 that the cost-effectiveness relates to the profitability of a medical intervention. Are the costs justifiable with regard to the intervention’s outcomes? If treatment A is cheaper than treatment B and in addition has the better results a good cost-effectiveness can be observed. If, however, treatment A provides the better results but is more expensive than treatment B, it has to be evaluated whether the additional costs are balanced by the positive results. To do so a cost-effectiveness ratio is calculated.
This is exactly what health economic evaluations look at. (cf. chapter 6)
The term ‘effectiveness’ alone has a different meaning (cf. 3.7). Thus attention has to be paid to clearly differentiate the terms. Some authors avoid confusion by using the term ‘efficacy’ when they mean the results or benefits and the term ‘cost-effectiveness’ when they want to point out the profitability.

Doublilet, Weinstein et Mc Neil, (1986) argue that cost-effectiveness is given if a program with the same effects costs less than the others or if one program is more expensive than the others but the higher costs can be justified by much better outcomes. In turn, programs with a lower efficacy but also clearly lower costs can be regarded as cost-effective.

3.8 Human-capital approach

The human-capital approach is based on the assumption that “healthy time” which someone gains through a successful medical treatment can be equated with cash value. For the calculation the current market values of labour time are used. This usually corresponds to a person’s earned income.
The human-capital approach can be applied in two respects. First, to attribute a certain value to specific cost parameters (e.g. time spent for a specific treatment, cf. also 4.1) and second, to express the total effects of an intervention in cash value, like it is done in cost-benefit analyses (CBA, cf. 6.3.1). (Drummond et al. 2005)
Since many groups of persons (retirees, children, housewives and housemen, unemployed persons, …) cannot be included in this approach, it can be assumed that this approach also gives rise to some criticism. (Lauterbach et al. 2006)

4 Costs

If costs are mentioned within the framework of a health economic evaluation, these costs usually describe the total costs that arise in the context of an intervention. This includes the treatment costs, expenditures by the patients and their families, consumption of resources in other areas and productivity changes. (Drummond et al. 2005)
In this context, it is indispensable to describe and explain the cost evaluation as precisely as possible. Are the costs incurred by the health insurance? Does the patient have to bear the costs? Which costs are included in the evaluation and which are not? Why some are cost factors not considered? Which costs are included in the analysis usually depends on the perspective of the respective study. This perspective also determines the point of view from which a health economic analysis is carried out (e.g. from a societal point of view, from the point of view of the health care system, from the point of view of health insurance providers,...).
For instance, the transport costs to the place where the treatment is carried out play a role if the analysis is carried out from the point of view of the patient or the general public but not if the analysis relates to the perspective of a country's Ministry for health. Chapter 6.6 will describe the various perspectives and their pros and cons in more detail.

In health economic analyses the costs are evaluated in three steps: (Lauterbach et al., 2006)

- all relevant resources whose consumption is affected through an intervention are identified
- the resource consumption is quantified
- costs of the previously identified resources are defined

Cf. Chapter 4.4 concerning the topic cost evaluation.

Chapters 4.1 and 4.2 below first look at direct and indirect costs that can arise:

4.1 Direct costs

The term "direct costs" usually implies that money is directly transferred but this is not the case in this context. The term rather refers to the resources that are immediately consumed within the framework of an intervention (time, money, personnel,…). (Lauterbach et al., 2006)
It has to be pointed out that when studies are compared attention is paid to the fact that costs are always immediately defined and written in the formula, because e.g. sometimes a patient's waiting time is qualified not as direct but indirect costs.

The US Panel on cost-effectiveness in Health and Medicine[1] recommends expressing the time deployed by the patient in money units. This recommendation is based on the fact that it is more difficult to include the deployed time in the evaluation of the decrease of quality of life than to express it as costs, for instance, in the context of the human-capital approach (described in chapter 3.8). If approximately the same time has to be deployed for the two alternative treatments that are evaluated, this factor can be neglected in the cost evaluation because it does not have an influence on the result. (Lauterbach et al., 2006)
What is important, however, is that the exclusion of certain cost factors has to be justified with valid arguments.

4.1.1 Direct medical costs

The direct medical costs represent the most obvious cost factor in health economic evaluations. They indicate the direct use of resources that are necessary for an intervention. This includes the labour force of the necessary personnel (doctors, therapists, hospital nurses,…), costs for clinical tests, medication, appliances, medical equipment and consumable material. (Lauterbach et al., 2006)

4.1.2 Direct non-medical costs (external costs)

The direct non-medical costs include, for instance, costs for child care for the time the mother is treated, transport costs even if the patients use their private cars or public transport, the time consumed by the transport,…(Lauterbach et al., 2006)

4.2 Indirect costs

Indirect costs are also called productivity costs because they represent the productivity loss due to illness (Gold et al., 1996). Either someone cannot live up to his/her normal performance or he/she is on sick-leave and thus does not work at all for a certain period of time.
With regard to society in general the indirect costs become virtual costs from the moment openings are filled with unemployed persons. (Zink A, 2004)
The loss of leisure-time pleasures can also be counted towards the indirect costs. This means, for instance, that due to a disease someone can no longer enjoy an activity or hobby that was important to him/her. (Lauterbach et al., 2006)

[1] The US Panel on cost-effectiveness in Health and Medicine was convened by the U.S. Public Health Service in 1993 and was co-chaired by Louise B. Russell and Milton C. Weinstein

The US Panel, however, recommends that the loss of leisure time (pleasures) be counted towards the loss of quality of life and that it should not be regarded as indirect costs.
The indirect costs can be divided into morbidity costs and mortality costs (Lauterbach et al., 2006)

4.2.1 Morbidity costs

Morbidity costs are caused by health impairments. This means a reduced ability to work, and limitations with regard to leisure time activities,… However, limitations of leisure time activities should rather be regarded as a reduction of quality of life.
(Lauterbach et al., 2006)

4.2.2 Mortality costs

Mortality costs describe the years of life lost through disease. They can be included in the analysis either as monetary loss or as reduction in utility or effectiveness. The only thing that has to be made sure is that they are not included twice, i.e. either they have to be included among the costs or they have to be counted towards the effectiveness or utility. In general, it is recommended that mortality costs should not be attributed a monetary value. (Lauterbach et al. 2006)

4.3 Opportunity costs

The opportunity costs do not play a role as cost factor as such but in the context of cost evaluation (cf. 4.4) the principle of ‘opportunity costs’ can be important with regard to the attribution of costs for time deployed.
Opportunity costs are also called alternative costs because they also express how the expended resources could have been used alternatively.
One example of where the opportunity costs are important would be the assessment of indirect costs due to production downtimes of a sick employee.
The money that has to be spent to cover the costs which arise in this context cannot be used for other things (“opportunities”), even though it possibly would have produced a better utility to use the money for something else. The monetary value of the utility loss is called opportunity costs. (Lauterbach et al. 2006)

4.4 Cost evaluation

The chapter cost evaluation deals with a quite extensive topic. Thus it will be discussed with the utmost accuracy. It has already been mentioned that it is important to justify why certain cost factors have not been included in the analysis. If the perspective of a specific study (cf. 6.6) serves as justification, the choice of perspective in turn has to be justified.

First, the expended resources are identified. After that the cost evaluation consists of two elements: the quantitative evaluation of the resources and the attribution of costs per unit. (Lauterbach et al. 2006, Drummond et al. 2005)

The **quantitative evaluation** of resources can for example be carried out on the basis of study records (in the case of analyses within the framework of clinical studies). Also data systems (e.g. in hospitals) or the case histories of patients can serve as a basis for this evaluation. Resources that were expended in the domestic context can be established by means of questionnaires or in the form of expenditure diaries.

The **attribution of costs /market prices** is possible for many resources (price per appointment with the doctor, price per x-ray image, price per medication,…). In this context it could be argued that the actual cost of a resource corresponds to the opportunity costs (cf. 4.3) but the attribution of existing market prices should be chosen as pragmatic approach. (Drummond et al. 2005)

If an approach via opportunity costs was followed, it could happen that the same amount of money could have been spent for something else, which could have produced more utility/effectiveness and this alternative could thus seem to be more valuable. But since that in turn would have to be proven in order to confirm this assumption, the attribution of prices/costs would become much more complicated. This is probably the reason why Drummond et al. (2005) suggest the attribution of market prices.

Another problem is posed by the assessment of ‘volunteer time’ and ‘patient/family leisure time’. In the case of ‘volunteer time’ there is the suggestion to use ‘unskilled wage rates’ but the market value of leisure time is much harder to assess. There exist several suggestions to overcome this problem: some say leisure time should be attributed the value 0, some suggest to use average wages as a guideline, and others argue to use double wages paid for overtime as the basis for attributing a price to leisure time because also the employer has to pay double for the free time of the employee. (Drummond et al. 2005)

Brouwer et al. (2001) argue that even before worrying about the value of time, it should be evaluated how much time volunteers and relatives really sacrifice and then it should be assessed whether this factor could influence the result considerably.

In the literature the common practice is to attribute the value 0 to the time factor, especially in analyses which are carried out from the perspective of decision makers. (Drummond et al. 2005)
In the case of a 'societal perspective' Gold et al. (1996) recommend to include this cost factor in any case.

Based on these different arguments and suggestions Drummond et al. (2005) propose that the time of volunteers and relatives could be documented separately in time units and that these records could be analysed parallel to the cost evaluation. It can still be decided afterwards whether the result would be considerably influenced by the efforts of volunteers and relatives.

4.5 Discussion

The most important thing in cost evaluations is a good transparency when the data concerning costs are collected. The methodology and the origin of the estimated values and costs should be clearly recognizable. The choice of perspective (cf. 6.6) and possibly factors excluded due to that should be well substantiated with good arguments.
The attribution of specific amounts of money to time deployed could potentially cause discussions not only regarding the time deployed by volunteers and relatives (as discussed in chapter 4.4) but also in the context of osteopathy in Austria.
If a certain money value is attributed to the time of a physical therapist or an osteopath, this would mean to make a generalization about the skills of the individual therapists. Differences are not considered (e.g. how much the physical therapist or osteopath has invested in his/her additional training in order to be more 'efficient' (cf. 3.7)).
The majority of osteopaths in Austria are physical therapists in their original profession. As physical therapists they can charge their fees directly to the statutory health insurance providers (at least under an optional settlement contract), so that the patients only have to pay part of the fees out of their own pocket. The time, however, is regarded as the time of a physical therapist, no matter whether the therapist just qualified from the Academy for Physical Therapy Training or e.g. has undergone an additional osteopathic training.
In the context of a health economic evaluation a question regarding studies that are carried out in Austria is, whether to use the prices estimated on average by the statutory health insurance providers or to ask a representative number of osteopaths how much they charge for one treatment and to calculate the mean value. Another question is, how the choice of method affects the choice of perspective. In any case it is important to make sure that the same kind of cost evaluation is used for the two treatment alternatives that are to be

compared, i.e. one must not use the average price of the therapists in the case of osteopathic treatment and the average rate of the statutory health insurance providers in the case of physical therapy.
In addition, therapists with a similar skill level have to be chosen. E.g. if a series of osteopathic treatments is to be compared with a series of physical therapy treatments in the case of patients with lumbar pain, it is recommendable to choose physical therapists who also have an additional training (e.g. in manual therapy concepts like Maitland or Cyriax etc.). Osteopaths in Austria are graduates from a physical therapy academy or graduate medical students, who underwent additional six to six-and-a-half years of osteopathic training. Thus the credibility of the study would suffer if the work of osteopaths would be compared with that of physical therapists who just graduated from the PT Academy.

Regarding the prices of medications the market values can be distorted because of price-fixing agreements between the pharmacology companies and the government. Also this should be considered and if necessary corrected in the attribution of values. (Drummond et al. 2005)
In this context the question is, however, in how far you can gain an insight in these mechanisms?

In summary, at first glance the topic cost evaluation seemed much less complex than it is in reality. Nevertheless, with a meticulous and consistent methodology regarding the analysis and a good argumentation, a valid cost evaluation should be possible.

5 Effectiveness and utility

It has already been mentioned in chapters 3.6 and 3.7 that many things can be said about how the terms effectiveness and utility are understood. This means that if utility and effectiveness are to be evaluated, a thorough methodology is necessary and the choice of evaluation method has to be clearly justified and explained (just like it is the case for cost evaluation).
Will the utility or effectiveness be measured from the point of view of the statutory health insurance providers? Do you look at the consequences for society or for the individual person?
Clinical osteopathic studies usually examine how the treatment affects the individual patient or test person. Exactly these results should be included in the evaluation of the effects.

When the clinical outcomes of an intervention are assessed a health economic analysis that adheres to a consistent terminology will either use the terms ‘utility’ and ‘utility measures’ in the case of a CUA or the terms ‘effectiveness’ and ‘efficacy’ in the case of a CEA.
It can also be that only general, uncontroversial terms like ‘consequences’ or ‘outcome measures’ are used.
In the German speaking literature the term ‘Nutzen’ (benefit) is often used in the sense of the general clinical outcome. All of this should be considered when reading primarily English or primarily German literature.

Regardless of which form of health economic analysis is applied, the evaluation of the clinical outcomes should consider both the intended and also the unintended effects of an intervention.
“These effects can include disease-related or treatment-related changes or the influence by the following target parameters:

1. *mortality*
2. *morbidity (complaints and complications)*
3. *HRQoL (Health Realted Quality of Life)*
4. *intervention-related and disease-related expenditures*
5. *patient satisfaction*

A (from the patient’s perspective) positive change of these aspects is defined as direct patient-related medical benefit, a negative change is regarded as direct patient-related medical damage.”
(IQWiG 2006, p. 38 Z 22-28 and p. 39 Z 1-3)

Clinical results can be divided into three groups: hard, soft and surrogate data. Hard data are e.g. disease remission, avoided organ failure or avoided complications. In contrast to this soft data are e.g. changes in the quality of life (cf. 3.2) or in the functional status. Changes in the blood pressure, glycosylated haemoglobin and similar things belong to the surrogate results.
In the evaluation of the outcomes of an osteopathic treatment the soft data are probably most relevant. These include parameters like pain, daily activities, mobility and also emotional or mental impairments. (Maetzel, 2004)

5.1 Effectiveness – Efficacy

Health economic studies can concern the effectiveness/efficacy or utility of an intervention. The difference is that effectiveness is measured as cost per unit of health outcome, while the utility is expressed in utility weights (cf. 5.2). The natural units in the case of effectiveness analysis are usually one-dimensional (lives saved, successfully treated diseases, reduced stay in hospital, working days won, lowered blood pressure in mmHg,…), while utility values usually are composed of several dimensions (cf. 5.2). (Drummond et al 2005)
This would be the theory! In practice, however, it happens time and again that effectiveness and utility analyses are mixed so that health economic studies are called cost-effectiveness studies even though the clinical results are assessed with methods which strictly speaking are methods to assess the utility.
One reason for this can also be that in contrast to e.g. Drummond et al. (2005) Gold et al. (1996) present the cost-utility analysis as subdivision of the cost-effectiveness analysis and thus do not draw such a clear boundary.

5.2 Utility

It has already been mentioned in chapter 5.1 that the evaluation of utility comprises several dimensions which have an influence on the quality of life.
In this context psychometric (multi-attribute health status classification systems with preference scores) and utility-theoretical (measuring preferences) methods can be differentiated.
The psychometric methods are represented by questionnaires, while the utility-theoretical methods include interviews, where the interviewees are asked to assess their state of health under different circumstances. (Lauterbach et al. 2006).
The result of a utility analysis is expressed as utility measures. Utility measures are represented on a scale from 0 to 1, where 0 means immediate death and 1 perfect health.

(www.healtheconomics.nl) Despite their “soft” character the utility measures are regarded as precondition to be able to compare health economic aspects. (Richardson 1997)
Health economists have agreed that the evaluation of utilities or utility measures best fulfil the purpose of the non-monetary evaluation of clinical outcomes. (Gold et al 1996)

5.2.1 Utility-theoretical methods

5.2.1.1 Standard Gamble

Today the “Standard Gamble” developed by Neumann and Morgenstern is regarded as gold standard in the evaluation of utility measures. It is the only method that is based on a sound theoretical model.
In this form of interview the patients are asked several times to decide whether they opt for their present sate of health or whether they want to participate in an imaginary game which with a certain probability will lead to complete recovery but with the complementary probability to unavoidable death. The proportion between death and recovery is changed with every repetition of the game until the patient has difficulties to decide which option to choose.
The value at which this is the case is regarded as the “optimum health equivalent” for the present condition.
On a 0-to-1 scale a threshold of 80% probability of death to 20% probability of recovery would correspond to the value 0.8.
The standard gamble method is considered as good method because the patient is forced to make a decision in a risk-laden situation which is similar to the situation, where he/she has to decide in favour or against a medical intervention.
Nevertheless, this method does not meet with a lot of acceptance among study supervisors and participants, on the one hand, because a lot of effort is necessary to collect the data and, on the other hand, because the direct confrontation with death is regarded very critically. (Maetzel A, 2004).

5.2.1.2 Time-Trade-Off

The TTO interview developed by Torrance offers an alternative to the above mentioned standard gamble.
The patients are repeatedly confronted with the choice of spending the rest of their lives with the present disease or living shorter but in perfect health. The decisive moment is the moment in which the patient has difficulties to make this decision. The duration of life (in years) with the disease is then compared and contrasted with the reduced life expectancy to

calculate the utility measure. Example: Life expectancy with disease: 25 years; in perfect health: 20 years. If 20 is divided by 25 the resulting value is 0.8.

With this form of interview it is easier to determine utility measures because the questions can be answered more intuitively and thus are cognitively less demanding.

In contrast to the all-or-nothing principle of the standard gamble the result of the TTO interview is save (no threat to life). Thus the TTO utility measures are slightly lower than the standard gamble utility measures. (Maetzel A, 2004)

5.2.1.3 Rating Scale

This is the simplest method to directly assess the utility but it is also the method that is the least substantiated. (Naylor CD et Llyewellyn-Thomas HA, 1998).

The patients are asked to rate their state of health on a 0-to-100 scale, with 0 equalling death and 100 representing perfect health. Since the patients neither have to decide between life and death like in the standard gamble method nor sacrifice a number of years of their life like in the TTO interview, the utility measures are even lower than in the above mentioned approaches because without pressure patients tend to rate their state of health even worse. (Maetzel A, 2004)

5.2.2 Psychometric methods

Psycometric methods are also regarded as indirect procedures to assess the utility. In contrast to the utility-theoretical (direct) evaluation methods they offer the advantage that they do not require so much personnel as needed in the interviews.

The advantage of an interview is that the patient can very well bring in his/her personal problem, while a questionnaire offers only a limited choice of health aspects. (Maetzel A, 2004)

We differentiate between generic and disease-specific questionnaires. The latter are adapted to a large number of different conditions (psoriasis, angina pectoris,…) where the questionnaires take the different characteristics of the condition into account. (Lauterbach et al., 2006)

This chapter will only present the most commonly used generic questionnaires (Maetzel 2004; Drummond et al 2005) because a description of different disease-specific questionnaires would simply go beyond the scope of this paper.

5.2.2.1 HUI-2 und HUI-3

Two of the oldest questionnaires are the Health-Utility Index 2 and 3. It comprises a ‘health classification system’ and a ‘utility scoring formula’. HUI 3 is the most recent version. The HUI 3’s ‚health classification system’ includes eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion and cognition. Every question has a scale of 5 to 6 degrees of impairment. The HUI 2 version differs in three aspects, which can be useful for specific questions: self-care, emotion with a focus on anxiety/worry and fertility. (Furlong WJ et al., 2001; Horsman et al 2003; Drummond et al. 2005)

The ‘scoring formula’ helps to attribute the utility values of 0 for dead to 1 for ‘perfect health’.

The English versions of HUI 2 and HUI 3 can be found in the annex (13.2).

5.2.2.2 EQ-5D

The EuroQol Group developed this questionnaire. Thus the name of the questionnaire: EQ stands for EuoQol Group and 5D for the five dimensions the questionnaire wants to explore. There is one question for each dimension, which can be rated with three degrees of impairment. The five dimensions are: mobility, self-care, usual activities, pain and anxiety/depression. The patients can choose between ‘no problems’, ‘some problems’ or ‘major problems’. (Essink-bot et al. 1993; Brooks 1996; Kind 1996)

This results in 243 different possible states of health. By means of a ‘scoring formula’ they can be translated into a utility value between 0 and 1. Also this questionnaire and its scoring formula can be found in the annex (13.2). (Drummond et al. 2005)

5.2.2.3 SF-6D

The Short Form 6D questionnaire (6D stands for 6 dimensions) is the short version of the SF-36 questionnaire, which evaluates the quality of life on the basis of 36 questions. The SF-6D summarizes 11 questions of the SF-36 into six dimensions: physical functioning, role functioning, social functioning, pain, mental health and vitality. The possible answers within these dimensions offer four to six degrees of impairment (Brazier J et al 1998, Brazier J et al 2002). With the relevant ‘scoring model’ a utility value between 0 and 1 can be calculated for the result of the questionnaire. (Drummond et al. 2005)

Also this questionnaire and its ‘scoring model’ can be found in the annex (13.2).

5.2.3 QALYs

The Quality Adjusted Life Year (QALY) is the unit which is most often used in the context of cost-utility analyses (CUA s. 6.2) to express the utility of an intervention.

Chapter 5.2 has already explained that the utility is expressed with a value between ‘0’ and ‘1’, with 0 meaning ‘death’ and 1 ‘perfect health’. QALYs describe life years in perfect health gained as a result of an intervention. They are also expressed as values between ‘0’ and ‘1’, with 0 again meaning ‘death’ and 1 ‘one year in perfect health’.

The QALYs are calculated by multiplying the ‘duration of survival’ with the ‘quality of life’, which has been determined before through one of the methods to calculate the utility.

The advantage of data expressed as QALYs is that with one value both the quantitative gain (reduction of mortality) and also the qualitative gain (reduction of morbidity) can be expressed. In the context of a CUA the results are expressed as costs per QALY, i.e. costs per life year in perfect health that could be gained. (Lauterbach et al. 2006, Schrauder , 2001 Breyer et al. 2005, www.healtheconomics.nl, Drummond et al. 2005)

A detailed description of how QALYs are determined can be found in the literature contained in the list of references, e.g. Drummond et al. 2005 or Breyer et al. 2005.

5.3 Discussion

The evaluation of utility or effectiveness can be discussed quite controversially. Regarding the effectiveness evaluation the main question to be asked is whether a usually one-dimensional clinical result can have the desired weighting to be of significance in a health economic study. What good is it if, for instance, a medication soothes back pain but produces stomach ache as side effect? If data concerning side effects or other accompanying changes are not explicitly collected, in theory the result of a study could be favourable for an intervention even though it has more negative side effects than beneficial effects.

In a multi-dimensional utility evaluation this problem is not so predominant. The disadvantage here is that the instruments to evaluate the utility are uniform and thus less tailored to the specific topic of the study. This could mean that possible disease-specific aspects which could be of decisive importance for the patient are not attributed enough importance because they are either not evaluated at all or because among all the questions of a questionnaire they only have a relatively small influence on the overall result.

The interview forms in utility evaluations have several advantages and disadvantages. First of all, the interviewer can relate to the interviewee and explain misunderstandings that might arise. Further, the patient can easily consider his/her personal problem and attribute a certain importance to it. Possible side effects which can affect the overall state of health or the

quality of life can also be included in the consideration. The downside of the interviews is that they require a lot of time and personnel. In addition, the personnel have to be well-trained to minimize the interviewer bias as pointed out by Krabbe PFM (2002). Another question is (e.g. in the standard gamble interview) whether the interviewees really are intellectually able to differentiate between a probability of death of 30% or 35%? How strongly is the decision influenced by the fact whether someone already had the experience of being at the verge of death? Can a 25-year-old patient with lumbar pain, who never was seriously ill before, really make such an important decision? And would he make the same decision in the morning without acute pain and in the evening after a long, exhausting day with severe pain?

Maetzel (2004) pointed out that the different utility-theoretical methods provide different results because the interviewees are confronted with a much more risky situation in the standard gamble interview than in an assessment by means of a Visual Analogue Scale, where they do not have to sacrifice life time or are faced with a certain probability of death. If utility values are calculated on the basis of such a utility evaluation and studies compared with each other, which used different methods to evaluate the utility, this can easily lead to misconceptions.

Also the utility evaluation based on questionnaires is not free of problems. The use of questionnaires is not so time-consuming and does not necessitate such a large number of personnel, but it has to be considered that the test persons usually have to fill in the questionnaires on their own which could produce problems. The question is whether every single participant can intellectually comprehend the different questions. Is the compliance of the participants good or do they answer the questions ‘in passing’? Will a sufficient number of questionnaires be returned to have enough data for evaluation? Of course, the quality of the data collected by means of questionnaires can be increased if more personnel are employed, but this again means that a large number of well-trained personnel are required. In a clinical study the weighting of disease-specific aspects can also not be influenced when the above mentioned questionnaires are used.

Several other aspects, which can have an influence on the result of effectiveness or utility evaluations, are discussed in the literature. Gudex et al. (1996) wonder about the influence of age, sex, smoking, marital status, economic activity, work experience, looking after ill people, reported experience of serious illness and geographic location. Krabbe PFM (2002) also looked at the influence of the weather and the seasons and discovered a considerable interviewer bias. In addition, Krabbe PFM (2002) wondered whether death can be regarded as ‚health state’ in the first place and whether survival has to be rated higher in any case than being dead. Richardson J (1997) discovered that his test persons on average presented a higher level of education that the average of the general public and that there existed ‘little variation’ in age, sex, socioeconomic status, previous health history and geographic location.

Given all these arguments it is easy to recognize how important it is to clearly define which tools are used and to justify this choice and make it transparent. It could also be important to collect data on some additional aspects like the age or the level of education of the study participants. When it comes to the description of the study population one should never collect too little data.

With regard to osteopathy this could imply that it would be good to look at what kind of persons usually come to an osteopath. It is possible that an osteopathic practice is frequented more often by patients of a higher social status than a statutory health insurance out-patient practice because in Austria the patients have to pay at least part of the osteopathic treatment themselves, while the treatment in the latter is free for everybody with social insurance coverage.

It has already been pointed out that it is important to carefully choose the instruments for a utility evaluation and to thoroughly research what kind of problems and catches can occur before the study is carried out.

6 Health economic evaluation

In order to identify the best possible way for osteopathy to gain a foothold in the science of health economics this chapter will explain the current evaluation methods. Particular emphasis will be put on the main forms of analysis: cost-effectiveness analysis and cost-utility analysis. For the sake of completeness the other forms of analysis will also be briefly explained.

In health economic evaluations the costs and effects of two or more treatment alternatives are assessed within the framework of a randomized controlled study (RCT) and then compared to find out which treatment approach is the most efficient or cost-effective (cf. 3.7 and 3.7.1). (Drummond et al. 2005; Lauterbach et al, 2006; Gold et al. 1996)
The evaluation of the costs is explained in chapter 4, the evaluation of effectiveness and utility in chapter 5.
The costs are observed during the whole study period, while the effects are evaluated at least at the beginning and at the end of the study period but usually also at certain moments defined in the study protocol. (Torrance, 1997)
This does not make sense in the case of placebo or sham interventions[2] because in this case no cost factor can be determined that could be compared and contrasted with the costs of the intervention. (Schrauder, 2001)
The choice of measuring instruments for the effects should be well-founded because there is a large number of different kinds of measuring instruments available and every instrument has its own particularities. (cf. 5.3). In addition, the choice of instrument or instruments has to be clearly comprehensible for the reader.
Further, all cost factors have to be considered or explicitly excluded. If certain factors are not included in the analysis, this has to be mentioned in the discussion.

6.1 Cost-effectiveness analysis (CEA)

The cost-effectiveness analysis is a form of analysis where two or more treatment alternatives are compared with each other with regard to their effects. The clinical results are

[2] placebo intervention = a treatment where inert medication is administered, thus the therapeutic effect is only feigned
sham intervention = a placebo intervention where no medication is administered

measured in **natural units** and are also described as ‘effectiveness’ or ‘efficacy’ (lives saved, successfully treated cases of disease, shortened stay in hospital, working days gained, lowered blood pressure in mmHg,…).
Then the costs and the effectiveness are put in relation to each other to obtain a ‘cost-effectiveness ratio’. The ‘cost-effectiveness ratios’ of the treatment alternatives are then put in relation to each other by means of an incremental analysis (cf. 6.5) to determine the most cost-effective among them. (Drummond et al., 2005; Lauterbach et al, 2006).

6.2 Cost-utility analysis (CUA)

In the cost-utility analysis the clinical results are expressed as **utilities**. A utility describes how much the patient gains or looses through an intervention. The value ‘0’ is attributed to death, while perfect health is represented by the value ‘1’ (cf. chapter 5.2).
This is the main but not the only difference between a cost-effectiveness and a cost-utility analysis. This is probably the reason why the CUA is described as subdivision of the CEA especially in the American literature which means that often there is no consistent nomenclature. In CEAs the ‘effectiveness’ is established by means of singular parameters and thus one-dimensional but often more specific for the problem that is studied. The utility in CUAs is established in more dimensions and thus easier to be related to the quality of life in general. (Drummond et al 2005)
In order to express the utility so-called “QALYs” (quality-adjusted life years; cf.5.2.3) are often used. They have become established as important measure in the context of utility evaluations and are used in many studies. QALYs are the product of the life expectancy multiplied by a value representing the quality of life. After costs and utility have been established, ‘ratios’ are formed like in CEAs. In the case of CUAs these ratios are called ‘cost-utility ratios’. By means of incremental analyses (cf. 6.5) they are compared with each other. (Lauterbach et al, 2006)
With the aid of CUAs different intervention strategies can be compared. This kind of analysis also lends itself particularly for questions related to quality of life, mortality and morbidity (IQWiG, 2005).

6.3 Other forms of analysis

6.3.1 Cost-benefit analysis (CBA)

In this form of economic evaluation also the utilities are expressed in monetary units. However, it is exactly the attribution of monetary values to clinical results or states of health

which represents the big difficulty in this form of analysis because health cannot be measured strictly in money values.
The so-called “human-capital approach” (3.10) is used to achieve this.

6.3.2 Cost of illness analysis

The cost of illness analysis aims to identify and measure the total costs attributable to a particular condition. In this context all arising expenses should possibly be considered. A more detailed discussion concerning the evaluation of costs can be found in chapter 4.
In such an analysis no comparisons between different diseases or interventions can be made. (Lauterbach et al, 2006)

6.3.3 Cost-minimisation analysis (CMA)

A cost-minimisation analysis serves to determine the more cost-effective treatment method for a disease or clinical condition out of two treatment methods which proved to be equally efficient in an equivalence study. (Lauterbach et al., 2006)

6.3.4 Outcome-maximization studies (OMS)

Outcome-maximization studies evaluate costs and effects. In this form of analysis the costs of all intervention groups are set at an equal level. The aim is to determine the maximum result that is possible with the set expenditure. Since specific amounts are invested in the health care system, it is important to know what the various intervention methods can achieve with a set budget. Since some interventions work more cost-effective with a higher number of patients, the individual interventions can be made comparable by bringing the costs to the same level. (Sutton, 1997)

6.4 Selection of a specific form of analysis

Since every form of analysis has its particularities, catches and debatable issues, one should carefully think about which form to choose.
Lauterbach et al (2006) recommend to prefer the CEA and the CUA over the CBA because it is ethically questionable to attribute a certain money value to utility.
Schrauder A (2001) recommends to choose the CUA for osteopathic studies because it looks at more dimensions and because at the beginning of a study neither effects nor costs can be compared.

Drummond et al. (1996) and the Hannoveraner Konsensus Gruppe (1999) recommend to provide an elaborate justification for the choice of analysis in any case, no matter what form of analysis is applied.

6.5 Incremental analysis

In the context of health economic analyses usually two possible scenarios are compared with each other. Each scenario is characterized by the variables cost and utility. Chapters 4 and 5 look at the terms “costs” and “utility/effectiveness” more in detail.

At the end, every health economic analysis provides the cost-utility ratio for each scenario.

An incremental analysis calculates the difference between the same parameters of the two scenarios, i.e. the costs of the first scenario are subtracted from the costs of the second scenario. The same is done with the utility of the first and second scenario. Depending on the kind of analysis these differences are used for further calculations.

In the case of health economic analyses the quotient of the two differences is calculated. The result is an incremental cost-utility ratio.

A positive quotient means that the more expensive intervention also provides the better utility, while a negative quotient means that the more beneficial scenario is also the cheaper intervention, which is the best precondition for a decision in favour of this intervention. (Lauterbach et al. 2006)

In other words, the incremental analysis determines the proportion of additional costs to additional utility. (IQWiG, 2005)

6.6 Perspective

The perspective determines the point of view from which a health economic analysis is carried out. This has a decisive influence on the cost and utility factors (or effectiveness factors) that are to be considered in the analysis.

Among others the following perspectives are possible: societal perspective, perspective of the statutory health insurance providers, health policy perspective (or the perspective of the government in general), or the perspective of hospitals, of employers, of the sponsors of the study, of patients or doctors. (Gold et al. 1996, Hannoveraner Konsensusgruppe 1999, Drummond et al. 2005)

6.6.1 Societal perspective

If a health economic analysis is carried out from a societal perspective, all changes in costs and utilities due to an intervention play a role. This perspective is the most comprehensive.
The societal perspective considers unwanted side effects, the costs, which the patients themselves or their relatives have to bear, as well as productivity losses.
It has to be mentioned also that the societal resources are limited and that money that is invested in health cannot be used for investments in other fields like education, research or culture. Therefore the results of health economic analyses can help to make decisions about the distribution of resources easier.
An advantage of the societal perspective is that from a very general perspective decisions can be made by those who neither win nor loose through this decision. They do not represent the interests of individual groups or investors and do not consider something as gain or profit that actually is a loss somewhere else.
The societal perspective is the perspective that is usually recommended for health economic analyses. Once all data have been collected for this perspective any other perspective can be calculated as well on their basis. (Gold MR et al. 1996, Lauterbach et al 2006, www.healtheconomics.nl 2007) When in doubt, Drummond et al. (2005) recommend to opt for the 'societal perspective' because it offers the broadest approach and is always relevant.

6.6.2 Health service perspective

The health service perspective includes the evaluation of data that are directly relevant for the health care system. This means that only those costs are included that are covered by the health care system. Costs like transport costs or all other indirect costs are not included in the analysis. (www.healtheconomics.nl 2007)

Since these two perspectives are those that are most frequently chosen they have been explained in more detail in this paper.
With regard to osteopathic studies it makes sense to choose the societal perspective because the cost factor for osteopathic treatment does not yet exist in the Austrian health care system, thus it cannot be attributed a certain value. In addition, it can be argued that that on the basis of an analysis from the societal perspective any other perspective can be calculated. In addition, the authors also avoid the criticism that something that could have been of importance was not considered.

6.7 Discounting

Since a study usually has a limited time frame but changes in costs and utilities can occur also after the study period the tool of discounting is used.
This important tool in health economic analyses is used to look at longer periods of time.
In the case of many diseases the costs and utilities can continue to change in the years after a study is completed. To assess these changes they are discounted.
Basically, discounting is the opposite of interest calculation. One could say that costs and utilities are assessed by discounting the interest.
The principle of discounting is that costs that arise earlier are more important than costs that arise in the years after the evaluated intervention period.
The same principle holds for the utility. (Lauterbach et al., 2006)
In general studies are carried out in a particular time frame. This time frame can be some months or also some years.
But in the case of health economic studies also the future costs and effects (after the end of the study) should be evaluated. Thus certain tools are necessary.
As stated before discounting is one of these tools. The thought behind this method is the following: money that I spend today e.g. to finance a treatment cannot be invested with profit elsewhere. If I underwent the treatment two years later, I could have at least gained the interest of the specific amount in the two years.
Drummond et al. (2005) follow a similar line of thoughts as can be seen in the example below where they try to illustrate the method of discounting in the case of interventions for heart problems:
"If two options for dealing with heart disease were (1) expanding funding for Coronary Artery Bypass Graft (CABG) and (2) a health education campaign to influence diet and lifestyle, we might expect option (1) to deliver benefits earlier. Therefore, if a positive rate of time preference were acknowledged, it would look more attractive, compared with the preventive option, that would otherwise be the case. …" (p 73, l 8-12)
In the beginning a preventive program is probably more expensive than a bypass operation because all potential candidates for a cardiac infarction have to be re-trained, while only those who truly have blocked vessels have to be operated.
From an economic point of view the question now is whether the money that is now invested in a prevention program could be regained in a few years through the decrease in the number of operations, or whether more operations than necessary could be financed with the funds if the money would be saved in the same time.

Through discounting the money that is necessary for a medical intervention is "reduced" which is the opposite of increased through interest. It considers the fact that 1000 Euros that

are spent today could be worth 1050 Euros (thus more) in one year's time. (Lauterbach et al. 2006)

The question is what percentage rate should be used for discounting.
Drummond et al (2005) present two different approaches:

1.) The government of a country (e.g. United Kingdom) determines a discount rate which applies to all projects within the public sector.
2.) A discount rate that is often mentioned in the literature is used. In the late 70s and early 80s many articles were published in the New England Journal of Medicine, where a discount rate of 5% was used. Such a standardized rate makes it easier to compare studies with each other.

The US Public Health Service Panel on Cost-Effectiveness in Health and Medicine argued that costs and consequences should be discounted at a rate consistent with the Shadow-Price-of-Capital Approach to evaluating public investments. (Gold et al. 1996)
"Under this approach, one first transforms the stream of program costs over time into the corresponding stream of consumption losses that would be induced by the foregone investment and consumption opportunities. Next, one transforms the stream of program benefits into the corresponding stream of consumption gains. Finally one discounts these streams to present value using the social rate of time preference (SRTP) – that is the rate at which the social decision maker is willing to trade off present for future Consumption" (Gold et al 1996, p 218, l 21-28)

Gold et al (1996) propose to adapt the generally valid discount rate always to the most recent state of knowledge. However, to make sure that existing studies can be compared, standardized rates should be used. They recommend a rate of 3% or alternatively the above mentioned 5% for a period of at least 10 years.

6.8 Confidence interval

The term confidence interval is a term applied in statistics. A confidence interval describes the probability with which a value lies within a specific fluctuation margin. Usually the probability is determined (e.g. 90 %) and the fluctuation margin is calculated symmetrically around the mean value. In contrast to exact parameter estimation the confidence interval allows to directly conclude the significance. A confidence interval that is too large indicates that the sample size is too small. (www.wikipedia.de, www.healtheconomics.nl)

6.9 Bootstrapping

In health economics the statistical method of bootstrapping is used to estimate the statistical uncertainty in the outcome of an analysis. This involves choosing random samples (= bootstrap sample) with replacement from a data set and analyzing each sample the same way.
(www.wikipedia.de, www.healtheconomics.nl, www.wiwi.uni-bielefeld.de/~wolf/lehre/ss07/statan/bootsrap.pdf)

6.10 Sensitivity analysis

Since health economic analyses are subject to various uncertainties, it is recommendable to carry out one or several sensitivity analyses to evaluate the robustness of the results of a CEA or CUA.
In sensitivity analyses variables are deliberately changed to see how much influence they have on the result. It thus makes sense to carry out a sensitivity analysis for all variables that are subject to fluctuations.
If not all "fluctuating" variables are evaluated by means of a sensitivity analysis, the authors of a study should justify why this was the case. One possible reason for omitting this analysis for one variable would be e.g. that in the literature there is proof of this variable's small influence on the end result.
Regarding the analysis the interval within which the variables are changed has to be determined. However, this is often the weakness of sensitivity analyses. Often no or hardly any reasons are presented for the choice of a specific interval. Most of the times just double or halve the value of the variable in question is used without really arguing why.
To determine a plausible interval you can either refer to the relevant literature or expert opinions or choose the limits of confidence intervals.

"The moral appears to be that if you do not shake your study too hard it is unlikely to fall apart." (Drummond et al.2005, p. 42, l.35-36)

Another criterion of quality is which kind of sensitivity analysis is chosen. There are one-way-analyses, multiway-analyses, scenario-analyses, threshold-analyses and probabilistic-sensitivity-analyses.
In the **one-way-analysis** only one variable is changed and its effect on the result is evaluated. This is the most frequently used form of sensitivity analysis.
Since it is more realistic that more than one variable change at the same time and thus have an influence on the result, the multiway-analysis is the better choice.

In the **multiway-analysis** several variables are changed at the same time in a pre-set and justified interval. Even if there are only a few uncertain variables a large number of combinations are possible, thus a considerable calculation effort can be necessary.

In order to limit the numerous possibilities of a multiway-analysis the method of **scenario-analysis** can be used. It is a special form of multiway-analysis which involves the creation of specific extreme scenarios. On the basis of a "base case" also the "best case" and "worst case" scenarios are constructed and subjected to a sensitivity analysis. Also scenarios that are important for the authors or seem to be reasonable can be included in this analysis.

The **threshold-analysis** is another form of sensitivity analysis. In this form of analysis a threshold value is determined which must not be exceeded. For example, this could be a cost-effectiveness ratio that determines whether an intervention will be paid or not.

In the threshold-analysis the variables continue to be changed until it is found out which combination of changes alters the final result so that the set threshold is exceeded.

The last variety of sensitivity analysis is the **probabilistic-analysis** (Monte-Carlo-Simulation)**.** It is used more and more often.

This method involves that "probabilistic distributions" are determined for the influencing parameters. This means that assumptions are made concerning the distribution of the results within the pre-set interval. Finally, random samples are created as examples. (Drummond et al 2005)

6.11 Summary and discussion

A health economic analysis which compares two or more treatment alternatives looks at the 'costs' on the one hand and the 'consequences', 'effects', or 'outcomes' on the other hand. Depending on the form of analysis either both or only one aspect are variable.

In the most common forms, the CEA and CUA (they are also most frequently recommended in the literature), both the 'costs' and the 'consequences' are unknown at first and need to be determined. The costs can be roughly divided into direct and indirect costs. Depending on the study design they have to be included more or less completely. Concerning the 'consequences' we differentiate between a one-dimensional and a multi-dimensional evaluation of utility. Within these categories there exist direct and indirect evaluation methods. On the basis of the results of the utility evaluation utility values (in many cases QALYs) between '0' and '1' are calculated in CUAs, with '0' representing death and '1' perfect health (1 QALY = 1 year in perfect health).

Cost-effectiveness ratios (C/E ratios) and cost-utility ratios (C/U ratios) are calculated on the basis of 'costs' and 'consequences'. By means of an incremental analysis these ratios are

put in relation with each other. Ideally this shows which treatment option is more cost-effective.

To make a health economic analysis as plausible and reliable as possible several other factors have to be considered.

First of all, the choice of perspective plays a decisive role for the collection of data and also for the significance of the results. An analysis from the ‘societal perspective’ comprises all possible parameters and presents the effects on society, while an analysis from the ‘health service perspective’ reflects the point of view of the health care system or politics and thus that of the decision makers.

Another aspect is that maybe you want to make a prognosis concerning the cost-effectiveness of an intervention beyond the study period. A way to do that is to use the method of discounting, which means that the costs and effects of the moment have a higher significance than the same costs/effects in the future. Thus they are “discounted” by a certain rate. There are different recommendations concerning this discount rate, usually a rate of 3% or 5% is suggested.

In addition, a health economic evaluation requires a statistical analysis, which in turn can be subject to uncertainties. Confidence intervals or non-parametric bootstrapping can help to determine these uncertainties.

Finally, not only the statistical analysis but also the health economic evaluation as such is subject to various uncertainties because of highly fluctuating variables. To evaluate the influence of such variables on the overall result and to identify “outliers” one ore several sensitivity analysis should be carried out. In such a sensitivity analysis the variables that are to be evaluated are altered within a set interval and their influence on the overall result is studied. There are various kinds of sensitivity analysis and usually there are also more than one variable that require a sensitivity analysis. The choice which sensitivity analysis is used and which variables are subjected to it (or not) needs to be well justified.

All of this suggests that a health economic analysis is not an easy undertaking and that many things have to be considered in this context. The following chapter will present a checklist which should help to read and evaluate health economic analyses.

7 Checklist for the evaluation of health economic studies

It has already been mentioned several times that a number of problems are involved in the execution of health economic studies. One difficulty is to carry out a study with the best possible quality and most unambiguous results. It is very difficult in particular for a layperson of health economics to evaluate the quality of the numerous publications.

Thus Drummond et al. (2005) have established a checklist in their book, which should help to evaluate the quality of economic evaluations. Already in 1996, Drummond had published a similar checklist together with other colleagues in the BMJ, but the checklist presented below is the more recent version.

In the literature, however, the 1996 checklist is often used as reference. (Drummond MF et al (1996): Guidelines for authors and peer reviewers of economic submissions to the BMJ; BMJ, 313:275-283).

The most recent version of the checklist (Drummond et al 2005) comprises ten main questions, which will be listed below with a brief explanation.

The complete checklist can be found in the annex 13.3.

1. Was a well-defined question posed in answerable form?

An explicatory introduction and a well defined problem definition are important with regard to a health economic evaluation because articles that are published in medical journals are also read by people who are not experts in health economics. Thus a clear introduction can facilitate a better understanding.

2. Was a comprehensive description of the competing alternatives given? (That is, can you tell who did what to whom, where and how often?)

The clinical study which is accompanied by a health economic evaluation has to be well explained or at least sufficient references have to be provided to emphasize that a good RCT served as a basis for the health economic evaluation.

3. Was the effectiveness of the programmes or services established?

To have good ground for a CUA or CEA similarly effective interventions have to be compared. It has already been pointed out under question 2 that a well documented RCT, whose results show that the alternatives that are to be compared are effective, is necessary for that.

4. Were all the important and relevant costs and consequences for each alternative identified?

In this context it is important to consider whether all ‘costs’ and ‘consequences’ have been evaluated with regard to the selected form of analysis. For example: which perspective was chosen and why? Were all costs that are relevant for this perspective collected? Do the authors justify why certain parameters were not considered? Is a decisive parameter missing? Etc.

The possible cost and effectiveness/utility parameters are explained in chapters 4 and 5. More detailed information on the perspective is presented in chapter 6.6.

5. Were costs and consequences measured accurately in appropriate physical units (for example, hours of nursing time, number of physician visits, lost work days, gained life-years)?

This question looks at how the collected data were measured. Was the time expended by the patient measured in units of time or money? Was the leisure time valued differently than working time? How were the efforts and expenses of relatives or volunteers valued? Did the authors consider whether an appointment at the doctor’s required more or less time than a visit at another therapist? Or did they use a set price per doctor’s or therapist’s appointment, and if yes, why?

6. Were costs and consequences valued credibly?

What prices were attributed to the individual services and what were they based on? This question wants to make sure that the values are based on common sources and were not just ‘estimates’.

7. Were costs and consequences adjusted for differential timing?

This question wants to clarify whether a discounting was carried out and if yes, with what rate and why? If not, why not? Chapter 6.7 of this paper looks at discounting in more detail.

8. Was an incremental analysis of costs and consequences of alternatives performed?

The calculation of incremental ratios shows how the treatment alternatives that are compared correlate with each other. Writing ‘costs’ and ‘consequences’ next to each other and comparing them is not sufficient for a health economic analysis. More details on incremental analysis can be found in chapter 6.5.

9. Was allowance made for uncertainty in the estimates of costs and consequences?

To obtain robust results possible uncertainties have to be considered. The description of sensitivity analyses or other statistical methods (e.g. bootstrapping) is the proof of whether and how this was done. The various possibilities of sensitivity analysis are explained in chapter 6.10, while bootstrapping is described in chapter 6.9.

10. Did the presentation and discussion of study results include all issues of concern to users?

The final discussion and conclusion should present the results for the reader in a clear and concise way. In this context it is also important to pay attention to possible questions that the reader could have and to answer them or at least to mention them in order to point out possible deficits or shortcomings of the study. Another important aspect is in how far the results can be generalized. It might be that the results are only the consequence of the scenario constructed for the specific study design.

This checklist comprises all aspects (mentioned in the chapters 4-6) concerning the evaluation of the costs, the 'consequences' and the methodology in the context of health economic analysis.
The questions have to be unambiguous and the treatment alternatives have to be described in a clear and comprehensive way. The measurements of the costs and consequences have to be sufficiently and understandably explained and the choice of parameters and methods have to be well justified.
Further, discounting, incremental analysis and sensitivity analyses have to be carried out and the results should be presented in a way that the reader can recognize in how far the results can be generalized.

In addition to this checklist the literature offers a number of guidelines published by several groups of authors. Below you can find some examples:

- Weinstein MC et al. 1996; Recommendations of the Panel on Cost-Effectiveness in Health and Medicine;JAMA:276 p 1253-8
- Siegel JE et al. for the Panel on Cost-Effectiveness in Health and Medicine 1996; Recommendations for reporting cost-effectiveness analyses; JAMA: 276 p 1339-41
- Hannoveraner Konsensus Gruppe 1999; Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – revidierte Fassung des Hannoveraner Konsens; Gesundh.ökon.Qual.manag.:4 p A62-5
- Mason J et Drummond MF 1995; Reporting guidelines for economic studies; Health Economics: 4 p 85-94
- Murray CJL et al. 2000;Development of WHO guidelines on generalised cost-effectiveness analysis; Health economics: 9 p 235-51

These guidelines do not concern the assessment of health economic studies, they provide recommendations how an evaluation should be carried out, e.g. with regard to perspective, form of analysis (CUA, CEA, ...), etc.. Weinstein et al., for instance, provide very concrete

recommendations (e.g. ‚societal perspective' and CUA are to be preferred), while Drummond et al. or also the Hannoveraner Konsensus Gruppe leave it to the authors e.g. to choose the perspective or form of analysis, but they point out that every decision in the context of a health economic study needs to be well substantiated. For example: Why did the authors chose a 'health care perspective' and not a 'societal perspective? Or: why was a CUA preferred over a CEA?

An article by Walker D (2001) provides a good overview of all the various guidelines: How to do (or not to do)…Cost and Cost-effectiveness guidelines: which ones to use?; Health Policy and Planning; 16/1:113-121

8 Study examples of other non-invasive fields

Since there is only a limited amount of literature on health economic evaluations in the field of osteopathy, I would like to present a few examples from other related fields.
The descriptions below do not discuss the clinical aspects in detail because the focus should lie on assessing the quality of the health economic evaluation. Thus this aspect will be emphasized in the analyses. The checklist mentioned above will serve as a basis for the assessments.

8.1 UK BEAM Trial Team – Cost-effectiveness of physical treatments for back pain in primary care

As the name of the team of authors implies the study took place in the UK. UK BEAM means "United Kingdom back pain exercise and manipulation".
The study is a stochastic cost-utility analysis alongside a pragmatic randomised trial with factorial design. The study lasted from August 1999 to April 2002 and the economic analysis was conducted from the perspective of the health care system.
The authors recruited 1334 participants from 181 practices around 14 centres across the UK. 1287 could finally be included in the economic analysis because they provided sufficient data.
The participants were divided in four different intervention groups:

- Best care in General Practice (GP) (326)
- Best Care plus exercise (297)
- Best care plus manipulation (342)
- Best care plus combined treatment (322)

Participants were randomised into these four groups and the test persons on the "Manipulation-Group" were further randomised into private and NHS (National Health-Care System) premises.
The main outcome measures were Healthcare Costs, QALYs and Costs per QALY over 12 months.
To collect the data for these outcome measures the patients completed the EQ-5D questionnaire at baseline, at three months and at twelve months. In addition to the questionnaire they recorded the times they used the health care system (hospital stays, visits to secondary or primary care and physical therapists private and within the NHS).
Physical therapists filled in record forms about the number of treatments.

For the cost evaluation the authors used unit costs in pounds sterling at 2000-1 prices. For NHS they took the national average costs of such services in England as basis. For private care the figures were based on data from a major insurance provider.

The costs were not discounted because the study period was only one year and the focus on the effects was within this period.

Finally, the authors carried out three sensitivity analyses. The first one examined the influence of cost "outliers", very large healthcare costs reported by a few participants.

With two other sensitivity analyses the influence of the unit costs of manipulation were assessed. The first one did so by costing the scenario in which the NHS buys all manipulation from the private sector, using private manipulation costs. The second one did so by costing the scenario in which the NHS buys half its manipulation from the private sector, using private costs when trial manipulation took place in private premises and NHS costs when in NHS premises.

The results of the evaluation over one year were that manipulation seems to be the most cost-effective addition to GP care. If manipulation was not available exercise would be the "better" alternative than combined treatment. That does not mean that combined treatment was not effective, but the costs were higher than in the other groups and the effects were not significantly better. This means that it would not be "worth" spending that much more money to have nearly the same effects on your back pain.

8.1.1 Quality assessment:

It has already been pointed out above that this assessment will be based on the aforementioned checklist (chapter 7). The ten main questions will be repeated and answered.

1. **Was a well-defined question posed in an answerable form?**
   The question of this study was which of the four treatment alternatives for 'back pain' that were compared is the most cost-effective. To justify the question the authors argued that there are proofs of effectiveness for each of the measures but the information in the literature on costs and outcomes is contradictory [3]. Thus the question is well-defined.
2. **Was a comprehensive description of the competing alternatives given?**
   The participants in the study were divided into four groups: 'best care in general practice', 'best care plus exercise', 'best care plus manipulation' and 'best care plus combined treatment'. The choice of these four groups was justified with the fact that proofs of effectiveness of each of these treatment alternatives exist in the literature.

[3] Waddell et al. 1999, Van Tulder et al. 1997, Assendelft et al. 2003, Van Tulder et al. 2000

Also comparative studies of these interventions with different results are available. But there is no study that evaluates the cost-effectiveness of each of the alternatives. No 'do-nothing' alternative was chosen because the standard intervention in the case of 'back pain' is the so-called 'best care in general practice', thus a treatment is carried out in any case. The procedures of the individual interventions were thoroughly described in the study that accompanied the health economic evaluation. Since this study focused on a cost-utility analysis and a separate article on the clinical analysis exists, the authors refer to this article for questions on details. (UKBEAM randomised trial: effectiveness of physical treatments for back pain in primary care; BMJ 2004; 329; p1377-84).

3. **Was the effectiveness of the programmes or services established?**
   In the clinical study mentioned under question 2 the effectiveness of the treatment in the individual interventions was measured separately in each group and then compared with each other. To assess the effectiveness of the clinical study several questionnaires were used: 'the Roland Morris disability questionnaire', 'the modified Von Korff Scales', 'the back beliefs questionnaire', 'the fear avoidance beliefs questionnaire' and two generic measures: 'the SF-36' and the 'EQ-5D'. Inclusion and exclusion criteria were defined with regard to age and medical history (various contraindications) and also the scores of the 'Roland Morris disability questionnaire'. A detailed list of all the criteria can be found in the report on the clinical study (cf. question 2).
4. **Were all the important and relevant costs and consequences for each alternative identified?**
   Only direct medical costs were evaluated, because the study was carried out from the perspective the health care system. All indirect costs and also direct non-medical costs (productivity losses, sick-leave days, transport costs,…) were not included. The choice of this limited cost evaluation can be justified with the choice of perspective. The choice of perspective was justified with the argument that the authors wanted to contribute to making allocation decisions in health care easier.
   In the identification of the effects only the questionnaires are mentioned but there is no detailed discussion of the individual parameters.
5. **Were costs and consequences measured accurately in appropriate physical units?**
   The costs were not really measured in physical units. Instead the number of visits to a doctor or therapist etc. were recorded without considering e.g. the expenditure of time or other resources like transportation costs etc. Again this can be justified with the choice of perspective (health care). In the clinical study the consequences were

6. **Were costs and consequences valued credibly?**

   To evaluate the consequences the participants had to complete the EQ-5D questionnaire at baseline, after three months and after 12 months. Over the same period of time the participants had to record all health care services they accessed, no matter whether these had to do with their back pain or not. In addition, the therapists documented the number of treatments they provided.
   The authors described how they collected the data, but they did not explain why the data was collected like that and why the EQ-5D and not another instrument was used to determine the outcomes. In the clinical study more than one kind of questionnaire were used (e.g. also a second generic questionnaire: SF-36). But later these other instruments were not mentioned any more. Since a CUA (cost-utility analysis) was carried out, a measuring instrument with results that could be scored was necessary. This is probably the reason why only the EQ-5D is mentioned in the documentation of the CUA. It would be interesting to know how the participants were instructed concerning the questionnaires (in particular the EQ-5D) and how it was ensured that they completed the questionnaires to the best of their knowledge. The study does not mention e.g. whether the questionnaires were completed under supervision during a treatment session or alone at home. Based on the results QALYs were calculated, which can be explained with the design of the CUA. The costs were indicated in pounds sterling based on the market prices of 2000-1, which is a current method applied in this context. The costs for private care were based on the data of a major insurance company.

   The fact that a cost-utility analysis was carried out is mentioned but not explained. It can be supposed that this kind of analysis was chosen because of recommendations found in the literature.

7. **Were costs and consequences adjusted for differential timing?**
   No discounting was carried out. This was justified with the fact that the study period was limited to one year.

8. **Was an incremental analysis of costs and consequences of alternatives performed?**

   Yes, an incremental analysis was carried out.

9. **Was allowance made for uncertainty in the estimates of costs and consequences?**

   Three sensitivity analyses were carried out. Their description indicates that they were

one-way-analyses. It is not discussed why exactly the selected variables were chosen and not others. In the first sensitivity analysis ‘outliers’ were identified, which distorted the result on the side of the costs. Therefore these ‘outliers’ were excluded from the CUA. The second sensitivity analysis looked at in how far the costs of private practitioners (spinal manipulation), who do not practice within the NHS, could disturb the cost-effectiveness ratio because they are more expensive. Due to this sensitivity analysis the result shifts in favour of the ‘combined treatment’. In the third sensitivity analysis all manipulations (of the pure manipulation group and the combined treatment group) were attributed the costs of private practitioners. With that the result shifted in favour of the exercise group.

10. **Did the presentation and discussion of study results include all issues of concern to users?**

    The conclusion of this study discusses that even though the result changed in the various sensitivity analyses, the change was so small that both the ‘combined treatment’ and the ‘manipulation alone’ can be recommended as cost-effective in addition to ‘best care’. The seemingly biggest problem is that, if one assumes that the treatment is provided by therapists practicing within the NHS (only that would guarantee the cost-effectiveness), not enough qualified practitioners would be available and thus this treatment variant could not be ubiquitously provided in England. While the training to provide a specified exercise program takes relatively short time, it takes years to train therapists in good ‘spinal manipulation’. Due to this fact it would be cost-effective up to a certain limit to buy personnel resources from the private sector.

Comment:

This study fulfilled the quality criteria in many respects. Nevertheless, a quite “convenient” way was chosen in particular regarding the choice of perspective and thus also the cost evaluation, because due the ‘health care perspective’ some cost factors could be generously omitted. From a societal perspective the cost evaluation would have been much more complicated because many small and big “cost producers” would have to be considered. In addition, it is quite short-sighted not to regard e.g. days of sick-leave as cost factors for the health care system.

The evaluation of the consequences seems to be a little intransparent because the choice of instruments (questionnaires) is not substantiated by valid arguments in the clinical study and also the use of the EQ-5D is not commented in the CUA.

The sensitivity analyses showed that due to cost outliers and also through consumption of private care the results could be altered relatively easy and thus their robustness could be called in doubt.

In the study's own final evaluation important questions concerning the availability of the various services of health care are discussed. In England many physical therapists, chiropractors and osteopaths do not practice within the NHS which causes additional costs for the patients. The necessary number of therapists cannot be provided by the NHS. This means in particular that this study will not entail any immediate consequences with regard to the decisions within the NHS concerning the recommendation of certain treatments, because there is a lack of quantitative and qualitative personnel resources in the NHS.

8.2 Patrick DL et al. – Economic evaluation of Aquatic exercise for Persons with Osteoarthtitis

This study took place in Washington State in the United States. It was a randomised trial involving 20 week aquatic classes. The study was carried out from the societal perspective.
249 adults aged 55 to 75 were recruited between March and December 1997. The main inclusion criterion was a doctor-confirmed diagnosis of osteoarthritis. The patients were randomised into a treatment and a control group using a stratified randomisation, because the authors wanted similar numbers of women and men in the two groups.
For the economic evaluation the Quality of Well-Being Scale (QWB) and the Current Health Desirability Rating (CHDR) were used as preference weighted health status measures. In addition to these two measures, three non-preference-weighted measures (Arthritis Specific Health Assessment Questionnaire (HAQ), Perceived Quality of Life Scale (PQLS) and Center for Epidemiologic Studies Depression Scale (CES-D)) were used. All these measures were collected at baseline and post-class.
The visit of other healthcare facilities was established on the basis of diaries/questionnaires so that the collection of cost data was comprehensive.
The costs for medical care goods and services were based on the 1997 Medicare reimbursement rates in Washington State. For non-traditional health care interventions like massage or acupuncture average costs of ten Seattle area providers were taken as a basis. Medication costs were based on the 1997 Drug Topics Red Book [4]. For costs of aids and devices (canes, walkers, wheelchairs,…) an average of ten Seattle area special equipment retail stores was taken as a basis. Costs for housekeeping were established on the basis of the cost-diaries of the participants.

[4] Medical Economics Company; 1997 Drug Topics Red Book;Montvale;NJ

Class fees (for renting the pool and the paying the instructors), transportation costs, time costs associated with travelling to and from and participating in the classes were also collected for the cost evaluation.

In the economic analysis QALYs were calculated and future costs and QALYs were discounted at 3%.

To derive a 95% confidence interval for the incremental cost-effectiveness ratio the bias-corrected nonparametric bootstrap method was used. To do so QALY results derived from the QWB and CHDR were applied.

The results proved that the participants of the exercise group had equal (QWB) or better (CHDR, HAQ, PQOL) health related quality of life and cost savings regarding other medical interventions, but the economic analysis showed, that the classes were so expensive, that exercise cannot be recommended as cost-effective.

8.2.1 Quality assessment

1. **Was a well-defined question posed in an answerable form?**
   The question of this study is whether 'warm water exercise' (that was proven to be effective in several studies [5]) is also cost-effective from a societal perspective. This is a good question for a health economic evaluation.
2. **Was a comprehensive description of the competing alternatives given?**
   The participants were divided into an intervention group and a control group. The participants in the intervention group received their usual treatment (medication,…) and in addition they frequented 'warm water exercise classes' for a period of 20 weeks, while the participants in the control group did not change anything in their normal treatment. An important exclusion criterion was that the participants of both groups did not pursue any other exercise programs.
3. **Was the effectiveness of the programmes or services established?**
   An RCT with a 'stratified randomization' was carried out because the authors wanted to make sure that both groups were attributed approximately the same number of men/women, because previous studies have shown that the outcomes for men and women could be quite different. The inclusion and exclusion criteria involved a confirmed diagnosis, an age between 55 and 75, no other current exercise program, permission of the attending doctor to participate in the program, no simultaneous participation in another study, no pending surgical interventions (prostheses), being resident in an area where water gymnastics was offered and compliance.

[5] Templeton etal.1996, Suomi et Lindauer 1997

4. **Were all the important and relevant costs and consequences for each alternative identified?**

   The authors have to be commended on their careful and meticulous handling of the cost and utility evaluation.

   For the cost evaluation many non-medical costs were collected e.g. on the basis of diaries, while five different questionnaires and scales were used to evaluate the utility.

   For the cost evaluation all costs for medical interventions (visits to doctors, chiropractors, massage, electro-therapy, thermo-therapy,…), costs of medications, costs for aids and devices (canes, walkers, wheelchairs,…), household aids, expenditures to publish and promote the water gymnastics classes, transportation costs, time expenditure and also the costs of the 'warm water exercise' classes (including pool rent and instructor fees,…) were included. As mentioned above, five different questionnaires were used to collect data for the utility evaluation. Two 'preference weighted' and three 'non-preference-weighted' measures were completed by the participants at baseline and at the end of the study period.

5. **Were costs and consequences measured accurately in appropriate physical units?**

   In the cost evaluation adequate units were used. The number of medical interventions was documented. Distances of transportation were measured and the expended time was recorded. Based on this the cost factors were determined. Due to the large number of different questionnaires many parameters were covered in the utility analysis.

6. **Were costs and consequences valued credibly?**

   The cost factors were valued very carefully. The values of medical interventions were based on the 1997 rates in Washington State; medications were attributed values according to the prices of the 1997 Drug Topics Red Book; the costs for aids and devices were calculated on the basis of an average of ten local retailers; the same procedure was applied to calculate the costs for home helps; a mileage allowance was set for transportation costs; the time expenditure was valued with the mean income of the average participant.

   Every participant received $ 10 for every completed questionnaire, which increased the compliance considerably. Thus the collected data was quite comprehensive.

   A CUA was chosen for evaluation but this choice was not substantiated by any argument.

7. **Were costs and consequences adjusted for differential timing?**

   Both the discounted and the not-discounted results were presented. For discounting a

rate of 3% was used as recommended by the US Panel for Cost effectiveness in health and medicine.

8. **Was an incremental analysis of costs and consequences of alternatives performed?**

   Incremental cost-effectiveness ratios were calculated.

9. **Was allowance made for uncertainty in the estimates of costs and consequences?**

   The result was statistically analysed by means of non-parametric-bootstrapping, but no sensitivity analysis was carried out. Also this decision was not explained. Therefore possible uncertainties of the results were not at all discussed.

10. **Did the presentation and discussion of study results include all issues of concern to users?**

    A big advantage of this study was the good compliance of the participants. This holds for the participation in the water gymnastics and also the completion of the questionnaires or the writing of diaries.

    The authors mention that this scenario usually does not correspond to reality and that with a worse compliance of the patients in an everyday situation the utility of the therapy would probably be not so good.

Comment:

The big advantages of this study are the very meticulous cost and utility evaluation and the extraordinary compliance of the participants.

The disadvantages are the lacking sensitivity analyses and the absence of explanations concerning the choice of analysis.

The results of this study show (provided that the results would not change in sensitivity analyses) that effective does not necessarily always mean cost-effective, since in this case the patients did have a better quality of life but the costs of the intervention clearly exceeded the gain in quality of life. The study mentioned in chapter 8.1 was a good example which showed that one or several sensitivity analyses can alter the result and change the significance of the study. This study shows that maybe costs are sometimes underestimated (especially if a societal perspective is used and thus many factors have to be considered), but the result can rightly be doubted due to the lack of sensitivity analyses.

Also the compliance of the patients does not reflect reality as correctly pointed out by the authors. This means that if such an exercise program is offered in reality, only part of the participants would take advantage of it, which means that the costs for each participant could increase further, e.g. because the pool rent and the instructor's fees would have to be distributed among fewer persons (higher per capita costs).

8.3 Moffett JK et al. – RCT of exercise for low back pain: clinical outcomes, costs and preferences

This randomised controlled trial (RCT) took place in the U.K. and the evaluation period was one year. The authors recruited 187 patients aged 18-60 years with mechanical low back pain of 4 weeks to 6 months duration. The participants were divided into two groups: one exercise group and one GP care group.

The main outcome measures were taken at baseline, after the intervention period that lasted four weeks, six months later and one year later. The authors used several questionnaires for the evaluation of the effects: the Roland disability questionnaire, the Aberdeen back pain scale, the EQ-5D, pain diaries, the fear and avoidance beliefs questionnaire (FABQ). The participants also had to record which other health care services they used within the study period and how many days off work they had.

Costs were evaluated by retrospective questionnaires and prospective diary cards both for the NHS and the societal perspective and the costs were checked with a non-parametric bootstrap.

A special “feature” of this study was the evaluation of patients’ preferences concerning the treatment. Before randomisation all participants were asked, which group they would prefer. Then they were fully randomised and after the study period the influence of patients’ preferences was evaluated separately. However, the analysis showed no significant influence of whether the patients received their favourite treatment or not.

The result of the economic analysis was that the exercise group was better in all aspects and therefore cost-effective.

8.3.1 Quality assessment

1. **Was a well-defined question posed in an answerable form?**
   The main questions of this study were whether a targeted exercise program could help patients with 4-weeks to 6-months back pain to return to their daily activities more easily and faster and to re-establish their trust in the functioning of their spine, and whether this exercise regime was cost-effective also from a ‘societal perspective’. This question can be regarded as sufficiently well-defined, because the question is clearly formulated regarding the health economic analysis.

2. **Was a comprehensive description of the competing alternatives given?**

   In this study the participants were divided into two groups. The intervention group observed a 4-week exercise and training program (eight sessions), while the participants of the control group continued to see their doctor and recorded the progress. Before the patients were randomised, they were asked about which treatment they would prefer to afterwards evaluate whether someone's personal preference would influence the result.

3. **Was the effectiveness of the programmes or services established?**

   Since the literature provides differing information on the effectiveness of an exercise program [6], the authors of the study developed their own exercise program based on existing studies to evaluate it in an RCT. The health economic evaluation was carried out at the same time. The inclusion criteria were: patients between 18 and 60 years of age with 'mechanical low back pain', living in a certain area (so that they could easily reach the site of the exercise classes) and qualified by their doctor as suited for participation in the study.

4. **Were all the important and relevant costs and consequences for each alternative identified?**

   Unfortunately, only little information is provided concerning data collection. The only fact that is mentioned is that that retrospective questionnaires and prospective diaries were used. The tables that are provided for illustration show that only the direct medical costs and sick-leave days were considered. It is not explained why no indirect costs were established and why no references concerning the set costs e.g. per visit to the doctor were provided. Other information that is lacking includes e.g. costs of medication, while costs of aids and devises are mentioned. If really not a single participant had taken any medication during the study period, this should at least be mentioned.

   The evaluation of the utility was based on different questionnaires, which is a positive aspect of this study. The evaluation shows that different questionnaires provide different results at different moments in the study. This suggests that if you rely on only one type of questionnaire, misleading results could be obtained.

   It is also briefly mentioned that both a NHS perspective and a societal perspective was chosen, but it is not explained why. Many cost parameters were not considered (e.g. all indirect costs) in particular in the context of the 'societal perspective'.

5. **Were costs and consequences measured accurately in appropriate physical units?**

   It has already been mentioned under question 4 that the direct medical costs and the

[6] Malmivaara et al. 1995; Faas et al. 1993; Frost et al. 1998a; Frost et al. 1998b

sick-leave days were evaluated. In the utility evaluation different kinds of questionnaires were used, which were analysed with the appropriate scoring formulas. No information is provided concerning the physical units used for measurement.

6. **Were costs and consequences valued credibly?**

   The information concerning this aspect is also quite imprecise. There is no mention of sources on which the costs were based, e.g. how was the value of therapy calculated? etc. In addition, there is no explanation of what kind of economic analysis was used and why.

7. **Were costs and consequences adjusted for differential timing?**
8. **Was an incremental analysis of costs and consequences of alternatives performed?**
9. **Was allowance made for uncertainty in the estimates of costs and consequences?**

   Neither discounting nor incremental analysis were performed, while a sensitivity analysis was carried out but without any explanation of the reasons. A non-parametric-bootstrapping was carried out to statistically analyse the costs.

10. **Did the presentation and discussion of study results include all issues of concern to users?**

    A conclusion along the lines of an economic analysis was lacking. Only the effects and results are described, but no cost-effectiveness ratios or similar are calculated to convert the results into figures. The discussion only mentions several other studies, which looked at similar questions (but did not include an economic analysis), but avoids direct comparisons. Also the feasibility in reality is only mentioned in passing and no reference is made concerning e.g. the program's practicability.

Comment:

An aspect that has to be regarded as positive is that the patients' preference of treatment was considered. It was established before randomisation. The patients were asked which study group they would prefer but then they were randomised in the two groups (without respecting this preference). However, the preference was considered in the analysis but the authors found out that it had no influence on the results.

The most interesting result is the difference in the sick-leave days over the year of the study period. With 378 the intervention group had much less sick-leave days than the control group with 607.

Nevertheless, overall the study has considerable deficits, in particular regarding the lack of discounting, incremental analysis and sensitivity analysis. It is not mentioned what kind of economic analysis is carried out. In addition, there is no explanation why certain cost

factors have not been included. Basically, this study is worthless as health economic study. Hardly any criteria of the checklist were fulfilled. Maybe an expert could provide a good health economic evaluation on the basis of the study's data, but the authors of the study did not do that. As far as a judgment is possible, theoretically an evaluation from a 'health care perspective' seems to be possible (even though the costs of medications are lacking) because only the direct medical costs are necessary for it. In any case it would be necessary to carry out more comprehensive sensitivity analyses to confirm or deny the robustness of the results. Whether discounting of the results makes sense or not is debatable because the study period is limited to one year and also other authors (cf. 8.1) abstain from discounting in this case. Summarizing it can be said that the authors tried to approach the field of health economic evaluations but they could not provide valid results.

8.4 Korthals-de Bos IBC et al. – Cost-effectiveness of Physiotherapy, Manual Therapy and General Practice Care for neck pain: economic evaluation alongside a RCT

Like the previous study this study is also based on an RCT that took place in the Netherlands. The study included 183 patients aged between 18 and 70 with neck pain for at least two weeks. The participants were recruited by 42 GPs and randomly allocated to three groups. A manual therapy group (mainly spinal mobilisation), a physiotherapy group (mainly exercise) and a GP care group (counselling, advice and medication).

Different scales and questionnaires were used to evaluate the outcomes measures. Six point scales were used to rate the perceived recovery (intensity of pain, functional disability and utility), while the mean pain during the preceding week was indicated by the patient on an eleven point scale. Two kinds of questionnaires were used: the neck disability index and the EQ-5D.

Outcome measures were assessed at baseline and at three, seven, thirteen and fifty two weeks after randomisation.

Costs were established from a societal perspective on the basis of cost diaries that were completed by the participants over a period of fifty-two weeks. Direct health care and non-health-care costs, indirect costs (loss of production, days off work, and days of inactivity for patients with or without a paid job) were taken into consideration.

A pair-wise comparison of the mean costs between the groups was carried out by means of bootstrapping. Bias corrected and accelerated bootstrapping helped to establish confidence intervals for the mean different costs. Also the cost-effectiveness and cost-utility ratios were

calculated with bootstrapping according to the bias corrected percentile method, by using the clinical outcomes.

Within the framework of a sensitivity analysis the data of two patients who were referred to hospital were excluded.

The results of the cost-effectiveness and cost-utility analysis were that manual therapy is more cost-effective than physiotherapy and GP care. Manual therapy had significantly lower costs and slightly better effects at one year.

8.4.1 Quality assessment

1. **Was a well-defined question posed in an answerable form?**

   The questions were which treatment option (general practitioner, physiotherapy, manual therapy) was effective for the treatment of non-specific neck pain and which of the treatment methods was more cost-effective. Again the question is suitable for the study.

2. **Was a comprehensive description of the competing alternatives given?**

   The participants were divided into three groups. The 'manual therapy group' was treated by chiropractors, osteopaths and physical therapists (with special manual therapy training). The treatment included specific mobilisation techniques without manipulation (low amplitude high velocity). The 'physiotherapy group' was treated by physical therapists. The therapy consisted of individual exercises and optional massages and/or tractions. The 'GP care group' received a 'typical treatment' by a physician, including advice, information, medication and waiting for spontaneous recovery. There was no do-nothing alternative because all study participants were patients who contacted a physician on their own accord.

3. **Was the effectiveness of the programmes or services established?**

   The health economic analysis was carried out within the framework of the RCT. The intensity of the patients' complaint, their age and their place of residence in relation to the place of treatment were considered in the randomization. Since only little information is available concerning the effectiveness of the therapy methods applied in this study [7], the objective was first of all to determine their effectiveness and secondly, to evaluate their cost-effectiveness.

4. **Were all the important and relevant costs and consequences for each alternative identified?**

   The 'societal perspective' was chosen for the cost evaluation. This means that direct medical costs (manual therapy, visits to the doctor, physical therapy, other services

[7] Gross et al.1996; Skargren et al. 1998

within the health care system, medication, home-based care and hospital stays), direct non-medical costs (out of pocket costs, costs for paid and unpaid help and travel expenses) and indirect costs (loss of production, absenteeism from work or days of inactivity for patients with or without a paid job) were considered. The utility and effectiveness evaluations were carried out with the aid of various instruments (two analogue scales and two questionnaires) according to the criteria: 'perceived recovery', 'intensity of pain', 'functional disability' and 'utility'.

5. **Were costs and consequences measured accurately in appropriate physical units?**

   It can already be recognized under question 4 that data were measured in different physical units (intensity of pain, utility measures, time loss,…) and thus adequately collected.

6. **Were costs and consequences valued credibly?**

   The costs of medication were valued according to the provisions of the Royal Dutch Society for Pharmacy, the costs of working time were based on the average income of the Dutch population with regard to age and gender, and the costs for unpaid work were valued with a 'shadow price' of € 7.94 per hour. Due to the different instruments that were used in the evaluation of the consequences also different data were measured. The analysis of the EQ-5D (with a uniform 'scoring formula' (cf. Annex 12.2.3.1)) provided the data for a CUA. The other outcome measures were used in a CEA. Thus both a CUA and a CEA were carried out.

7. **Were costs and consequences adjusted for differential timing?**

   Discounting lacked and was not mentioned. Due to the study period of one year it can be assumed that the authors abstained from discounting like e.g. the authors of the study described under chapter 8.1.

8. **Was an incremental analysis of costs and consequences of alternatives performed?**

   An incremental analysis was carried out in the context of the CUA and the CEA. It was calculated separately for each outcome measure instrument.

9. **Was allowance made for uncertainty in the estimates of costs and consequences?**

   Unfortunately only one sensitivity analysis was carried out. It looked at the 'cost outliers' but it is not described in detail. Only the result, which was statistically not significant, was documented. Thus it cannot be reconstructed how much possible uncertainties were researched. Since the results of the CUA/CEA were clearly in favour of 'manual therapy', a substantiation with well documented sensitivity analyses would be desirable.

10. **Did the presentation and discussion of study results include all issues of concern to users?**

    The results show that ‘manual therapy’ outclasses the other treatment options (GP care and physiotherapy) both in effectiveness and also costs. This holds for the intervention period and also for the observation period afterwards (the overall study period was one year). The result is a good cost-effectiveness of ‘manual therapy’ above all because of its superiority regarding costs. Nevertheless, more patients are referred to physical therapists by the general practitioners in the Netherlands rather than to manual therapists, maybe because the doctors do not know enough about the different options. This is at least what the authors of the study suspect. The literature provides clinical studies on ‘neck pain’, which look at different treatment methods but only to compare their effectiveness. There is no study that evaluates the cost-effectiveness at the same time. Unfortunately, the discussion did not talk about the feasibility of the interventions in reality nor did it look at whether the study scenario could be valid for patients in general, e.g. regarding compliance. Further, the reader does not know how the Dutch health care system works because there is no description of it. And there is no mention whether such a treatment regime could be integrated in the health care system or whether it is already offered but not used.

Comment:

This study fulfils almost all criteria of the checklist. The only things to be criticised are the lack of discounting which is not further explained and the sensitivity analyses which are not carried out satisfactorily and are only insufficiently documented. Since the study period was only one year and neither costs nor effects were studied long-term the lack of discounting is a negligible shortcoming.

Positive aspects are that the utility was evaluated with different instruments, that direct and indirect costs were collected, that the authors provided a comprehensive description of how they calculated the values of different services, that they carried out both a CEA and a CUA, that they effected an incremental analysis and a sensitivity analysis of the costs, bootstrapping and provided comparisons with studies dealing with a similar topic.

The unsubstantiated lack of discounting and the insufficiently presented and justified sensitivity analysis are points of criticism. Further, the discussion of the study could be more elaborate and look more at in how far the options are realizable in general and what consequences they could produce.

9 Study examples from the field of osteopathy

9.1 Williams NH et al. – Cost-utility Analysis of osteopathy in primary care

This Welsh trial was conducted between September 1997 and March 2001. The participants were between 16 and 65 years old and were referred from 14 different general practices. Symptoms were neck, upper or lower back pain of 2-12 weeks duration.

The trial team used a secure method of stratified randomisation and carried out a CUA alongside this pragmatic single-centre RCT.

The study included 200 participants who were divided into two groups. One osteopathy plus GP care group (= intervention group) and one group with GP care alone.

For the utility measurement the authors used the EQ-5D questionnaire at baseline, after two months and after six months and calculated QALYs afterwards.

Costs were measured from the NHS perspective. Therefore the authors collected only direct medical costs and applied no discount rate because the study period was only six months.

For the statistical analysis they used a non-parametric bootstrap method. A sensitivity analysis was carried out to be able to exclude cost outliers.

The intervention group showed a relative improvement in mean QALY and a small increase in mean health service costs. Thus it seemed that the addition of osteopathic treatment is cost-effective, but the authors emphasize that this conclusion is subject to considerable random error.

9.1.1 Quality assessment

1. **Was a well-defined question posed in an answerable form?**
   Since there are several studies that look at the effectiveness of different measures to treat 'back pain' and some even evaluated the costs involved [8], the question is what result a CUA in this context would provide. Such a CUA is carried out on the basis of a 2001 RCT (published 2003) by Williams et al. Thus the study's question is appropriate.
2. **Was a comprehensive description of the competing alternatives given?**

[8] Meade et al. 1990; Cherkin et al. 1998; Waterworth et al 1985; Skargren et al. 1997; Niemisto et al 2003; Korthals-de Bos et al 2003

The participants were divided into two groups. Both groups continued their GP care, while the patients of the intervention group were additionally referred to an osteopathic institution to receive an osteopathic treatment. The osteopathic treatment consisted mainly of 'spinal manipulation', but it also included advice concerning activities, exercises and the avoidance of too much immobility. A disadvantage of this study is that the osteopathic treatment was delivered by only one osteopath, who is also a doctor.

3. **Was the effectiveness of the programmes or services established?**
   It has already been mentioned under question 1 that the CUA of this study was carried out on the basis of an RTC that took place in 2003 (Williams NH et al. : Randomised osteopathic manipulation study (ROMANS): pragmatic trial for spinal pain in primary care; 2003; Family Practice; 20:6; p 662-9)
   Within this 'pragmatic RCT' slight but nevertheless significant advantages of osteopathic treatment with regard to the effects of the intervention could be detected. Different instruments were applied in the original study to evaluate the outcomes (Extended Aberdeen Spine Pain Scale (EASPS), Short-Form McGill Pain Questionnaire (SMPQ) and two generic measures, the SF-12 and the EQ-5D) but the EQ-5D was the only one used for the CUA (cf.4.).
   Inclusion criteria were the patients' age between 16 and 65 years, mechanical back pain (neck, upper back, lower back) of 2-12 weeks duration, no serious pathology, no 'red flag' symptoms and no nerve root pain.
4. **Were all the important and relevant costs and consequences for each alternative identified?**
   Regarding the cost evaluation only the direct medical costs were evaluated (like in many other studies) but the "lack" of other cost factors was not explained. However, the selected NHS perspective could be an argument for this choice of cost evaluation. The choice of perspective was substantiated with the argument that the authors wanted to present the NHS point of view.
   When the utility was evaluated the CUA was only based on the EQ-5D without explaining this any further. It has already been mentioned in chapter 8.3.1 that in the utility evaluation different instruments can provide different results at different moments, which can either speak in favour of or against a certain intervention.
5. **Were costs and consequences measured accurately in appropriate physical units?**
   The costs were documented on the basis of how many times medical care was used. The number of visits at doctors, therapists, diagnostic measures, etc. was recorded and the prescribed and consumed medication and other means were documented.

The utility evaluation was carried out by means of the EQ-5D, on the basis of which QALYs were calculated. The authors did not explain why exactly this questionnaire was used.

6. **Were costs and consequences valued credibly?**

   The costs per unit were calculated on the basis of 'national sources' and 'finance officers of local provider units' for the years 1999/2000. The EQ-5D had to be completed at baseline, after two months and after six months. It is not explained whether the participants did complete the questionnaires under supervision and with the appropriate instructions or whether the forms were sent by mail and the participants filled them in at home. A CUA was chosen as form of analysis. In this context also QALYs were calculated. However, no reasons were given why this form of analysis was chosen.

7. **Were costs and consequences adjusted for differential timing?**

   The lack of discounting was explained with the short study period.

8. **Was an incremental analysis of costs and consequences of alternatives performed?**

   An incremental analysis was carried out and incremental cost-utility-ratios were also calculated.

9. **Was allowance made for uncertainty in the estimates of costs and consequences?**

   A sensitivity analysis was carried out for the costs and three 'outliers' were excluded. However, the authors did not describe what kind of sensitivity analysis they chose and which parameter was used. In addition, they did not explain why only participants with extraordinary high costs distort the result and other parameters are not regarded as uncertain. The statistical analysis also included a 'non-parametric percentile bootstrapping'.

10. **Did the presentation and discussion of study results include all issues of concern to users?**

    The most outstanding feature of this study is that it describes in detail how the economic analysis was carried out.

    In the conclusion the authors also refer to the study described under 8.4 and they criticise precisely the aspect of the intransparent economic analysis.

    In their own study the authors see the aspect of the generalization of the data as weak point, because the treatments were carried out in only one institution by only one therapist/doctor. They argued that no comparable institutions were available, which again gives rise to the question whether the result has a significance for the NHS if there are not enough institutions available.

Comment:

In this study the cost evaluation (only direct costs) and the utility evaluation (with only one instrument) are questionable. Even though the authors describe the economic analysis very much in detail, they lack this detail in the description of how the data were collected. Also the sensitivity analysis is discussed only in passing and is hard to comprehend.

In summary, this study was very meticulously executed with regard to certain aspects, but there are also some aspects that are quite uncertain and difficult to generalize.

9.2 Gamber R, DO, MPH et al. – Cost-effective Manipulative Medicine: A literature review of Cost-effectiveness Analyses for osteopathic manipulative Treatment

First of all, I have to point out that this is an American study, which means that some particularities have to be considered. In America a DO (Osteopathic Physician) is a fully licensed physician like an MD (allopathic physician) but with special training in manual medicine. When the authors use the terms OMT or OMM they mean 'Osteopathic Manipulative Treatment' or 'Osteopathic Manipulative Medicine'.

Despite the different terminology, the problems in America and Europe seem to be similar. Many studies look at the effectiveness of osteopathic treatment, but there is no or only little information available about the costs or the cost-effectiveness of OMM.

The authors searched the OSTMED and the MEDLINE databases for articles in English published between 1966 and 2002. The key-terms were cost-effectiveness, osteopathic medicine, workers' compensation, length of stay in hospital, healthcare providers and manipulative medicine. The authors reviewed the references of the articles and extended their search.

The main inclusion criterion was that the cost of OMM compared with standard medical care must have been analysed and discussed.

16 published papers and reports were identified that met the inclusion criteria. 11 of them were published in peer-reviewed journals and five in non-peer-reviewed journals. Additionally the authors found four government reports that deal with the cost-effectiveness topic.

The articles were divided into two groups, because not all of them included direct measurement of cost data but imputed cost data.

Imputed cost data means that the effort was evaluated with other parameters like 'length of stay in hospital', 'effort invested by the provider', 'time lost from work' and so on.

The usage of imputed costs is problematic, because of the risk to be imprecise. Thus the authors divided the articles into the aforementioned two groups.

The first group (direct measurement of costs) consisted of nine and the second group (imputed cost variables) of eleven reports.

After describing these 20 reports the authors concluded that many studies exist that have an acceptable measurement of clinical effectiveness but that the quality and the amount of the measurement of cost data respectively cost-effectiveness is weak.

For the credibility of cost-effectiveness analyses it is indispensable to have a well designed cost evaluation. If imputed cost variables are used, it would be helpful to use standards to make them more valid.

Finally the authors argue that a cost-effectiveness analysis (CEA) of good quality consists of a careful correlation between cost data and clinical outcomes, the use of statistical methods to manage limitations and definition of all used parameters (actual and imputed).

The authors point out that the strongest research design would be produced by collaborations between OMM providers, third-party payers, biostatisticans and healthcare economists.

9.2.1 Quality assessment

Since this study is not a clinical study but a literature review, it cannot be evaluated on the basis of the checklist.

The authors carried out a systematic search of relevant studies. But since they did not find too many they also included studies whose cost evaluation seemed to be questionable. They proceeded correctly by dividing the studies included in the review into two groups to point out the differences in the cost evaluations.

Unfortunately, the authors did not refer to any checklist, guidelines or recommendations in their description and evaluation of the studies, which could have helped to carry out a systematic analysis. The result is that it is not clear in the analysis of the clinical outcomes which study used which instruments and why. The final commentary only points out that the measurement of clinical efficacy was largely satisfying.

The authors identify considerable shortcomings in the collection of the data concerning costs, in particular in the group of ‘imputed costs’.

Again hardly any concrete details are provided regarding the evaluation of the cost data collection. Thus it cannot be excluded that only the most obvious shortcomings were described.

Comment:

This review illustrates the main problem of health economic evaluations in the field of osteopathy. It is used too rarely and with not enough professionalism and thus looses a lot of its credibility. In their conclusion the authors point out correctly that an interdisciplinary cooperation is necessary to carry out a health economic analysis as best as possible.

Despite the above mentioned points of criticism, this review clearly shows that osteopathy still needs to learn a lot with regard to the field of health economics.

10 Summary

Health economics is a relatively young science which tries to provide help in word and deed in the discussion about costs in the health care system.

The starting point is the definition of what can be regarded as “economic” in the health care system.

So-called “health economic analyses” were developed to evaluate various interventions and to compare them with each other. These types of analyses are referred to as cost-effectiveness analysis or cost-utility analysis.

As the names suggest they try to put costs and effectiveness or costs and utility in relation with each other to calculate cost-effectiveness or cost-utility ratios.

Health economic analyses look at e.g. two or more treatment alternatives to compare their results not only from a clinical perspective but also from an economic point of view.

Is the expensive treatment A worth the money in comparison with the less expensive treatment B because it provides much better results? Or could you say: huge expense, no recompense?

In order to carry out such an analysis a large amount of all sorts of data has to be collected.

On the one hand the costs have to be established and on the other hand the effectiveness or utility.

When costs are established, it is important to first define the perspective from which the study is carried out. Many recommend the “societal perspective” but in this case the collection of cost data is the most difficult and comprehensive because many cost factors have to be considered.

To establish the utility many instruments can be used. Basically we differentiate between different interview forms and questionnaires. Both have considerable advantages and disadvantages. The interviews require a lot of personnel but take the situation of the individual patient better into account. They also require seemingly quite serious decisions (probability of death, loss of years of life) of patients who do not suffer from life-threatening diseases (e.g. lumbago) and the patients might have cognitive difficulties to make adequate decisions. The questionnaires require the patients’ understanding of the questions. In addition, they have a set weighting of the individual parameters which might not reflect the individual problem of the patient.

Another aspect is that there is not only one kind of health economic analysis but several which all have their particularities and individual characteristics.

The most commonly known are the cost-effectiveness analysis and the cost-utility analysis, but there are also cost-benefit analyses, cost-minimisation analyses, outcome-maximization analyses, etc.

It seems, however, that for the field of osteopathy only the first two types are relevant due to their design. Since neither the outcome nor the costs are known at the beginning of a clinical study (which is carried out frequently) both aspects have to be evaluated, which corresponds to the design of either a CEA or CUA. A CBA also looks at both aspects, but it is often criticized because it attributes a monetary value to utility and thus is recommended rather rarely.

Once all the data are collected and the above mentioned cost-effectiveness or cost-utility ratios are calculated, the analysis is not finished.

The differences of the interventions that are to be compared have to be subjected to an incremental analysis and their “future prospects” have to be calculated by means of discounting. Finally, all the uncertain variables have to be subjected to a sensitivity analysis.

This description illustrates that a health economic analysis is quite a complex and laborious task that requires a lot of accuracy. All these things should not deter people from carrying out such analyses but everybody should realize the complexity of the work in reality. Maybe all this can be an incentive for osteopaths who want to carry out research projects and like a little bit of challenge.

Regarding the studies provided as examples many aspects can be criticized, but some things can be learned and deduced from this. Despite the points of criticism one cannot say that the results of the studies are not true.

No matter whether these health economic analyses have one or the other methodological weakness, the collected data (costs and consequences) are available in any case. Thus equal parameters of different interventions can be directly compared with each other: e.g. which group had more days of sick-leave? The patients of what intervention did need less health care services in the observation period? How did the outcomes of the utility evaluation present?

Given all the criticism regarding utility evaluations, the conditions are still the same for all participants and thus they can be compared with each other.

Nevertheless, the final result of a health economic analysis should only be considered as robust if the main criteria of the checklist are basically fulfilled. If, for instance, no sensitivity analysis was carried out, the results cannot be regarded as robust.

In addition, it is important to evaluate whether the “environment” in which the study was carried out can be generalized. The question is whether a constructed scenario was created, tailored to the specific problem or whether the scenario was really in step with actual practice. In how far did country-specific parameters (e.g. prices) influence the result?

A very crucial aspect in the critical review of health economic studies is that one can learn in particular from mistakes – from your own mistakes and also from those of others. Therefore methodological shortcomings will for sure be progressively improved in the future.

11 Conclusion

This paper shows clearly that the field of health economic evaluations provides a lot of topics for discussion, whether these concern the field of cost or utility evaluation, ethical objections to the concept of health economics as such, problems regarding the efforts and feasibility of the evaluations or doubts regarding the significance of the results.

In the context of osteopathy health economic analyses seem to be even more difficult because often there is a lack of resources to facilitate such a complex and laborious work. In any case such efforts are not possible regarding studies that are carried out within the framework of someone's osteopathic training.

Nevertheless, it has to be pointed out that health economics is an upcoming discipline of economics which is gaining in importance. Thus it deserves attention also from the osteopathic profession. Osteopaths are striving for official recognition in Austria thus they will need to face and deal with this "laborious" field of research. In particular since Austria now disposes of a Federal Institute for Quality in the Health Care System (Bundesinstitut für Qualität im Gesundheitswesen, BIQG).

The question is how osteopathy can avail itself of this kind of research with the means it can dispose of.

Basically, one should only try to carry out accompanying health economic analyses for studies that are implemented on a large scale and have a "good stance" as randomised controlled trial (RCT).

It also has to be considered that from the beginning of the planning phase of the study, the team has to include a specialist in health economics in order to carry out the analysis in the best possible way.

Also the osteopath should look into the subject, but not for nothing are health economics a separate branch of science.

If someone does not have the possibilities to carry out a large-scale study or to recruit a whole team of researchers, one can still contribute to the collection of additional data. It will probably not make sense to carry out a complete CEA in a one-man-study, but some of the collected raw data might be useful in the future.

Which data, however, could be of importance has to be discussed with experts. If various measurement instruments would be regularly applied or data regarding number of treatments, frequency of treatments or number of days of sick-leave would be consistently collected, the result would be a considerable volume of data which could be used for analysis to maybe find out significant things.

Factors which do not speak in favour of osteopathy are long waiting times or in some regions of Austria a small number of osteopaths in the area and thus maybe long travel times to reach the osteopathic practice and lack of availability.
Nevertheless, also these parameters have to be documented to guarantee credibility.
The checklist (chapter 7) points out that an important aspect is the degree to which the results can be generalized and to which the services are available. What benefit does a patient have if osteopathy can successfully treat lumbar pain but the practitioner-patient ratio is 2 : 500, which means long waiting times and thus an influence on the number of sick-leave days and/or productivity.?
This does not mean that this necessarily is the effect but despite good and convincing clinical results you must never be too sure. This is another reason to get an expert on board.

With this paper I have tried to provide an overview of the extremely complex subject of health economic evaluation methods and I hope I could raise the interest of one or the other colleague to look a little bit further into the subject.

12 List of references

Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG Spinal manipulative therapy for low back pain: a meta-analysis of effectiveness relative to other therapies; Ann Intern Med; 2003; 138:871-81

Brazier J, UsherwoodT, Harper R, Thomas K Deriving a preference-based single index from the UK SF-36 Health survey; J Clin Epidemiol;1998; 51: 1115-1128

Brazier J, Roberts J, Deverill M The estimation of a preference-based measure of health from the SF-36; J Health Econ; 2002; 21:271-292

Breyer F et Kolmar M Grundlagen der Wirtschaftspolitik; Mohr Siebeck; Tübingen 2001; Kap. 2

Breyer F, Zweifel P et Kifmann M Gesundheitsökonomik; 2005; 5. Auflage; Springer; Berlin, Heidelberg, New York; ISBN 3-540-22816-0

Brooks R et EuroQol Group EuroQol: The current state of play; 1996; Health Policy; 37; p 53-72

Brouwer W, Rutten F et Koopmanschap M Costing in economic evaluations; 2001; In: Economic Evaluations in Health Care: merging theory and practice (ed. Drummond and McGuire);Oxford University Press, Oxford; p 68-93

Cherkin DC, Deyo RA, Battie M, Street J, Barlow W A Comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain; N Engl J Med; 1998; 339:1021-29

Culyer AJ The normative economics of health care finance and provision. In: McGuire A, Fenn P, Mayhew K (Hrsg.); Providing Health Care; Oxford University Press; Oxford, England; 1991

DNEbM Deutsches Netzwerk für Evidenz basierte Medizin; http://www.ebm-netzwerk.de/grundlagen/definitionen/

Dolan P, Gudex C, Kind P et Williams A A social tariff for EuroQol: Results from a UK general population survey, Discussion paper No. 138. Centre for Health economics, University of York, York; 1995

Doubilet P, Weinstein MC et Mc Neil BJ Use and misuse of the term “cost effective” in medicine; New England Journal of Medicine; 1986; 314/4:253-56

Drummond MF et al Guidelines for authors and peer reviewers of economic submissions to the BMJ; BMJ; 1996; 313:275-283).

Drummond MF, Sculper MJ, Torrance GW, O’Brien BJ et Stoddart GL Methods for the Ecomomic Evaluation of Health Care Programmes Third Edition; 2005; Oxford University Press; New York; ISBN 978-0-19-852945-3

Essink-bot M, Stouthard M et Bonsel G Generalizability of valuations on health states collected with the EuroQol-questionnaire; 1993; Health Economics; 2; p 237-46

Faas A, Chavannes A, van Eijk JTM, Gubbles J A randomised placebo-controlled trial of exercise therapy in patients with acute low back pain; Spine; 1993; 18:1388-95

Feeny D, Furlong W, Torrance GW et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system; Medical care; 2002; 40:113-28

Friedmann M Price Theory; Aldine Press: Chicago, IL, 1962

Frost H, Klaber Moffett J, Moser J, Fairbank J Evaluation of a fitness programme for patients with chronic low back pain; BMJ; 1995; 310: 151-4

Frost H, Lamb S, Klaber Moffett J, Fairbank J, Moser J A fitness programme for patients with chronic low back pain: 2 year follow-up of a randomised conrolled trial; Pain; 1998; 75:273-9

Furlong WJ, Feeny DH, Torrance GW, Barr RD The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies; Annals of Medicine ;2001; 33:375-384

Gamber R, Holland S, Russo DP, Cruser A et Hilsenrath PE Cost-effective Osteopathic manipulative Medicine: A Literature Review of Cost-effectiveness Analyses for Osteopathic Manipulative Treatment; JAOA; 2005; 105/8: 357-67

Gold MR, Siegel JE, Russell LB et Weinstein MC Cost-effectiveness in health and Medicine; 1996; Oxford university press; New York, Oxford; ISBN 0-19-510824-8

Gross AR, Aker PD, Goldsmith CH, Peloso P Conservative management of mechanical neck disorders. A systematic overview and meta-analysis; Online J Curr Clin Trials; 1996; Document No 200-1

Gudex C, Dolan P, Kind P et Williams A Health state valuations from the general public using the visual analogue scale; Qual Life Res; 1996; 5:521-531

Hannoveraner Konsensus Gruppe 1999; Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – revidierte Fassung des Hannoveraner Konsens; Gesundh.ökon.Qual.manag.:4 p A62-5

Horsman J, Furlong W, Feeny D et Torrance G The Health Utility Index (HUI): concepts, measurement properties and applications; Health and Quality of Life Outcomes; 2003; 1:54 (electronic version available free at http://www.hqlo.com/content/1/1/54)

Hurley J An overview of the normative economics of the health sector; In: Culyer AJ, Newhouse JP Handbook of Health economics; Elsevier Science; 2000; 55-118

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Bastian H, Bender R, Ernst AS, Kaiser T, Kirchner H, Kolominsky-Rabas P, Lange S, Sawiki PT et Weber M; Methoden, Version 1.0 vom 1. März 2005, Köln

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Bastian H, Bender R, Ernst AS, Kaiser T, Kirchner H, Kolominsky-Rabas P, Lange S, Sawiki PT et Weber M; Methoden, Version 2.0 vom 19.12.2006, Köln

Kind P The EuroQol Instrument: An index of health related quality of life; In: Quality of life and Pharmacoeconomics in clinical trials (2nd ed)(ed. by Spilker);1996;Lippincott-Raven, Philadelphia; p 191-201

Korthals-de Bos IBC, Hoving JL, van Tulder MW, Rutten van Mölken MPMH, Adèr H, de Vet HCW, Koes BW, Vondeling H et Bouter LM Cost effectiveness of physiotherapy, manual therapy and general practitioner care for neck pain: an economic evaluation alongside a randomised controlled trial; BMJ; 2003; 326: 911-6

Krabbe PFM Good Day Sunshine: About Biases, Irregularities and inconsistencies in the valuation of health states; 2002; Paper presented at the 19th plenary meeting of the EuroQol Group; York, UK; p 139-146; available at: www.euroquol.org

Lauterbach KW, Stock S et Brunner H Gesundheitsökonomie – Lehrbuch für Mediziner und andere Gesundheitsberufe; 1. Auflage 2006; Verlag Hans Huber; Schweiz

Lauterbach KW et Hinzpeter B Effizienz-Analysen: Eine ethische Notwendigkeit; Deutsches Ärzteblatt; 1999; 96/17:778

Leidl R Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. In: Schwartz FW et al. Das Public Health Buch. Urban&Fischer; München und Jena; 2003 (2. Auflage); S. 461-484

Maetzel A Der Gebrauch von Nutzwerten im gesundheitsökonomischen Vergleich von Interventionen bei verschiedenen Krankheitsbildern; Zeitschrift für Rheumatologie; 2004; 63:5; 380-384

Malmivaara A, Hakkinen U, Aro T, Heinrichs M, Koskenniemi L, Kuosma E et al. The treatment of acute low back pain – bed rest, exercises or ordinary activity?; N Engl J Med; 1995; 332: 351-5

Mason J et Drummond MF 1995; Reporting guidelines for economic studies; Health Economics: 4 p 85-94

Meade TW, Dyer S, Browne W, Townsend J, Frank AO Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment; Br Med J; 1990; 300: 1431-37

Mehrholz J Studiendesign: Kosten-Nutzwert-Analyse; physiopraxis 7-8/06

Mehrholz J Studiendesign: Kosten-Effektivitäts-Analyse; physiopraxis 6/06

Moffett JK, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Phillips H, Farrin A et Barber J Randomised controlled trial of exercise for low back pain: clinical outcome, costs and preferences; BMJ 1999;319:279-283

Murray CJL et al. 2000;Development of WHO guidelines on generalised cost-effectiveness analysis; Health economics: 9 p 235-51

Naylor CD et Llewellyn-Thomas HA Utilities and preferences for health states: Time for a pragmatic approach?; 1998; J Health Serv Res Policy; 3:129-131

Niemisto L, Latinen-Suopanki PT, Rissanen P, Lindgren K-A, Sarna S, Hurri H A randomised trial of combined manipulation, stabilising exercises, and physican consultation compared to physican consultation alone for chronic low back pain; Spine; 2003; 28:2185-91

Patrick DL, Ramsey SD, Spencer AC, Kinne S, Belza B, Topolski TD Economic Evaluation of Aquatic Exercise for Persons With Osteoarthritis; Medical Care; 2001; 39/5: 413-424

Reinhardt UE Abstracting from Distributional Effects, This Policy is Efficient. In: Barer ML, Getzen TE, Stoddart GL *Health, Health Care and Health Economics*; John Wiley&Sons; Chichester, England; 1998;1-52

Richardson J et Nord E The Importance of Perspective in the Measurement of Quality-adjusted Life Years; Medical Decision Making; 1997; 17:33-41

Schlander M Kosteneffektivität und Ressourcenallokation: Gibt es einen normativen Anspruch der Gesundheitsökonomie? In: Kick HA, Taupitz J (Hrsg.):

Gesundheitswesen zw. Wirtschaftlichkeit und Menschlichkeit; LIT-Verlag; Münster 2005; S. 37-112

Schrauder A Erstellen eines Methodenhandbuches: Die Evaluation des Faktors “Kosten” bei vergleichenden Studien in der Osteopathie; Diplomarbeit; C.O.E. Osteopathie Colleg München; 2001

Schwartz FW et Dörning H Evaluation von Gesundheitsleistungen. In: Andersen H, Henke KD, Schulenbeurg JM Graf von (Hrsg.) Basiswissen Gesundheitsökonomie (Bd. 1) Berlin: edition sigma;1992; 175-200

Sen A The possibility of Social Choice; The American Economic Review; 1999; 89(3):357ff

Siegel JE et al. for the Panel on Cost-Effectiveness in Health and Medicine 1996; Recommendations for reporting cost-effectiveness analyses; JAMA: 276 p 1339-41

Skargren EI, Oberg BE, Carlsson PG, Gade M Cost-effectiveness of chiropractic and physiotherapy treatment for low back pain and neck pain: six-month follow-up; Spine; 1997; 22: 2167-77

Skargren EI, Carlsson PG, Oberg BE One-year follow-up comparison of the cost and effectiveness of chiropractic and and physiotherapy as primary management for back pain; Spine; 1998; 17:1875-84

Suomi R, Lindauer S Effectiveness of Arthritis Foundation aquatic program on strength and range of motion in women with arthritis; J Aging Phys Activ; 1997; 5:341-51

Sutton M How to get best health outcome for a given amount of money; BMJ; 1997; 315:47-49

Szucs T Medizinische Ökonomie – Eine Einführung. Urban und Voges GmbH; München;1997; ISBN 3-86094-089-9

Templeton MS, Booth DL, O'Kelly WD Effects of aquatic therapy on joint flexibility and functional ability in subjects with rheumatic disease; J Orthop Sports Phys Ther; 1996; 23(suppl6):376-81

Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y et Wang Q Multiattribute utility function for a comprehensive health status classification system: health utilities index mark 2; Medical care; 1996; 34:702-22

Torrance GW Preferences for Health Outcomes and Cost-Utility Analysis; American Journal Man Care; 1997; 3(Suppl):8-20

UK BEAM Trial Team United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care; BMJ; 2004; 329:1377-; BMJ, doi:10.136/bmj.38282.669225.AE (published 29 November 2004)

UK BEAM Trial Team United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care; BMJ,doi:10.1136/bmj.38282.607859.AE (published 29 November 2004)

Van Tulder MW, Koes BW, Bouter LM Conservative treatment of acute and chronic non-specific low back pain: a systematic review of randomised controlled trials of the most common interventions; Spine; 1997; 22:2128-56

Van Tulder MW, Malmivaara A, Esmail R, Koes B Exercise therapy for low back pain: a systematic review within the framework of the Cochrane Collaboration Back Review Group; Spine; 2000; 25: 2784-96

Waddell G, Feder G, McIntosh G, Lewis M, Hutchinson A A low back pain evidence review; London: Royal college of General Practitioners; 1999

Walker D How to do (or not to do)…Cost and Cost-effectiveness guidelines: which ones to use?; Health Policy and Planning; 2001; 16/1:113-121

Waterworth RF, Hunter EA An open study of diflunisal conservative and manipulative therapy in the management of acute mechanical low back pain; NZ Med J;1985; 98:972-5

Weinstein MC et al. 1996; Recommendations of the Panel on Cost-Effectiveness in Health and Medicine;JAMA:276 p 1253-8

Williams NH, Wilkinson C, Russell I, Edwards RT, Hibbs R, Linck P et Muntz R Randomised osteopathic manipulation study (ROMANS): pragmatic trial for spinal pain in primary care; Family practice; 2003; 20/6: 662-9; available at: www.fampra.oupjournals.org

Williams NH, Edwards RT, Linck P, Muntz R, Hibbs R, Wilkinson C, Russell I, Russell De et Hounsome B Cost-utility analysis of osteopathy in primary care: results from a pragmatic randomized controlled trial; Oxford university press, Family Practice Vol.21, No.6;2004

Zink A Methodische Aspekte der Gesundheitsökonomie; Zeitschrift für Rheumatologie; 2004; 63:5; S369-371

Internetreferenzen:

www.nice.org
www.iqwig.de
www.biqg.org
http://hta.lbg.ac.at/de/index.php
www.eurogol.org
www.pubmed.de
www.medline.de und www.medline.com
www.osteopathic-research.com
http://ostmed.hsc.unt.edu
www.cochrane.org
www.bmj.com
www.healtheconomics.com
www.healtheconomics.nl
www.wikipedia.org

http://www.ebm-netzwerk.de/grundlagen/definitionen/
www.drnawrocki.de/empfehlung/lebensqualität%20.html
www.irs-net.de/anzeigen.php?choice=monitoring&choice2=monitoring2
www.klinik-angermühle.de/buch/buchband/a-band5a.htm
http://encarta.msn.com/dictionary/effectiveness.html
http://dict.leo.org/forum
www.wiwi.uni-bielefeld.de/~wolf/lehre/ss07/statan/bootsrap.pdf

13 Annex

13.1 Schools of thought within health economics

This chapter wants to present additional economic underpinnings which are sometimes mentioned in publications. For the understanding of health economic analyses, however, it is not necessary to explain the economic theories which are the basis of health economics. Thus this chapter is included in the annex to this paper because it only represents complementary knowledge.

13.1.1 Welfare economics

Welfare economics is a branch of economics that uses microeconomic techniques to simultaneously determine allocative efficiency within an economy and the income distribution associated with it. It analyzes social welfare, however measured, in terms of economic activities of the individuals that comprise the theoretical society considered. As such, individuals, with associated economic activities, are the basic units for aggregating to social welfare, whether of a group, a community, or a society, and there is no "social welfare" apart from the "welfare" associated with its individual units. Here, 'welfare' in its most general sense refers to well-being.
http://en.wikipedia.org/wiki/Welfare_economics

As first representative of this branch of economics Kenneth J. Arrow received the nobel prize in economics in 1972; in 1998 also Amartya Sen was awarded.
http://de.wikipedia.org/wiki/Wohlfahrts%C3%B6konomie

The claim of classical welfare economics is that it brings together the utility of the individual persons in a way that the result is an optimum for society. Utility maximization (also referred to as satisfaction of needs by Hoyer und Rettig (1983)) is paramount. Thus it is assumed that utility is cardinal, i.e. measurable on a scale and thus can be compared among persons. (Lauterbach et al., 2006; Schlander M, 2005; Gold MR, 1996; Leidl 2003)

Utility can best be determined by the individuals themselves. This corresponds with the overall concept of welfare economics: consumer sovereignty (sometimes also called “self-determination criterion”). (Schlander M, 2005)

According to the principle of utility maximization the instrument of choice in welfare economics is the CBA (cost-benefit analysis) in which the utility is represented in monetary units. (Breyer et al., 2005)

Every individual thus contributes to the collective decision with his/her individual utility. (Breyer F et Kolmar M, 2001)

This would mean that it could be easily determined who would benefit most from a specific intervention.

It could be determined that a treatment provides a utility gain of 1.0 for patient A and 2.0 for patient B. Thus the utility gain for the whole society would be the sum of the individual utility gains. (Lauterbach et al. 2006)

Welfare economics postulates that the total utility of a person has to be considered as basis for decision-making. The welfare of a person, however, does not depend only on health but also on other goods. For collective decisions thus the overall welfare needs to be considered. (Breyer et al., 2005)

There were, however, immediate objections to these basic assumptions of classical welfare economics. One of the first critics of this approach was the Italian economist Vilfredo Pareto (1848-1923). He objected to the cardinal measurability of utility and to the comparability of patients with the same condition. Could not the renunciation of a pain killer have more serious consequences than the renunciation of an operation? Thus the Pareto criterion was developed. (Lauterbach et al. 2006)

13.1.1.1 Pareto criterion

It indicates that an outcome is more efficient if at least one person is made better off and nobody is made worse off. (Lauterbach et al., 2006)

The evaluation of an individual's utility can be carried out through e.g. direct or indirect measurement of the willingness to pay. However, the inevitable result is that wealthy people naturally have a better willingness to pay than less wealthy persons. (Schlander M, 2005)

Since the Pareto criterion was hardly practicable because it did not tolerate that any person is made worse off, it was further developed into the Kaldor – Hicks criterion (also referred to as potential Parteo criterion). (Lauterbach et al., 2006; Schlander M, 2005)

13.1.1.2 Kaldor-Hicks criterion

According to the Kaldor-Hicks criterion some people can be left worse off if the benefit for others is accordingly higher to compensate the loss. This compensation does not necessarily be effected, it merely requires that the possibility for compensation exists.

The theory says that even if someone who is made worse off does not receive compensation, the overall situation of society can still improve. (Lauterbach et al., 2006; Schlander M, 2005)

Theoretically a punch in the nose of someone can have a welfare increasing effect if this person was offered an accordingly higher compensation - regardless of whether the person receives the compensation. (Reinhart UE, 1998)[9]

In the technical sense thus also a less desirable change can be “efficient”. (Schlander M, 2005)

A particularity in the health care system is that a compensation for a loss of utility (e.g. through service that is not provided) can theoretically take on the form of money but usually not another form of health care service. This contradicts the exchangeability of health and any other form of consumption postulated by welfare economics. (Schlander M, 2005)

13.1.1.3 Comment

The individual preferences of a person are regarded as the predetermined starting point for the theory of welfare economics. On the basis of these preferences – whose origin and value often are not further examined – welfare economics deduces individual benefit and thus also welfare of the whole society. (Schlander M, 2005)

“The economist has little to say about the formation of wants; this is the province of the psychologist. The economist’s task is to trace the consequences of any given set of wants.” (Friedmann M, 1962, p. 13)

Culyer AJ (1991) looks at the problem of different preferences and uses an example of Amartaya Sen for better illustration:

A society consisting of two persons is confronted with the problem that one person suffers from hunger while the other has food in abundance. This can result in two possible scenarios (according to the welfare economics approach):

a) Re-distribution of the food supply

or

b) The hungry sadist tortures the filled up non-masochist

Actually both results have to be evaluated equally.

[9] Reinhardt UE, 1998 describes the Kaldor-Hicks criterion also as “punch-in-the-nose-criterion”

This should illustrate the shaky foundations on which the strictly normative approach of welfare economics is built.

"It becomes problematic if economists deduce prescriptive assumptions on "efficiency" without explaining the theoretical basics and the technical meaning of the term." (Schlander M, 2005, p. 48)

13.1.2 Extrawelfarism

Extrawelfarism tries to respond to the deficits and problems of welfare economics. It does that in particular by separating health and income or consumption. The evaluation focuses on the effects on health and no longer uses monetary units as substitutes. The effect on health is measured in natural units and is no longer considered as dependent on the willingness or ability to pay. (Schlander M, 2005)
In addition, the evaluation no longer focuses on the maximization of welfare but on the maximization of health and with regard to the consumers/patients not demand is the decisive factor but their needs. (Hurley J, 2000)
Extrawelfarists rate a certain treatment result always identically regardless of income, social status, age, ethnic origin, religion, etc. of the patient in question. The measure they use to assess effectiveness is the QALY, which is always weighted equally. (Schlander M, 2005)

"A QALY is a QALY is a QALY – regardless of who gains and who loses it." (Boyle et al. 1983; Williams, 1985; Feeny et Torrance 1989)
The extrawelfarism apporach thus uses the CUA (cost-utility analysis) because it puts the emphasis on health and not on the benefit like the CBA used in welfare economics. (Breyer et al., 2005)

13.2 Questionnaires

13.2.1 Health Utilities Index mark 2 classification system

Attribute	Level	Level description
Sensation	1	Ability to see, hear, and speak normally for age
	2	Requires equipment to see or hear or speak
	3	Sees, hears, or speaks with limitation even with equipment

	4	Blind, deaf, or mute
Mobility	1	Able to walk, bend, lift, jump, and run normally for age
	2	Walks, bends, lifts, jumps, or runs with some limitations but does not require help
	3	Requires mechanical equipment (such as canes, crutches, braces, or wheelchair) to walk or get around independently
	4	Requires the help of another person to walk or get around and requires mechanical equipment as well
	5	Unable to control or use arms and legs
Emotion	1	Generally happy and free from worry
	2	Occasionally fretful, angry, irritable, anxious, depressed, or suffering "night terrors"
	3	Often fretful, angry, irritable, anxious, depressed, or suffering "night terrors"
	4	Almost always fretful, angry, irritable, anxious, depressed
	5	Extremely fretful, angry, irritable, anxious, or depressed usually requiring hospitalisation or psychiatric institutional care
Cognition	1	Learns and remembers schoolwork normally for age
	2	Learns and remembers schoolwork more slowly than classmates as judged by parents and/or teachers
	3	Learns and remembers very slowly and usually requires special educational assistance
	4	Unable to learn and remember
Self-care	1	Eats, bathes, dresses, and uses the toilet normally for age
	2	Eats, bathes, dresses, and uses the toilet independently with difficultly
	3	Requires mechanical equipment to eat, bathe, dress, or use the toilet independently
	4	Requires the help from another person to eat, bathe, dress, or use the toilet
Pain	1	Free of pain and discomfort

	2	Occasional pain. Discomfort relieved by non prescription drugs or self-control activity without disruption of normal activities
	3	Frequent pain. Discomfort relieved by oral medicines with occasional disruption of normal activities
	4	Frequent pain, frequent disruption of normal activities. Discomfort requires prescription narcotics for relief
	5	Severe pain. Pain not relieved by drugs and constantly disrupts normal activities
Fertility*	1	Able to heave children with a fertile spouse
	2	Difficultly in heaving children with a fertile spouse
	3	Unable to heave children with a fertile spouse

*Fertility attribute can be deleted if not required. Contact developers for details. From Torrance *et al.* (1996*a*), Table 1.

13.2.1.1 Health Utilities Index mark 2 scoring formula

Sensation		Mobility		Emotion		Cognition		Self-care		Pain		Fertility	
x_1	b_1	x_2	b_2	x_3	b_3	x_4	b_4	x_5	b_5	x_6	b_6	x_7	b_7
1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00
2	0.95	2	0.97	2	0.93	2	0.95	2	0.97	2	0.97	2	0.97
3	0.86	3	0.84	3	0.81	3	0.88	3	0.91	3	0.85	3	0.88
4	0.61	4	0.73	4	0.70	4	0.65	4	0.80	4	0.64	4	n/a
5	n/a	5	0.58	5	0.53	5	n/a	5	n/a	5	0.38	5	n/a

Formula: $u^* = 1.06(b_1 \times b_2 \times b_3 \times b_4 \times b_5 \times b_6 \times b_7) - 0.06$ where u^* is the utility of the health state on a utility scale where dead has a utility of 0.00 and health a utility of 1.00. Because the worst standard error of u^* for estimating validation states within the sample is 0.015 for measurement error and sampling error, and 0.06 if model error is also included. x_i is attribute level code for attribute *i; b;* is level score for attribute *i*.
From Torrance *et al.* (1996a).

13.2.2 Health utilities Index mark 3 classification system

Attribute	Level	Level description
Vision	1	Able to see well enough to read ordinary newsprint and recognise a friend on the other side of the street, without glasses and contact lenses
	2	Able to see well enough to read ordinary newsprint and recognise a friend on the other side of the street, but with glasses

	3	Able to read ordinary newsprint with or without glasses but unable to recognise a friend on the other side of the street, even with glasses
	4	Able to recognise a friend on the other side of the street with or without glasses but unable to read ordinary newsprint, even with glasses
	5	Unable to read ordinary newsprint and unable to recognise a friend on the other side of the street, even with glasses
	6	Unable to see at all
Hearing	1	Able to hear what is said in a conversation with at least tree other people, without a hearing aid
	2	Able to hear what is said in a conversation with one other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least three other people
	3	Able to hear what is said in a conversation with one other person in a quiet room with a hearing aid, and able to hear what is said in a group conversation with at least tree other people with a hearing aid
	4	Able to hear what is said in a conversation with one other person in a quiet room without a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid
	5	Able to hear what is said in a conversation with one other person in a quiet room with a haring aid but unable to hear what is said in a group conversation with at least three other people even with a hearing aid
	6	Unable to hear at all
Speech	1	Able to be understood completely when speaking with strangers or friends
	2	Able to be understood partially when speaking with strangers but able to be understood completely when speaking to people who know me well

	3	Able to be understood partially when speaking with strangers or people who know me well
	4	Unable to be understood when speaking with strangers but able to be understood partially by people who know me well
	5	Unable to be understood when speaking to other people (or unable to speak at all)
Ambulation	1	Able to walk around the neighbourhood without difficultly, and without walking equipment
	2	Able to walk around the neighbourhood with difficultly; but does not require walking equipment or the help from another person
	3	Able to walk around the neighbourhood with walking equipment, but without the help of another person
	4	Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood
	5	Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood
	6	Cannot walk at all
Dexterity	1	Full use of two hands and ten fingers
	2	Limitations in the use of hands or fingers, but does not requires special tools or help of another person
	3	Limitations in the use of hands or fingers, is independent with use of special tools (does not require the help of another Person)
	4	Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with use of special tools)
	5	Limitations in use of hands or fingers, requires the help of another person for most tasks (not independent even with use of special tools)
	6	Limitations in use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools)
Emotion	1	Happy and interested in life

	2	Somewhat happy
	3	Somewhat unhappy
	4	Very unhappy
	5	So unhappy that life is not worthwhile
Cognition	1	Able to remember most things, think clearly, and solve day-to-day problems
	2	Able to remember most things, but heave a little difficultly when trying to think and solve day-to-day problems
	3	Somewhat forgetful, but able to thing clearly and solve day-to-day problems
	4	Somewhat forgetful, and have a little difficultly when trying to think or solve day-to-day problems
	5	Very forgetful, and have great difficultly when trying to think or solve day-to-day problems
	6	Unable to remember anything at all, and unable to think or solve day-to-day problems
Pain	1	Free of pain and discomfort
	2	Mild to moderate pain that prevents no activities
	3	Moderate pain that prevents a few activities
	4	Moderate to severe pain that prevents some activities
	5	Severe pain that prevents most activities

13.2.2.1 Health Utility Index mark 3 scoring formula

Vision		Hearing		Speech		Ambulation		Dexterity		Emotion		Cognition		Pain	
x_1	b_1	x_2	b_2	x_3	b_3	x_4	b_4	x_5	b_5	x_6	b_6	x_7	b_7	x_8	b_8
1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00
2	0.98	2	0.95	2	0.94	2	0.93	2	0.95	2	0.95	2	0.92	2	0.96
3	0.89	3	0.89	3	0.89	3	0.86	3	0.88	3	0.85	3	0.95	3	0.90
4	0.84	4	0.80	4	0.81	4	0.73	4	0.76	4	0.64	4	0.83	4	0.77
5	0.75	5	0.74	5	0.68	5	0.65	5	0.65	5	0.46	5	0.60	5	0.55
6	0.61	6	0.61	6	n/a	6	0.58	6	0.56	6	n/a	6	0.42	6	n/a

Formula (dead-perfect health scale): $u^* = 1.371(b_1 \times b_2 \times b_3 \times b_4 \times b_5 \times b_6 \times b_7 \times b_8) - 0.371$, where u^* is the utility of the health state on a utility scale where dead is a utility 0.00 and healthy ha a utility of 1.00. State worse than dead has negative utilities. x_i is attribute level code for attribute *I;* b_i is level score for attribute *i*. For the attribute "Cognition", the score for level 3 is greater than the score for level 2. This is not a typo, but reflects that level 3 was seen as preferable to level 2.

The standard error of u^*, including model error, is 0.08 for estimating validation states within the sample.

For estimating validation states (based on $n = 73$) from an independent sample, the standard error is 0.10 if the states are

unweighted, 0.006 if the states are weighted by prevalence excluding the state of perfect health, and 0.004 if the states are weighted by prevalence including the state of perfect health.

From Feeny *et al.* (2002), Table 3.

13.2.3 EQ-5D classification system

Mobility

1. No problems walking
2. Some problem walking about
3. Confined to bed

Self-care

1. No problems with self-care
2. Some problems washing or dressing self
3. Unable to wash or dress self

Usual activities

1. No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)
2. Some problems performing usual activities
3. Unable to perform usual activities

Pain/discomfort

1. No pain or discomfort
2. Moderate pain or discomfort
3. Extreme pain or discomfort

Anxiety/depression

1. Not anxious or depressed
2. Moderately anxious or depressed
3. Extremely anxious or depressed

Note: For convenience each composite health state has a five-digit code number relating to the relevant level of each dimension, with the dimensions always listed in the order given above. Thus 11223 means

1 No problems walking about

1 No problems with self care

2	Some problems performing usual activities
2	Moderate pain or discomfort
3	Extremely anxious or depressed

From Dolan *et al.* (1995), Fig. 1.

13.2.3.1 EQ-5D scoring formula

Coefficients for TTO tarifs

Dimension	Coefficient
Constant	0.081
Mobility	
level 2	0.069
level 3	0.314
Self-care	
level 2	0.104
level 3	0.214
Usual activity	
level 2	0.036
level 3	0.094
Pain/discomfort	
level 2	0.123
level 3	0.386
Anxiety/depression	
level 2	0.071
level 3	0.236
N3	0.269

From Dolan *et al.* (1995), Table 1.

EuroQol time trade-off scores are calculated by subtracting the relevant coefficients from 1.000. The constant term is used if there is any dysfunction at all. The N3 term is used if any dimension is at level 3. The term for each dimension is selected based on the level of that dimension. The algorithm for computing the tariff is quite straightforward. For example, consider the state 11223:

Full health	= 1.000
Constant term (for any dysfunctional state)	- 0.081
Mobility (level 1)	- 0
Self-care (level 1)	-0
Usual activities (level 2)	-0.036
Pain or discomfort (level 2)	-0.123
Anxiety or depression (level 3)	-0.236
N3 (level 3 occurs within at least one dimension)	-0.269
Therefore, the estimated value for 11223	= 0.255

13.2.4 SF-6D classification system

Physical functioning

1. Your health does not limit you in *vigorous activities*.
2. Your health limits you a little in *vigorous activities.*
3. Your health limits you a little in *moderate activities.*
4. Your health limits you a lot in *moderate activities.*
5. Your health limits you a little in *bathing and dressing.*
6. Your health limits you a lot in *bathing and dressing.*

Role limitations

1. You have no problem with your work or other regular daily activities as a result of your physical health or any emotional problems.
2. You are limited in the kind of work or other activities as a result of your physical health.
3. You accomplish less than you would like as a result of your physical problems.
4. 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.

Social functioning

1. Your health limits you in social activities *none of the time.*
2. Your health limits you in social activities *a little of the time.*
3. Your health limits you in social activities *some of the time.*
4. Your health limits you in social activities *most of the time.*
5. Your health limits you in social activities *all of the time.*

Pain

1. You have no pain
2. You have pain but it does not interfere with your normal work (both outside the home and housework).
3. You have pain that interferes with your normal work (both outside the home and housework) *a little bit.*
4. You have pain that interferes with your normal work (both outside the home and housework) *moderately.*
5. You have pain that interferes with your normal work (both outside the home and housework) *quite a little bit.*
6. You have pain that interferes with your normal work (both outside the home and housework) *extremely.*

Mental health

1. You feel tense or downhearted and low none of the time.
2. You feel tense or downhearted and low a little bit of the time.
3. You feel tense or downhearted and low some of the time.
4. You feel tense or downhearted and low most of the time.
5. You feel tense or downhearted and low all of the time.

Vitality

1. You have a lot of energy all of the time.
2. You have a lot of energy most of the time.

3. You have a lot of energy some of the time.

4. You have a lot of energy none of the time.

From Brazier *et al.* (2002), Table 1.

13.2.4.1 SF-6D utility scoring model

General terms		Physical functioning		Role limitations		Social functioning		Pain		Mental health		Viality	
Term	Score	Level	Score	Level	Score	Level	Score	Level	Score	Level	Score	Level	Score
C	1000	PF1	-0.000	RL1	-0.000	SF1	-0.000	PAIN 1	-0.000	MH1	-0.000	VIT1	-0.000
MOST	-0.070	PF2	-0.053	RL2	-0.053	SF2	-0.055	PAIN 2	-0.047	MH2	-0.049	VIT2	-0.086
		PF3	-0.011	RL3	-0.055	SF3	-0.067	PAIN 3	-0.025	MH3	-0.042	VIT3	-0.061
		PF4	-0.040	RL4	-0.050	SF4	-0.070	PAIN 4	-0.056	MH4	-0.109	VIT4	-0.054
		PF5	-0.054			SF5	-0.087	PAIN 5	-0.091	MH5	-0.128	VIT5	-0.091
		PF6	-0.111					PAIN 6	-0.167				

Utility = C + PF + RL + SF + PAIN + MH + VIT + MOST

where utility = utility on 0-1 dead healthy scale, C = constant item, PFx = level x on the physical functioning dimension, same for other dimensions, MOST = term to use if any dimension is at its most severe level.

From Brazier *et al.* (2002), Table 6, Column 5 (Model 10).