Critical appraisal of economic evaluations

Quantifying the quality of economic evaluations on obesity

Zòmelie Kortram *272598*
Master Thesis
May 2007

Master Health Economics, Policy and Law

Acknowledgements

This thesis was written as a final assignment in order to graduate with a degree in Master Health Economics, Policy and Law.

I would like to thank my tutor, Dr. Ken Redekop, for supporting me throughout the past year and giving me the necessary guidance in order to complete this thesis. I would also like to thank L. Niessen and C. Bouwmans for co-evaluating this thesis.
I am also grateful to my parents for enabling me to study at the Erasmus University Rotterdam and for giving me the mental support throughout the years.

Abstract

Objective: Various economic evaluations have been conducted on interventions aimed at decreasing obesity. These interventions are changing dietary patterns, behavioural modification, increasing physical activity, surgery or a programme including the previous mentioned interventions.
Yet it is hard to assess what the actual quality is of these economic evaluations as most assessment tools are of a qualitative nature. In this thesis the quality of the economic evaluations on obesity is assessed in a quantitative manner.
Methods: The economic evaluations (abstracts) on obesity are selected in the NHS Economics Evaluations Database and the quality of these studies is assessed by using the assessment tool of Chiou et al. (2003). This quality assessment tool consists of sixteen criteria which all have their own weight. If a study could fulfil a criterion which was set, the entire weight was appointed. If the criterion could be addressed partially, half of the weight was appointed. None of the weight was appointed if the study could not fulfil the criterion. The sum of the weight of the sixteen criteria reflects the quality of the economic evaluation.
To test the association between the characteristics of the economic evaluations and the quality, univariate, bivariate and multivariate analyses were performed. The characteristics are the year of publication, the category of intervention, the type of study, the country in which the study was performed, the location in which the study was performed and the resource of funding.
Results: In the NHS EED 34 economic evaluations on interventions aimed at decreasing obesity were selected. These studies were published from 1995 until 2006 and had a mean quality score of 51.9 with a range of 24.1 – 75.8 (SD 13.28).
Most studies were not able to fulfil the criteria entirely, because valuable information necessary to draw a conclusion was missing, the choice of the economic model was not discussed, statistical or sensitivity analyses were not performed and the resource of funding was not mentioned.
Multivariate analyses showed that the year of publication is positively associated with the quality of the publication, meaning that if time progresses with one year, the quality of the article increases with 1.5 points. The location in which the study was performed was also associated with the quality of the publications. European publications on obesity interventions had a higher quality score, of 14.46 points on average, than non-European countries and thus were of better quality.
In the univariate analyses the type of study was associated with the quality of the economic evaluations. The resource of funding was not associated with the quality of the studies, but articles which were funded by pharmaceutical companies did have a higher score than the studies funded by health research organisations. The category of intervention did not have an influence on the quality of the study.

Statistical analyses show a correlation between certain items and the characteristics, but the associations are mostly due to a natural association with the characteristics. A separate item cannot determine the quality of the publication.

Conclusion: The main conclusions which can be drawn are the following:

- The quality of the economic evaluations on obesity increases as time progresses.
- Studies which are based on previous publications have a higher quality than those which are based on a single study.
- The location in which an evaluation is performed also affects the quality of the evaluation.
- A single criterion of the assessment tool can yet not be used separately to determine the quality of the economic evaluation.

The mean quality score of the articles shows that on average the economic evaluations on obesity interventions are not of high quality. But the use of the quality assessment tool can give more insight in the areas in which improvements can be made.

The quality assessment tool can be useful before a research is conducted or an article is submitted, but afterwards the tool is also useful to assess the quality of the actual publication. This tool can also affect current policies and improve the way in which interventions aimed at decreasing obesity are implemented in the current interventions programmes.

Discussion: There are several recommendations which can be made. More information is needed on the type of scale which is most useful when appointing the weight. It can also be relevant to know what information can be gathered when the assessment tool is used in other disease areas. Perhaps it would be useful to determine if the assessment tool can be adjusted to the disease or risk factor which is examined.

Index

Introduction ... 6
1. Obesity ... 8
1.1 Description of obesity ... 8
1.2 Causes of obesity ... 9
1.3 Costs and effects ... 10
1.4 Prevention and treatment strategies ... 12
2. Methods to evaluate quality ... 14
2.1 Grading system ... 14
2.2 The grading system and disease areas ... 15
3. Methods ... 16
3.1 NHS Economics Evaluation Database (NHS EED) ... 16
3.2 General rules ... 17
3.3 Methods of analyses ... 18
4. Results ... 20
4.1 Database ... 20
4.2 Description of the economic evaluations ... 21
4.3 Description of quality scores ... 22
4.4 Characteristics of the economic evaluations ... 24
5. Analysis ... 28
5.1 Relationship between study characteristics and total quality score ... 28
5.2 Relationship between variables and item score ... 32
Conclusion ... 35
Discussion ... 37
References ... 43
Annex 1 ... 45
Criteria selected for the Grading System (Chiou et al. 2003) ... 45
Estimated weights for criteria (Chiou et al. 2003) ... 45
Annex 2 ... 46
Summaries of NHS EED ... 46
Annex 3 ... 58
Quality scores of economic evaluations ... 58

Introduction

In this thesis the focus will be on obesity and economic evaluations. Key in this thesis was to determine in what way economic evaluations could benefit from quantifying qualitative assessments of these economic evaluations.

Obesity is considered to be a risk factor for multiple diseases, such as diabetes type II and heart diseases. Since the prevalence of obesity is increasing dramatically in today's society, it is considered to be a problem that is in need for more attention and solutions which can decrease the prevalence of obesity. Some of the solutions are provided by economic evaluations, where several treatment options are tested and cost-effectiveness is determined.
However, there are a lot of different treatments and programmes and it is not always clear, especially from a policy making view, which studies have a better quality and are more useful for implementation in practice. By measuring the quality of these cost-effectiveness studies, these studies could be of more use in determining whether a certain treatment or programme could be useful in decreasing obesity.

There are several ways in which the quality of economic evaluations can be assessed, but most of these methods are more of a qualitative nature.
This thesis will show in what way qualitative assessments of economic evaluations on interventions aimed at decreasing obesity can be quantified in order to lead to better decision-making.

The main questions of this thesis are:
How can the quality of economic evaluations be assessed and how good is the quality of economic evaluations of interventions aimed at decreasing obesity?

The subquestions of this thesis were:

- *What is obesity?*
- *What are the consequences of obesity?*
- *Which strategies are currently used to prevent and/or treat obesity?*
- *Which methods exist to quantify the quality of the economic evaluations and which of these methods are included in this thesis?*
- *Which economic evaluations have been performed regarding the prevention and treatment of obesity?*
- *Which associations can be found between the quality of economic evaluations and the characteristics of these economic evaluations when a validated grading system is used?*
- *Which recommendations can be made to improve the quality of similar studies?*

The thesis is subdivided as follows. In the next chapter ’Obesity’ the most relevant information on obesity will be discussed, such as what obesity really is and what the possible consequences of obesity are. Several treatment options will also be discussed.

The quality of the economic evaluations on obesity was assessed by using a grading system. Therefore the grading system and case studies in which these grading systems were used will be presented in the second chapter.

In the chapter ‘Methods’ the methods which are used to select the economic evaluations in this thesis and methods in which the quality assessment tool will be used to quantify the quality of these studies will be discussed. The studies taken into account in this thesis are selected, by using the NHS Economics Evaluation Database, known for its accurate collection of economic evaluations on multiple subjects. The grading system tool was presented by Chiou et al. (2003) as a ‘practical, quantitative tool for appraising the quality of cost-effectiveness studies’ and consists of criteria which are used to grade cost-effectiveness studies.

The results of the selection and grading of the articles according to the assessment tool will be presented and the results will then be analysed. A descriptive analysis of the quality of the articles will be given, after which several characteristics of the study will be distinguished to determine the association between these characteristics and the quality of the studies.
The association between the criteria of the grading system en the quality of the articles will also be determined.

This thesis will be ended with a conclusion and a discussion in which the limitations of this study will be discussed and recommendations will be made for future research.

1. Obesity

In the past decades the number of obese people has exploded and this has a large effect on health and health care. In this section some of the details related to obesity will be discussed and a general overview of the most relevant aspects related to obesity will be given.
Several aspects will be considered such as what obesity actually is, in which way it is related to other diseases, how it can be prevented or treated and which costs are related to obesity.

1.1 Description of obesity

Obesity can be described as a disorder where a person has extreme overweight. There is an international system for the classification of overweight and obesity in adults (WHO 2000). This classification is based on the Body Mass Index (BMI) which can be calculated by dividing the weight (in kilograms) by height in m^2. A BMI over 25 kg/m^2 is defined as overweight.
A person is obese when he/she has a BMI over 30 kg/m^2 (Pushka et al. 2003). Obesity itself is also categorised. A BMI between 30.0 and 34.9 is considered *moderate*. A BMI between 35.0 and 39.9 is considered *severe* and a BMI of ≥40 is categorised as *very severe* (WHO 2000).

In most studies BMI is used as an indicator for obesity (Battacharya & Bundorf 2005). But there are other measures that can indicate obesity, such as the waistline (WHO 2000).

Overweight and obesity lead to adverse metabolic effects on blood pressure and cholesterol and can lead to non-fatal health problems such as respiratory difficulties and musculosketal problems.
More life-threatening problems such as type 2 diabetes, cardiovascular diseases, hypertension and stroke and certain types of cancer, can also occur due to overweight and obesity (Pushka et al. 2003). As obesity is related to so many disorders and treatments, it is considered to be a risk factor.
Yet, one can also state that obesity is a disease due to its severity and psychological background. In most studies however, obesity is treated as a risk factor for other disorders or diseases and not as a disease itself.

Obesity can affect all individuals, but certain individuals have a higher risk of developing obesity. Among these individuals you can consider children for both genetic and environmental reasons. Especially when one parent is obese, a child is more at risk of becoming obese.

In the past view years obesity rates have increased rapidly and obesity is becoming more of an epidemic. In 2005 the WHO estimated that 1.6 billion adults (age 15+) are overweight of whom at least 400 million adults are obese. It is expected that by 2015, 700 million adults will be obese and 2.3 billion adults will be overweight (WHO 2006).

Over 2.5 million deaths can be attributed to a higher BMI. This figure is expected to double by 2030 (IOTF 2006). However the increase of obesity also differs across continents. In Europe and North America the mean levels of BMI are much higher than in Asia and Africa. In Europe and North America the mean levels of BMI are approximately between 25 and 27 kg/m^2, while in Africa and Asia the mean levels of BMI are between 20 and 23 kg/m^2. Even though these figures are below the threshold of 30 kg/m^2, they indicate that due to a higher BMI a large part of the world population is exposed to health risks (IDF 2006).

1.2 Causes of obesity

Genetic, environmental and socio-cultural or behavioural factors play a role in the development of obesity (WHO 1998). In recent years it has become clear that certain gene deficiencies can cause an increase in weight. Yet genetic factors are not able to explain the rapid growth of obesity across the world population.

The cause of obesity is considered to be an increased consumption of food which is more energy-dense and nutrient poor. These foods contain high levels of sugar and saturated fats. The physical inactivity of people has also contributed to the increase of obesity (Pushka et al. 2003).

Besides studies investigating the determinants of obesity, studies have researched the growth of obesity. Most of these studies are related to obesity in the US, because the prevalence of obesity has grown more rapidly there compared to other countries. Chou et al. (2004) found that obesity rates of the US have a strong positive association with the per capita number of restaurants. They addressed the growth of obesity to the downward food prices and the emphasis on the anti-smoking campaign by the US government. Philipson (2001) provides an alternative explanation for the growth of obesity and states that this growth can be related to technological change. According to Winson (2004) political and economic factors can be part of the explanation for the increase of obesity and colonization of pseudo foods by supermarket merchandisers has added to this rapid growth of obesity. Colonization of pseudo foods relates to the focus of supermarket merchandisers on foods which contain a high level of sugar and/or fat, because these products lead to a larger profit. In a recent article Rashad (2006) states that caloric intake, activity level and smoking are the main contributors to the increase of obesity in the US.

The WHO (2006) has also categorised the determinants of obesity and the evidence of which risks decrease or increase the risk of weight gain and obesity (see table 1).

Table 1: Summary of strength of evidence on factors that might promote or protect against weight gain and obesity

Evidence	Decreased risk	No relationship	Increased risk
Convincing	Regular physical activity High dietary intake of NSP (dietary fibre)[b]		Sedentary lifestyles High intake of energy-dense micronutrient-poor foods[c]
Probable	Home and school environments that support healthy food choices for children[d] Breastfeeding		Heavy marketing of energy-dense foods[d] and fast-food outlets[d] High intake of sugars-sweetened soft drinks and fruit juices Adverse socioeconomic conditions[d] (in developed countries, especially for women)
Possible	Low glycaemic index foods	Protein content of the diet	Large portion sizes High proportion of food prepared outside the home (developed countries) "Rigid restraint/periodic disinhibition" eating patterns
Insufficient	Increased eating frequency		Alcohol

[a] *Strength of evidence: the totality of the evidence was taken into account. The World Cancer Research Fund schema was taken as the starting point but was modified in the following manner: randomized controlled trials were given prominence as the highest ranking study design (randomized controlled trials were not a major source of cancer evidence); associated evidence and expert opinion was also taken into account in relation to environmental determinants (direct trials were usually not available).*
[b] *Specific amounts will depend on the analytical methodologies used to measure fibre.*
[c] *Energy-dense and micronutrient-poor foods tend to be processed foods that are high in fat and/or sugars. Low energy-dense (or energy-dilute) foods, such as fruit, legumes, vegetables and whole grain cereals, are high in dietary fibre and water.*
[d] *Associated evidence and expert opinion included.*

1.3 Costs and effects

Pushka et al. (2003) recognize that the actual costs of obesity have not been calculated, because of the complexity. To calculate the actual costs, costs of illnesses caused by obesity also have to be taken into account.

In fact, significant cost- effectiveness can be gained if one considers the costs which are spent on obesity today. The costs of obesity are expected to be between 2% and 7% of all health care costs (IDF 2006). In 1995 it was estimated that in the US a total of USD 99.2 billion was spent on overweight and obese people. USD 51.6 billion was spent on direct costs such as preventative, diagnostic and treatment services. USD 47.6 billion was spent on indirect costs such as the value of wages lost by people who were unable to work due to illness or disability or premature deaths. Besides the costs, it was also estimated that approximately 280.000 deaths occur annually due to obesity (Vaidya 2006).
In 2001 the combined direct and indirect costs were USD 123 billion (IDF 2006).

In other countries the health care costs due to obesity are not as large as those of the USA, but one also has to take into account the costs related to the reduced quality of life and life expectancy due to obesity. When the BMI-level is increased, the global burden of disease due to obesity, expressed as a percentage of disability-adjusted life years (DALY's), can increase up to 16% of the total burden of disease (IDF 2006). This indicates that the total burden of disease due to obesity can increase in all countries, even when the health care costs due to obesity are lower.

Table 2: Obesity Costs in Relation to the Co-Morbidities (1999 dollars in billions) Source: The Lewin Group, 1999. (AOA 2006)

Disease	Direct Cost of Obesity	Direct Cost of Disease	Direct Cost of Obesity as a Percentage of Total Direct Cost of Disease
Arthritis	$7.4	$23.1	32%
Breast Cancer	$2.1	$10.2	21%
Heart Disease	$30.6	$101.8	30%
Colorectal Cancer	$2.0	$10.0	20%
Diabetes (Type 2)	$20.5	$47.2	43%
Endometrial Cancer	$0.6	$2.5	24%
ESRD	$3.0	$14.9	20%
Gallstones	$3.5	$7.7	45%
Hypertension	$9.6	$24.5	39%
Liver Disease	$3.4	$9.7	35%
Low Back Pain	$3.5	$19.2	18%
Renal Cell Cancer	$0.5	$1.6	31%
Obstructive Sleep Apnea	$0.2	$0.4	50%
Stroke	$8.1	$29.5	27%
Urinary Incontinence	$7.6	$29.2	26%
Total Direct Cost	**$102.2**	**$331.4**	**31%**

The Lewin Group (1999) examined the direct costs of obesity in relation to the co-morbidities. Even though some double counting could have occurred, the results still show that obesity plays a significant part in the costs of co-morbidities. The direct health care costs of obesity were estimated at $102.2 billion 1999 (see table 2) (AOA 2006).

Thus if obese people are able to maintain their strategy, the costs and mortality due to obesity can decrease. Costs can even decrease further if obese are able to lose weight without high-cost interventions, such as drugs or surgery. And if those who are overweight are able to decrease their weight it is also likely that many health care costs can be saved.

1.4 Prevention and treatment strategies

Vaidya (2006) describes the several strategies currently used to decrease obesity in individuals. These strategies include exercise, behavioural modification, diet therapy, pharmacotherapy and surgery.

Physical activity or exercise is very important when it comes to losing weight. That is why physical activity is almost always part of weight management programs (Vaidya 2006). However, the WHO state that multiple randomized trials of exercise programmes show mixed results in achieving long-term results. The focus should be on ongoing physical activity rather than previous physical activity or enrolment in exercise programmes. Therefore the recommendation of at least 30 minutes of moderate-intensity activity a day is more aimed at reducing the cardiovascular diseases and overall mortality. But to prevent unhealthy weight gain, it is likely that a lager amount of exercise is necessary. This can be placed in the range of 45 to 90 minutes a day. Currently the consensus on the amount of physical activity needed to prevent unhealthy weight gain is 45 to 60 minutes of moderate activity on most days or every day of the week (WHO 2006).

Behavioural modification or lifestyle modification relies on analysing the behaviour that is identified with inappropriate eating, exercise or thinking habits. Certain primary strategies have been found useful to help people change their behaviour. Some examples are self-monitoring, stress management and social support (Vaidya 2006). Most primary strategies are focused on increasing the knowledge on healthy diets and the negative side effects of obesity.
According to Cutler and Glaeser (2006), who studied the difference in smoking rates between Americans and Europeans, such a strategy is effective. They found that Americans have stronger beliefs about the health risks of smoking and the beliefs result in lower smoking rates among Americans.
Kan (2004) has also found a positive, significant relationship between obesity and health risk knowledge for males under the mid-range of BMI distribution. The relationship between health risk knowledge and obesity was insignificant for females at all percentiles.

There are several forms of diet therapy. They can be distinguished in Low Calorie Diets (LCD) and Very Low Calorie Diets (VLCD). VLCD can produce very large weight losses (15-25%), but the costs of these programs are very high (≈ USD 3000). VLCD are thus not very cost-effective. Usually these diet therapies combine the decrease of calorie intake with exercise and behavioural modification. This is especially the case with popular diets such as Weight Watchers and Atkins. However weight loss seems to be the greatest if individuals adhere to the diet, no matter what diet it is (Vaidya 2006).

Pharmacotherapy has been somewhat disappointing in proving cost-effectiveness as they often show poor-side effects or limited long-term efficacy. But the FDA (USA - Food and Drug Administration) has approved of two medicines, namely sibutramine and orlistat (ibid. 2006). Sibutramine has proven to be effective as those patients taking sibutramine had a likelihood of loosing 20-30% more weight than those patients taking a placebo (and loosing 5% weight). Orlistat has shown to decrease weight (compared to a placebo 5.9-10% vs. 4.6-6.4%), but has also shown that chances of regaining weight is reduced.
Currently another medicine is on the horizon; rimonabant. Studies have shown that a large weight loss can be achieved in a short period of time, but that there was also a slight incidence of adverse events. It has yet to be seen if rimonabant will be approved by the FDA (ibid. 2006).

Surgery is reserved for those individuals with a BMI of 35 or higher and is often referred to as bariatric surgery. The type of surgery performed today can be categorized in restrictive surgeries and malabsorptive surgeries. Restrictive surgeries include vertical banded gastroplasty (VBG) and adjustable silicone-banded gastric banding (ASGB).
Malabsorptive procedures include mostly gastric bypass procedures, such as Roux-en-Y (RYGB) and biliopacreatic diversion (BDP). Often patients have to modify their eating behaviour and diet to assist with weight loss and to prevent potential complications (ibid. 2006).

The cost-effectiveness of these strategies has been demonstrated for some strategies more than others, but was not convincing in all cases. What becomes clear from the literature described above is that cost-effectiveness very much depends on the *individual's* own ability to follow the strategy chosen to decrease his' or hers obesity. Especially for strategies such as physical activity or diet therapy, which partially depend on behavioural modification, the individual's own ability to be consistent is important (ibid. 2006).

2. Methods to evaluate quality

There are several methods and tools which can be used to assess the quality of economic evaluations. These methods can be grouped into questionnaires or grading systems.
The grading systems make it possible to translate the quality of study into a grade or score.
In this thesis a grading system will be used, so the focus of this chapter will be on these grading systems.

2.1 Grading system

Chiou et al. presented a quality assessment tool in 2003. This grading system is presented as a practical, quantitative tool for the appraisal of the quality of cost-effectiveness studies.
A steering committee of five health economists selected a set of criteria for the instrument from an item pool.
A conjoint analysis survey on 120 health economists was used to estimate the weight for the sixteen criteria of the grading system. For estimation of each criterion a random effects regression model was used. This led to a coefficient and a weight for every criterion (Annex 1). If a study does not apply to a certain criterion the coefficient will not be appointed, thus leading to the possibility to grade the study by adding up the weights appointed to each of the sixteen criteria.

The grading system was validated by collecting data from sixty health economists and performing statistical analysis such as Spearman rho, Wilcoxon tests and analysis of covariance (ANCOVA).
The health economists had to review three CE-studies on a visual analogue scale (VAS) and use the grading system. The Spearman rho (coefficient = 0.78, $P < 0.0001$) and Wilcoxon ($p= 0.53$) tests were used to test the convergent validity and ANCOVA ($F_{3,146} = 5.97$, $P = 0.001$) was used to test the discriminant validity of the test.

The results show that the grading system had a high correlation between the global score and the weighted score, indicating a good convergent validity. The discriminant validity was proven by the ability of the grading system to distinguish between the good quality and poor quality of an economic study. This grading system is thus of importance when making decisions on f.e. allocation of medical resources, because it creates the possibility to quantify the quality of the studies.

2.2 The grading system and disease areas

The quality assessment tool that was presented by Chiou et al. was presented as the Quality of Health Economic Studies (QHES) instrument by Ofman et al. (2003). But Ofman et al. had not studied the evaluation of this tool by health economist and they had not taken into account the use of a constant in the calculation of the grade. They did study the perceived value of 156 experts for the QHES instrument and they performed a case study to test the use of the QHES instrument in practice.

Of the 156 experts, 84 experts stated that they would use the tool or recommend it to others and 39 experts would not use or recommend the tool to others.
In the case study 30 cost-effectiveness publications in gastroesophageal reflux disease (GERD) were graded. These studies were published since 1985 and had a mean score of 63.6 with a standard deviation of 14.7. View studies stated the perspective of the study or performed an incremental analysis.
Ofman et al. concluded by establishing that the QHES would make it possible to focus on higher quality studies. At the same time the tool would also be useful for non experts, because the tool made it easier to assess the quality. The QHES would also have an impact on clinical and political decisions, because it would make it possible to use economic analyses of a higher quality in economic analyses.

Spiegel et al. (2004) also performed a systematic review to identify and assess the quality of health economic analyses on digestive diseases, similar to the case study of Ofman et al. 2003. Of the 186 economic analyses on digestive diseases, published since 1986, less than one third was able to meet the criteria for high quality studies (29%). Most studies did not include a description of the model biases and a description of the methods for deriving model assumptions was also not given.
The lower quality of most studies could be remedied according to Spiegel et al. and if editors and peer reviewers would use the results of their studies, the high quality of publications on digestive diseases could be ensured.

3. Methods

In this section a description of the methods which have been used to select and analyse the economic evaluations will be given. In the first paragraph information will be given on the database from which the economic evaluations are extracted. In the second paragraph the general rules for the selection of the studies will be discussed, after which a description of the methods for analysis of the studies will be given.

3.1 NHS Economics Evaluation Database (NHS EED)

The NHS Economic Evaluation Database contains reliable information on costs, effects and cost-effectiveness of health care interventions (e.g. drugs, treatments and procedures).
The NHS EED was set up in 1995 by the CRD and has been funded by the Departments of Health of England and Wales to assist decision-makers. It is able to do so, due to the systematic identification of papers and providing abstracts on the description of economic evaluation and quality appraisal. In comparison with databases such as Medline or PubMed, where articles are not reviewed, NHS EED can save the decision-maker a lot of time (CRD 2006).

The CRD also contains two other databases, namely DARE and HTA. The DARE database provides summaries of systematic reviews which are assessed according to strict quality criteria and also contains a critical commentary on the quality of the reviews. These reviews all have to be on the subject of effectiveness of interventions, but the content may differ from health to social care topics.

The HTA database is produced in collaboration with the INATHA Secretariat, based at the Swedish Council of Healthcare Technology Assessments and records on the projects of the members of INAHTA and other health technology assessments. Aside systematic reviews, the database also contains research based trials, questionnaires and economic evaluations.

NHS EED includes economic evaluation papers, papers on cost or burden of illness, methods of economic evaluation and reviews on economic evaluation. There is much effort put in the assessment of economic evaluations. The economic evaluations are assessed by a health economist to provide a description and a detailed evaluation of the economic evaluation.
These are referred to as abstracts and currently the NHS EED contains over 6000 abstracts of economic evaluations (ibid. 2006).

In a commentary the reliability and generalisability of the paper is also addressed. The abstract is then checked and edited by another health economist and the original authors are also invited to reply with any further information or corrections. All other papers on costs or burden of illness, methods for economic evaluations and reviews are only entered in the database with bibliographic details.

The structured abstracts will give descriptions of the following issues related to the study:

- health technology (e.g. disease, intervention, hypothesis)
- key elements such as the type of economic study, the setting, the sources of effectiveness data
- details on the clinical evidence related to e.g. the study sample and the study design
- the way in which the paper uses previously published articles and describes which methods are used
- methods used to derive estimates and key assumptions on effectiveness
- economic analysis; direct or indirect costs and the use of statistical analysis
- the way in which the costs and benefits are reported and related to each other
- the conclusion of the authors and whether or not they provided any information on issues such as the choice of the comparators, the validity of the cost and benefit estimates and other issues related to the methodology and inclusion of issues such as side-effects of treatment etc.
- implications of the study on f.e. clinical practice, health policy or research or equity implications. Links to other studies in the NHS EED and comments are also mentioned (ibid. 2006).

3.2 General rules

Several methods and search strategies were applied to select the articles which were later graded according to the criteria of Chiou et al. (2003).

The search strategies were applied in the NHS EED. Though the CRD contains other databases, such as DARE and HTA, the choice was made to use the NHS EED, because this database contains economic evaluations which are also assessed by health economists.
The DARE also contains reviews which consist of critical commentary, but these reviews are not all on cost-effectiveness of studies.

The reason why the grading system of Chiou et al. is chosen, is because even though other grading systems are also available, such as the British Medical Journal (BMJ), Canadian Collaborative Workshop for Pharmacoeconomics and the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine, these grading systems (1) have not been formally validated for internal and external validity of the economic evaluations and (2) these instruments are qualitative and do not provide the possibility to compare studies by their scores.

The general rules for selection of the economic evaluations on obesity were:

- The articles which were selected were gathered from the NHS EED.
- The record type of the articles was limited to economic evaluations.
- The search terms were based on obesity and intervention strategies.
- The search terms were also combined to decrease the number of records as much as possible.

These articles were then tested on their quality according to the grading system presented by Chiou et al. (2003). The criteria which were used to determine the quality of an article and the weights which can be appointed to each of these criteria can be seen in Annex 1.
The appointment of the weights to the articles occurred by checking the consensus between the NHS EED – abstracts and the criteria of Chiou et al. (2003).

If the criteria (questions) could be answered with 'yes', the entire weight (1) was appointed.
If the authors only addressed the criteria partially, half of the weight was appointed (0.5).
Naturally, if the authors did not address the criteria, no weights (0) were appointed for that particular criterion.
When the articles were reviewed according to all the criteria, a total amount of the weights could be appointed which reflects the quality of the article. Because half of the weight could be appointed, use of the coefficients and the constant was made to decide the final weight of the study.

3.3 Methods of analyses

In the chapter 'Results' a descriptive analysis is given of the economic evaluations on obesity interventions. This analysis contains some general information on the economic evaluations, after which the association between the quality of the articles and the criteria used by Chiou et al. (2003) is discussed.

To assess which information can be gathered by using Chiou et al. (2003), several characteristics of economic evaluations were distinguished to determine the association between these characteristics and the quality of the economic evaluations on obesity interventions.

The characteristics which were distinguished are:

- the year of publication
- the category of intervention
- the type of study
- the country in which the study was performed
- the location in which the study was performed
- the resource of funding.

To test for the relationship between the characteristics of the economic evaluations and the quality of the evaluations, several analyses were performed. SPSS was used to perform the statistical analyses. The quality of the evaluation is then equal to the sum score of the articles. To test if there was an association between the quality of the economic evaluations and a single characteristic, several univariate analyses were performed. The association between the year of publication, the type of study and the location in which the study was performed was tested by performing regression analyses. To test for the association between the other characteristics and the quality of the economic evaluations, bivariate analyses were performed. This is because in the other characteristics more than two categories can be distinguished.

The bivariate analyses might show associations with the quality of publications which are not clear when all the characteristics interact. This is why a multivariate analysis was also performed to test for the association with the quality of the publications in a model.
To test whether there is a correlation between the several variables and the items, several tests were performed. A Spearman's test was conducted to test between the score of the items and the year of publication. The Kruskal–Wallis test was used to test the correlation between the other variables and the items.

In all cases an α of 0.05 was used to assess whether there was an association between the quality of the studies and the characteristics.

In the next chapter the results of using the grading system of Chiou et al. (2003) and the grades which could be appointed to the different economic evaluations are described. A summary of all the studies included in this thesis can be found in Annex 2.

4. Results

In this chapter, the results of searching the database and the characteristics of the publications will be described. First the results gathered from selection of the economic evaluations will be discussed, after which the scores of the evaluations will be shown.

4.1 Database

The NHS EED (2006) has been used to select the articles which are included in this thesis. Using all CRD databases and obesity as a search term led to 388 records. This result contains the records of all databases, so including DARE, HTA and NHS EED. But the focus was on economic evaluations and therefore the NHS EED was searched with the search term obesity. Selecting the NHS EED led to 187 records. These records include economic evaluations, papers on cost or burden of illness, methods of economic evaluation and reviews on economic evaluations. To include only the economic evaluations on obesity interventions, the type of study had to be selected in the NHS EED as *economic evaluation*. This led to the result of 70 records (see table 3).

However, since some of the records appeared twice in this list, there were actually fewer than 70 studies which were useful. Some studies also showed a limited relationship with this thesis' subject and focused more on other diseases, disorders or treatment strategies. The studies which were useful were too minimal for the scope of this thesis, so the focus was broadened to all treatment strategies including surgery and mix intervention programmes and not merely on diet, exercise or physical activity and behaviour.

Table 3: Results of search in NHS EED

Search terms used in NHS EED	**Number of records**
Obesity & all type of studies	187
Obesity & type of study = economic evaluation (EE)	70
Obesity & Diet therapy (EE)	0
Obesity & Diet (EE)	16
Obesity & Physical Activity (EE)	3
Obesity & Behavioural modification (EE)	0
Obesity & Behaviour (EE)	15
Obesity & Exercise (EE)	28
Childhood obesity	2

In order to retrieve only the useful publications the NHS EED database was searched again and 76 records relating obesity were found. The reason why the second search resulted in more records was because the database was updated. From the previous search I learned that the records in the NHS EED database tend to overlap. That is why the records were manually selected by judging the abstracts. Again, the records were only selected if they were on economic evaluations papers.

Of the 76 records only 35 records were useful. The other 41 records related mostly to other diseases. One record was excluded, because this record related to an abstract which was not drawn up by a NHS EED health economist. The results of the remaining 34 records will be discussed in the following paragraph, using the grading system of Chiou et al. (2003).

4.2 Description of the economic evaluations

The 34 studies which were included in this thesis were published from 1995 until 2006. These articles were published in Australia and various European countries, but most of the studies were published in the USA, namely 18 of the 34 studies (see table 4).

Table 4: Number of studies per country

Countries	Studies
Australia	3
Belgium	1
Denmark	2
Finland	1
Germany	1
Greece	1
Ireland	1
The Netherlands	1
Spain	1
Sweden	1
UK	3
USA	18

The economic evaluations were based on various interventions against obesity, such as surgery, pharmacotherapy, diets, behavioural treatment and mixed intervention programs. There were articles which were more specific and focused on diabetes type 2 patients suffering from obesity.
As mentioned before, a summary of the articles and the actual results of the various studies can be found in Annex 2.

Some economic evaluations were (partially) based on studies where results from other published articles were used to gain results. These are referred to as studies based on multiple studies. Of the 34 articles included in this study, there were 21 articles which were based on a single study and 13 articles which were based on multiple studies. The funding of these studies was either from health research organisations or pharmaceutical companies (see table 5).

Table 5: Number of studies per variable

Variables	Number of studies
Categories	
Surgery	13
Medicines	4
Diets	2
Behaviour	4
Mixed Intervention programmes	4
Type 2 Diabetes	7
Year	
1995-1999	13
2000-2003	11
2004-2006	10
Location	
European countries	13
Non - European countries	21
Funding	
Research organisations	11
Pharmaceutical companies	6
Non mentioned	17
Type of studies	
Based on a single study	21
Based on multiple studies	13

The quality of the publications and the fulfilment of the criteria is discussed in the next paragraph. After which a more detailed description of the quality of the various economic evaluations will be given.
The characteristics of a publication, such as the type of intervention or study, the country in which it is published, the publishing year and the funding of the study, are used to grade the articles with the help of Chiou et al. (2003).

4.3 Description of quality scores

The 34 studies which were judged according to Chiou et al. (2003) had a mean sum score of 51.9 and a standard deviation of 13.28. The minimum sum score of all publications was 24.1 and the maximum score of all publications was 75.8. As mentioned before in paragraph 2.3, when complete fulfilment of a criterion of Chiou et al. 2003 had occurred, the entire weight would be appointed. This would be reflected in an item score of (1). Partial fulfilment of the criterion would be reflected in an item score of (0.5) and half of the weight would be appointed. No fulfilment of the criterion would be reflected in an item score of (0), in which case none of the weight would be appointed. The average relative sum score of the criteria or items per abstract was 9.3.
The total score for each item was calculated by a sum of (1), (0.5) or (0), depending on whether an evaluation was able to fulfil the criterion set by Chiou et al. (2003). The total sum score was calculated by adding up the weights of the criteria which were completely or partially addressed. The weights which were appointed to each of the studies can be found in Annex 3.

Table 6 shows the percentage score of all items. The columns no fulfilment, partial fulfilment and complete fulfilment of the criteria show the percentages of the studies which were able to fulfil that particular item. The sum of the item score represents the sum of (1) and (0,5) on that particular item. Since there are 34 studies, the maximum potential score per item of a particular criterion is 34.
The fulfilment score in percentages represents the percentage in which each criterion was fulfilled.

Table 6: Quality score per items (criteria)

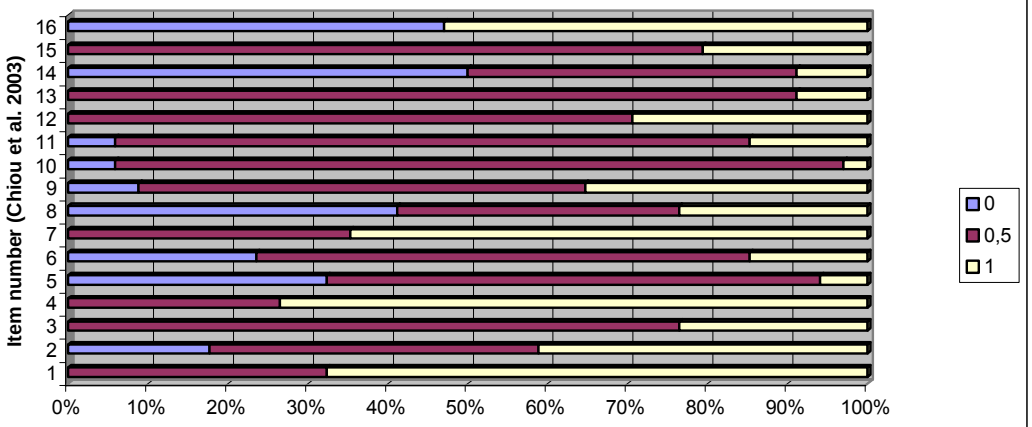
Item number	Criteria of Chiou et al. 2003 (see Annex 1)	No fulfilment of criteria (0) in %	Partial fulfilment of criteria (0,5) in %	Complete fulfilment of criteria (1) in %	Sum of fulfilment score per item	Fulfilment of criteria in %
1	Clear presentation of study	0	32.4	67.6	28.5	84
2	Stating objective	17.6	41.2	41.2	20.5	60
3	Best available source study	0	76.5	23.5	20	59
4	Prespecification of groups	0	26.5	73.5	28.5	84
5	Performing statistical and sensitivity analyses	32.3	61.8	5.9	12	35
6	Incremental analysis for resources and costs	23.5	61.8	14.7	15.5	46
7	Stating methodology for data abstraction	0	35.3	64.7	27	79
8	Appropriate analytic horizon, discounting and discount rate	41.2	35.3	23.5	13.5	40
9	Appropriate measurement of costs, description of estimation	8.8	55.9	35.3	20.5	60
10	Stating primary outcome including short, long, negative outcomes	5.9	91.2	2.9	16	47
11	Valid and reliable health outcomes, justification for measurement and scale	5.9	79.4	14.7	18	53
12	Stating the economic model	0	70.6	29.4	21.5	63
13	Stated and justified choice of the economic model	0	91.2	8.8	17.5	51
14	Discuss potential and magnitude of biases	50	41.2	8.8	9.5	28
15	Justified conclusion based on study results	0	79.4	20.6	20	59
16	Resource of funding	47.1	0	52.9	18	53

Table 6 shows that none of the criteria were fulfilled entirely. On item number 1, the criterion which measures whether an article and the aim of the study is presented in a clear manner, most publications scored well and the percentage score of that item is also the highest (84%). But in most publications valuable information to draw a correct conclusion was missing. This shows in the percentage of item number 14 (28%), which also has the lowest fulfilment score.

There are seven items which publications were able to fulfil partially or even entirely. The percentages of the partial fulfilment score are often also quite large, which indicates that publications were able to fulfil some part of the criteria, but that did not present sufficient information to appoint the entire weight.

Very few publications were able to fulfil item 5, 10, 13 and 14 entirely. The percentages of these items on entire fulfilment of these items were smaller than 10%. But most items have high percentages at fulfilling part of the criteria. Item number 16 is the only item which scored either (0) or (1), because that criterion of Chiou et al. does not give the possibility to appoint a partial weight. The information on the fulfilment of the criteria can also be found in figure 1.

Figure 1: Item score



4.4 Characteristics of the economic evaluations

Type of studies

The effectiveness data of the studies was either based on a single study, so the actual research which is described in the articles, or on multiple studies. When the authors used multiple studies, they often used effectiveness data of other studies or combined data of other studies with their own effectiveness data to gain results.

Of the 34 articles included in this study, there were 21 articles which were based on a single study and 13 articles which were based on multiple studies. The average sum score of the articles based on a single study was lower (45.2) than the average sum score of articles based on multiple studies (55.8).

Countries

The various economic evaluations were also differentiated by countries in which the studies were conducted. As can be seen in table 7.1 most of the studies were conducted in the USA, followed by the UK and Australia. Ireland had the highest sum score (70.95) and Finland had the lowest sum score (36.36). However, both results were based on the quality of one study.

Table 7.1: Description countries

Countries	# studies	Mean
Australia	3	49.73
Belgium	1	52.02
Denmark	2	40.44
Finland	1	36.36
Germany	1	55.74
Greece	1	31.14
Ireland	1	70.95
The Netherlands	1	56.01
Spain	1	52.81
Sweden	1	26.40
UK	3	52.88
USA	18	54.55

Since a small number of economic evaluations in most countries reflected the quality of publications, the grouping of countries was used to give a better picture of the quality across countries.
The geographical region was used to group countries and differentiation was made between European countries and non-European countries, in this case the USA and Australia.
Table 7.2 shows that the mean scores for the two regions are not very different.

Table 7.2: Description countries

Location	# studies	Mean	Standard Deviation
European countries	13	50.61	13.49
Non-European countries	21	52.61	13.43

Categories of interventions

The 34 economic evaluations were divided in several categories to distinguish between various interventions. These categories were *surgery*, *medicines*, *diets*, *behavioural treatment* and *mixed intervention programmes*. The mixed intervention programmes were a combination of all or several interventions. In addition to these categories a separate category could be distinguished, namely *type 2 diabetes*. The studies of this category focused on patients who suffered from type 2 diabetes and were also overweight or suffered from obesity (see table 8).

The category which had the highest mean sum score is the category *medicines* (65.65). The lowest mean sum score was found in the category *diets* (40.44), but it should also be noted that this score was based on two publications. So the finding of the lowest score in this category could be a chance finding. If one looks at the average sum score of all studies, it becomes clear that only the category of medicines and mixed intervention programmes have a higher score than the mean sum score of all studies.

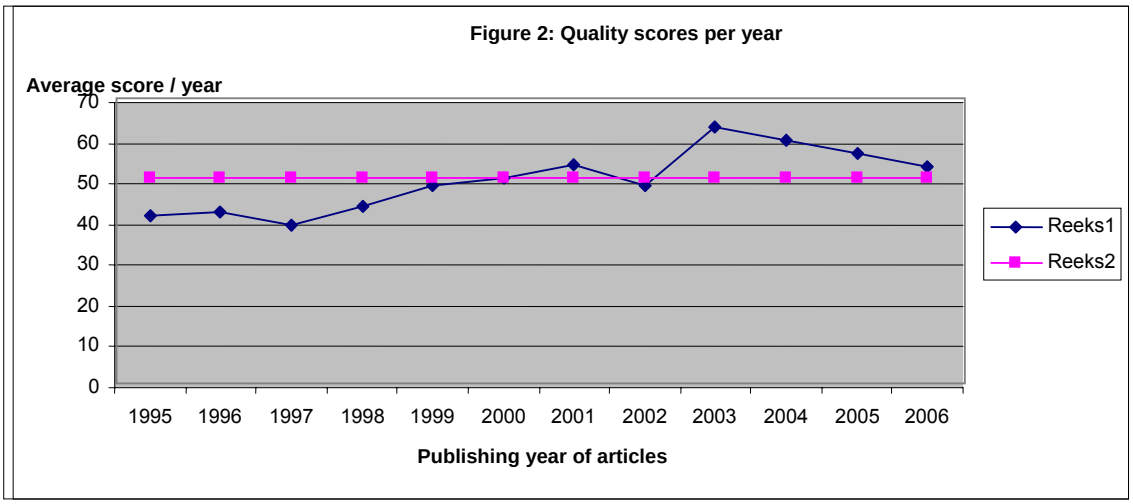
Table 8: Description categories

Category	# studies	Mean	Standard Deviation
Surgery	13	47.34	11.19
Medicines	4	65.65	10.59
Diets	2	40.45	23.12
Behaviour	4	45.99	7.77
Mixed intervention programmes	4	56.00	18.82
Type 2 Diabetes	7	51.91	9.44

Year of publication

To take a closer look at the quality of the economic evaluations, the sum score of the studies was also categorised in publishing year. The studies which were included were evaluations from 1995 until 2006. In figure 2 can be seen that the articles which were published from 1995 until 2000 have a lower average score than the articles which were published from 2001 until 2006.

Figure 2: Quality scores per year



Funding

The publications also include whether the authors stated the source which funded the research. Overall, the funding resources which were mentioned were health research organisations, installed by governments or universities, and pharmaceutical companies. To examine whether the quality of the articles was better depending of the funding resource, the funding resources were divided in three categories, namely 'research organisations', 'pharmaceutical companies' and 'none mentioned'.

From the results of the quality score, it was clear that the mean score of the articles which were funded by pharmaceutical companies was higher (62.7) than the articles which were funded by health research organisations (mean sum score = 48.7). The articles which did not mention the source of funding (mean sum score = 49.5) were close to the mean sum score of all articles (51.91).

5. Analysis

This chapter includes the results of several statistical analyses. The general aim of these analyses was to determine if the quality of the economic evaluation of obesity treatments was associated with a number of study characteristics. These analyses will give more insight in the patterns which can be found in the economic evaluations on obesity and what kind of information can be gained from using Chiou et al. 2003.

5.1 Relationship between study characteristics and total quality score

The characteristics or specific details of the studies, which were presented in the previous chapter, can also be seen as variables, including the sum score and the criteria used by Chiou et al. (2003).
In some of the cases there appears to be a relationship between the variables and the sum score and item score. The variables were Year, Category, Type, Country, Europe, Funding, in addition to the various items and sum score.

First the relationship between the characteristics of the studies and the sum score of the studies was tested. The aim was to determine whether there was an association between the quality of the studies and the characteristics.

To test whether there is a linear relationship between the variables and the sum score, regression analyses on Year, Type and Europe were performed.
The results can be seen in table 9.

Table 9: Results regression analyses (SPSS)

Variable	P-value	B-coefficient	R Square
Year	0.003	1.80	0.25
Type	0.04	9.47	0.13
Europe	0.68	-2.00	0.005

The year of publication was significantly associated with the quality of the articles (at α=0.05). This meant that the more recent publications received higher scores for quality than the older publications. In addition, the overall quality score of the study was higher if the study was based on multiple studies rather than a single study. The location of the study (Europe vs. non-Europe) showed a high p-value, indicating that the location of a publication was not significantly associated with the quality of a publication.

The previous analyses have given some insight in the significance of the various variables. But, because some variables were grouped in a random way, dummy variables were created to test for the significance of the variables.

The r-square of the variable Category was 0.287. In table 10 can be seen that the category medicines scored approximately 8.8 point higher in quality, while the other categories have a lower average sum score. The category which scores the least is the category with the articles on interventions which are aimed at changing dietary patterns.

Table 10: Parameter Estimates of bivariate analysis of variance (SPSS); variable CATEGORY

Dependent Variable: Sum score (Quality)

Parameter	B	P-value	95% Confidence Interval	
			Lower Bound	Upper Bound
Intercept	56.86	.00	47.43	66.28
Surgery	-9.52	.11	-21.21	2.18
Medicines	8.79	.26	-6.84	24.43
Diets	-16.41	.10	-36.41	3.59
Behaviour	-10.87	.17	-26.51	4.76
Mixed Intervention Programmes	-.85	.91	-16.49	14.78
Reference category: *Type 2 Diabetes*	0	.	.	.

The variable Country shows that there is a difference between the quality per article and country. The r-square is 0.249. For example, looking at table 11, the quality of the article from Greece is approximately 21.9 point lower than articles from USA. But it should be noted that the quality is often based on the score of one abstract.

Table 11: Parameter Estimates of bivariate analysis of variance (SPSS); variable COUNTRY

Dependent Variable: Sum score (quality)

Parameter	B	P-value	95% Confidence Interval	
			Lower Bound	Upper Bound
Intercept	54.55	.00	48.28	60,82
Australia	-4.82	.55	-21.40	11.76
Belgium	-2.53	.85	-29.85	24.79
Denmark	-14.11	.15	-33.93	5.72
Finland	-18.09	.18	-45.41	9,23
Germany	1.19	.90	-18.63	21.01
Greece	-23.41	.09	-50.73	3.91
Ireland	16.40	.23	-10.92	43.72
The Netherlands	1.46	.91	-25.86	28.78
Spain	-1.74	.90	-29.06	25.58
Sweden	-28.15	.04	-55.47	-.83
UK	3.25	.74	-16.57	23.07
Reference category: *USA*	0	.	.	.

Table 12: Parameter Estimates of bivariate analysis of variance (SPSS); variable FUNDING
Dependent Variable: Sum score (Quality)

Parameter	B	P-value	95% Confidence Interval	
			Lower Bound	Upper Bound
Intercept	49.86	.00	43.40	56.31
Research organisations	-1.19	.81	-11.30	8.92
Pharmaceutical companies	11.84	.05	.14	23.54
Reference category: *Non mentioned*	0	.	.	.

The r-square of the variable Funding was 0.147. Table 12 shows that there was a large difference between the quality score of the publications when the publications were split according to the source of funding. Studies which were financed by pharmaceutical companies scored an average of 11.8 points higher than the studies which did not mention any source of funding and approximately 13 points higher than the studies funded by health research organisations. However, better quality of the studies based on funding by pharmaceutical companies is based on six publications, which could indicate a chance finding. Especially since seventeen publications did not mention a source of funding.

Table 13 shows the multivariate analysis of all variables in one model. When the previous results are compared to the results of this table, there are some significant differences.

Table 13: Parameter Estimates of multivariate analyses of variance (SPSS)
Dependent Variable: Sum score (Quality)

Parameter	B	P-value	95% Confidence Interval	
			Lower Bound	Upper Bound
Intercept	-2958.09	.02	-5400.91	-515.27
Type of intervention				
Surgery	-7.41	.42	-25.89	11.06
Medicines	11.20	.13	-3.37	25.78
Diets	-1.60	.90	-28.13	24.93
Behaviour	-5.29	.60	-25.75	15.17
Mixed Intervention programmes	4.93	.61	-14.48	24.33
Type 2 Diabetes	0	.	.	..
Location				
European countries	14.46	.01	4.48	24.44
Non - European countries	0	.	.	.
Source of funding				
Research organisations	-3.72	.52	-15.49	8.05
Pharmaceutical companies	1.08	.90	-16.24	18.41
Non mentioned	0	.	.	.
Type of study used				
Based on a single study	-6.22	.24	-16.90	4.45
Based on multiple studies	0	.	.	.
Year of publication	1.50	.02	.28	2.72

The B-coefficient shows the direction of the association of the different variables. Previously the regression analysis showed that the variable Year was significant. In the table 13 can be seen that the association is a positive one, meaning that as time progresses with one year, the sum score of the publications increases with 1.5 points. The results of the regression analyses showed that the B-coefficient was slightly higher, namely 1.8 points.

The B-coefficients of the other variables also seems to become smaller when the multivariate analysis is compared to the bivariate analyses. This can be seen when you can compare both the variables Category and Funding. The table also shows that studies which are based on a single study score respectively 6.2 points lower than those studies which are based on multiple studies.
The difference between the sum score of studies supported by pharmaceutical companies and other organisations can also be seen.
Comparing the scores of the different intervention categories shows that the publications based on mixed interventions and medicines score higher than the other categories.

However, these differences of the B-coefficient do not correspond with a low p-value in the entire model. The variable TYPE, which was earlier found to be associated with the quality score of the studies, is not associated with the quality of the studies in the model with all variables.
Instead, the variable of Non-European countries does seem to be associated with the quality and shows a positive B-coefficient, which means that articles of Non-European countries have a higher sum score of almost 14.5 points.

Because both variables related to the location of Non-European countries and the publication year are associated, it could be that these variables interact. This could lead to a higher B-coefficient than can actually be found in practice.

Table 14: Parameter Estimates of univariate analyses (SPSS) variables EUROPE * YEAR

Dependent Variable: Sum score (Quality)

Parameter	B	P-value	95% Confidence Interval	
			Lower Bound	Upper Bound
Intercept	-4064.99	.04	-7848.15	-281.84
European countries vs non-European	196.98	.93	-4621	5015
Interaction between non -European countries and year of publication	2.16	.006	.67	3.65
Interaction between European countries and year of publication	2.06	.03	.17	3.95

The analysis in table 14 shows that the effect of year of publication does not vary substantially from the studies performed in Europe and the studies performed elsewhere. In both cases, the B-coefficient shows approximately a 2-point improvement for every increase in the year of publication.

5.2 Relationship between variables and item score

The previous paragraph showed analyses which were based on the absolute sum score of the economic evaluations. This is the most clear reference to the quality of a publication, because a better sum score also refers to a better quality. But one could also argue that not all criteria have to be included in the analysis in order to determine the quality of the publication.
A researcher could solely include the criteria of interest. If this choice were to be made, it is important to test whether the same association exists for the individual item score as exists for the sum score. The correlation between the individual items and the various characteristics was tested to see if the same association exists.
The results are shown in table 15. In this table only the items and the characteristics which show an association were included.

Table 15:

Correlation between items and parameters

Item	Criteria Chiou et al. 2003 (see Annex 1)	Parameter	P-value
3	Best available source of study	Type	0.042
4	Prespecification of groups	Country	0.048
5	Performing statistical and sensitivity analyses	Year	0.001
8	Analytic horizon, discounting and discount rate	Type	0.005
10	Stating primary outcome including short, long, negative outcomes	Year	0.02
14	Discuss potential and magnitude of biases	Type	0.044
		Year	0.02
16	Resource of funding	Category	0.00
		Funding	0.00

To test the correlation between Year and the several items, a Spearman's test was conducted. Given an $\alpha = 0.05$, a correlation seems to exist between the year of publication and item 5, item 10 and item 14. Item 5 relates to whether any statistical analyses are performed. Item 10 relates to whether the authors clearly discuss the primary outcomes and the major short, long and negative outcomes.
Item 14 is related to the direction and magnitude in which potential biases are discussed.
The correlation between a characteristic feature and these items is not directly clear. But it is probably the case that as the publishing year becomes more recent, the authors become more accurate in stating their methods and all relevant information such as outcomes and potential biases.

The Kruskal-Wallis – test was used to test the relationship between the other variables and the items. Item 16 was associated with source of funding and category of intervention. For the source of funding this was very logical since item 16 relates to mentioning what the source of funding is. The correlation between the category of intervention and item 16 is probably due to the category of surgery. The studies in this category rarely referred to funding, unlike the studies of other categories.

The type of study (single study vs. multiple studies) correlated with item 3, item 8 and item 14. Type of study was correlated with item 3 because item 3 refers to the type of analysis which is performed. The B-coefficient of publications which were based on single studies were lower than publications which were based on multiple studies. However, this is not what was to be expected from this item, because a randomised controlled trial is considered to be the best available source for results. The publications which were based on multiple studies were mostly not randomised controlled trials, but because the type of study does have an effect on the quality of the studies, there is still a correlation between item 3 and the type of study.
The correlation with item 8 is more evident, because this criterion relates to the use of discounting. Studies which were partially based on previous publications used discounting to synthesize previous results, often in relationship with multivariate analyses. The category of surgery scored well on this criterion since this category contained many publications that were based on single studies.
The correlation therefore shows the contrast between the category surgery and the others.
The correlation with item 14 could result from the fact that publications based on multiple studies discuss potential biases more extensively. Again, the category surgery seldom fulfilled this criterion.

The country in which a study is performed only correlates with item 4 ($p=0.048$), which refers to prespecifying the groups. The studies with a higher score were often more specific in mentioning the details of their methods. This could explain the correlation with this item. Surprisingly, an association between the geographical location in which the study was performed cannot be found in any of the items. Since this characteristic is derived from the variable Country, one would also expect to find an association with item 4. Especially since the countries which scored lower on this criterion were mostly European countries.

In summary, certain characteristics of economic evaluations, such as year of publication and whether the study was based on a single study or multiple studies, were significantly associated with the quality of the articles. However, knowledge about these characteristics is not enough to determine the quality of a study based on Chiou et al.'s tool, since the overall quality score could still be lower than what one would predict based on year of publication or use of a single study.

The correlation between the characteristics and certain items is also a part of this study which needs more research. Some items correlate with the characteristics of the publications, but these results could differ when a similar research with another subject than obesity is performed. The patterns in the characteristics of other economic evaluations may differ which would lead to a correlation with other items.

Conclusion

Obesity is currently a fast growing global problem, where more and more individuals have an extreme overweight or a BMI above 30. Because obesity is a risk factor for multiple diseases, such as diabetes type 2 and cardiovascular diseases, the consequences are large for both public health and health care costs. To decrease the impact of obesity, there are multiple strategies which can be implemented, namely exercise, behavioural modification, diet therapy, pharmacotherapy and surgery. Often, these strategies are combined to gain more weight loss.

The 34 economic evaluations studied in this thesis were gathered from the NHS EED and graded to determine the quality of these economic evaluations.
The methods which exist to qualify economic evaluations are mostly qualitative and describe which elements are included in a certain publication. In this thesis a quality assessment tool, developed by Chiou et al. (2003), was used to assess the quality of articles by using sixteen quantitative weights.
If the study was able to fulfil the criterion, the entire weight was appointed. Half of the weight was appointed when the criterion was partially addressed and none of the weight was appointed when the criterion was not fulfilled.
Economic evaluations were performed on surgical interventions, pharmacotherapy, interventions aimed at changing dietary patterns or behavioural modifications. But there were also evaluations on mixed interventions programmes which combined these strategies and interventions especially aimed at diabetes type 2 patients suffering from obesity.

The 34 economic evaluations on interventions aimed at decreasing obesity were published from 1995 until 2006. The mean quality score of these studies was 51.9 with a range of 24.1 - 75.8 and a standard deviation of 13.28.
There were a number of criteria which were not fulfilled entirely, but on four criteria less than 10% of the criteria were fulfilled. These criteria were related to including valuable information to draw a conclusion, discussing the choice of an economic model, the primary outcomes (including the short, long and negative outcomes) and including relevant statistical and sensitivity analyses. More than half of the studies also failed to mention the source of funding.

Using this tool showed that the publishing year of the article and the type of study can be associated with the quality of studies on obesity. The more recent publications were published, the better the quality was in comparison to older publications. The multivariate analysis showed that the quality of a study could increase with 1.5 points if one year had progressed.
If a publication was based on the results of multiple studies rather than a single study, the likelier it was that the quality of this publication would be better.

The quality of articles does not seem to differ between countries, but more on the level of Europe versus other continents. European studies had a higher quality score of 14.46 points on average.
The funding of articles did not seem to be associated with the quality of the studies, but studies which were funded by pharmaceutical companies did have a better score than studies which were funded by health research organisations.
However, the significance of the country of publication or funding only becomes apparent when year of publication or type of study is taken into account.
Grouping articles in categories of intervention does not show an association with the quality, which means that the quality does not differ between the categories.

The associations which were mentioned before are based on the sum score of the articles. The analogy is that the higher the sum score is, the better the quality of the publication will be. The individual item score which is gained by using the assessment tool could also give some indications of the quality of an economic evaluation. But there does not seem to be a relationship between all the items and the characteristics of the publication. There are only some items which correlate with a particular characteristic of a study. Therefore it is not clear if it is possible to select only the items of interest to reflect the quality of a publication. It could also be possible that the correlation between certain items and characteristics will differ when another disease area is examined.

Thus, the quality of economic evaluations can be assessed with the quality assessment tool that Chiou et al. (2003) present. The economic evaluations on interventions aimed at decreasing obesity are not of high quality if the mean quality score is taken into account, but this is only an average.
The associations which were made and the conclusions on the manner in which all criteria were fulfilled provide insight in the possibilities to increase the quality of the economic evaluations on obesity interventions.

Discussion

In this study there are some issues which need to be addressed and for which certain recommendations can be made.

Shortcomings of the quality assessment tool

Some of the issues which Chiou et al. 2003 discussed also appeared to be an issue in this thesis.

The first issue which was raised in their study was that it might be better to use a 3- or 5- point scale rather than a 'yes/no' response scale. Chiou et al. 2003 pointed out that the complexity of a 3- or 5- point scale could reduce the potential value of a simple grading system.

However, they did recommend more research on this topic in order to determine the value of a more complex grading system. I agree that this is an issue which needs more research.

The problem with the criteria of Chiou et al. is that most criteria are difficult to answer using a 'yes/no' response. Therefore, it is also difficult to decide whether a study should be appointed the entire weight of the criterion or none of the weight.

Some of the criteria also addressed multiple subjects which should be reflected in the study. This made it more difficult to appoint the weight of a particular criterion. Chiou et al. did not specify if a subject of a particular criterion was more important than another and what effect a subject would have on the fulfilment of that criterion. As a solution to this issue, I decided to add the option of appointing half of the criterion weight whenever part of the criterion had been fulfilled. The addition of this option made it easier to appoint the weights of the different criteria and looking at the sum scores of the various economic evaluations, it also made it easier to differentiate between the quality of the economic evaluations. If the addition of this option was not added, certain economic evaluations would have a much lower or higher score which would indicate less or better quality of an economic evaluation than in fact was the case. If part of the criterion was fulfilled it made it easier to choose what the appointment of the weight should be and if this appointment of the weight was in fact a reflection of the quality of the economic evaluation on that particular criterion.

The second and third issue that Chiou raised was the use of academic health economists to validate the criteria system and expert characteristics which were therefore added to this tool. The use of health economist does create a gold standard, but I agree with Chiou et al. that it remains unclear whether expert opinions are reflected in the criteria and whether that causes a conflict for other researchers. It could be possible that the weight of certain criteria is too high or too low for a particular researcher.

For example, criterion 1 has a weight of 7, but in this thesis it became clear that most studies score well on this criterion. A researcher could decide to decrease the weight of this criterion and increase the weight of another criterion. If a 3- or 5- point scale would be used to decide how much of the weight would be appointed, creating a higher weight for another criterion would lead to a better possibility to differentiate between the score of studies on 'more important' criteria. This would than lead to better differentiation between the quality of the economic evaluations.

Possible shortcomings of the study

In this study, the information about economic evaluations was primarily taken from the NHS EED rather than the actual publications. It is therefore possible that this approach biased the assessment of the quality of the economic evaluations. However, it could be argued that the use of a database gives a better indication of the actual score when Chiou et al. is used.
The NHS EED abstracts are written by health economists and have a clear structure and method in which all relevant details of the studies are summarised. The details of the studies can be missed if one were to determine the quality of every study based only on the original publications.

Since all of the economic evaluations were scored by one person, it is possible that this led to a lower validity and reproducibility than a method involving two or more persons.

The introduction of giving half-points for partial fulfilment of criteria could be an issue. However, since it is unclear what the result can be of using different scales, further research should be done in order to determine what type of scale is useful.

Several characteristics of the studies were selected to determine their association with the quality of economic evaluations. However, other study characteristics could have been selected, for example the impact factor of the journal in which the article was published. However, the primary aim of this study was to gain some insight into the usefulness of this quality assessment tool rather than to assess the association between the specific study characteristics and the quality of economic evaluations.
The differences in the results between the univariate, bivariate and multivariate analyses were striking. The reason for this fact could be that when the characteristics are tested in a model, the association of a characteristic with the quality becomes significant, while it is not significant in the univariate or bivariate analyses. In the univariate analysis the use of other publications in the economic evaluations was associated with the quality, while this association did not appear in the p-value of this characteristic in the multivariate analysis. The geographical region was not associated with the quality of the economic evaluations, while the model did show an association with the quality according to the p-value.

The only characteristic, namely year of publication, remained significant in both the univariate and multivariate analyses. However, the publication year alone cannot indicate what the quality of an economic evaluation is. More recent publications can still have a lower quality score than older publications, even though there is an association between the year of publication and the quality of economic evaluations. It is also likely that the increase of quality score according to publication year reflects a natural progression in the quality of economic evaluations in general and not simply an improvement only found in cost-effectiveness studies of obesity interventions.

The quality of publications might improve, because more knowledge becomes available through the years. This may also be the reason why in the univariate analyses, publications which are based on multiple studies appeared to have a higher quality than those based on a single study. Another issue which can be addressed regarding the benefit of the results is the sum score or quality score.
When several studies are evaluated and graded the result can be a total sum score, but it is not immediately clear how this score can be used. The sum score, in my opinion cannot solely be used as a threshold for the quality, because it is difficult to draw the line on a particular score. For example, when a study has a sum score of 65, it is easy to say that this study scored well and therefore the intervention generates good results. But because the sum score is based on so many criteria, it is possible that a study which appears to have a good score, has not taken issues into account which are relevant, such as the inclusion of all biases or negative outcomes.

Utility

This assessment tool can be very useful when one wants to evaluate the quality of different cost-effectiveness studies which have been published, for example on certain disease areas. Quality assessment of existing studies would be helpful when a similar research will be performed, because it would give a good indication about areas where the quality could be improved in future studies. If the quality of economic evaluations is improved, this could lead to better chances of implementation of those interventions that are truly effective and cost-effective.

The comparison and quality of other studies which have done similar methods versus the methods that I used in this thesis are also issues which need to be discussed.
First, the choice to use the Chiou scoring method seemed to be a clear one because, unlike other methods, the Chiou method provides the possibility to quantify the various aspects of quality and determine the overall quality. Moreover, the criteria included in the Chiou's tool correspond with the list of criteria seen in other quality assessment criteria. The possibility of quantifying the methods gives an extra dimension to critiquing existing studies and makes it less abstract than methods such as Drummond et al. (2005).

Drummond et al. (2005) present a checklist which could be used for the critical appraisal of a published article. It consists of ten questions which can be answered with 'yes', 'no' or 'can't tell'.
The aim of this checklist is to assist the reader in the interpretation of the economic evaluations and to determine how useful the economic evaluations are for health care decisions.
Although Drummond et al. (2005) present criteria which are similar to those of Chiou et al, it does not give the possibility to present the quality of articles in a quantitative manner.
However it does give the possibility to create a detailed review, similar to the NHS EED abstracts. But detailed reviews cannot be easily mapped out when various studies are discussed and need to be compared.

The research done by Pignone et al. (2005) points out that the use of systematic reviews could create a greater dialogue between model developers and policy makers and could also provide decision makers with important information to make policy decisions. Systematic reviews could also hand better mechanisms for the presentation of modelling results.

The two studies which were discussed in paragraph 2.2 were somewhat similar to the research of this thesis.
Ofman et al. (2003) presented the quality assessment tool as the Quality of Health Economic Studies (QHES) instrument, but they did not take into account the use of a constant when grading the cost-effectiveness studies on gastroesophageal reflux disease (GERD). However, the similar advantages and limitations were found.
Some limitations of the tool were (1) the requirement of published evaluations,
(2) yes/no response scale rather than a continuous scale and (3) the experience a non – user would need to determine whether an article could fulfil the various dimensions. The last limitation was not a problem in this thesis, but it is understandable that appointing weight to several articles could be difficult when one does not have any experience with judging the quality of studies.

Spiegel et al. (2004) performed a systematic review to identify and assess the quality of health economic analyses on digestive diseases by using the QHES instrument. Spiegel et al. also found that the QHES does not give the possibility to measure the external validity or generalisability of health economic analyses. This limitation was also found in this thesis when assessing the quality of economic evaluations on obesity with the use of Chiou et al.
Thus it would be a recommendation that an addition would be made to the current tool which would assess the external validity of the tool. This could be done by adding a criterion to the tool and also validating what weight this criterion should have.

Recommendations

Further research should be done to decide whether the use of partial scores of (0.5), to indicate the partial fulfilment of a criterion, is sufficient to differentiate between the quality of an economic evaluation. Perhaps it is also necessary to study whether a larger scale will be more capable of differentiating between the quality of articles.
The NHS EED was also used in this thesis since it facilitated the search for economic evaluations of interventions for obesity. However, this database also includes a brief assessment of the quality of these evaluations and therefore makes the appointment of the weights of Chiou et al. less subjective. Therefore, use of this database is recommended in any assessment of the quality of economic evaluations.

The results of this thesis showed that it is not entirely possible to use the item score instead of the sum score, because the interaction between the various characteristics of the economic evaluations and the item score differs. However, it would be useful if future research would give more insight in whether the use of a limited number of items would be sufficient to assess the quality, rather than calculating the sum score.

The focus of this thesis was on economic evaluations discussing interventions aimed at decreasing obesity. However, because this is a specific subject, it is possible that the assessment tool should be adjusted to review the economic evaluations better. Perhaps it should be discussed which criteria are important for the evaluation of obesity interventions and if adjustments of these criteria or even the height of the weights could be made.
In the past, a criteria list for a specific problem has been applied. In the UK several researchers felt that there was a lack of high quality studies on self care (Gravelle et al. 2003). That is why they created a specific quality assessment tool which is similar to the criteria of Chiou and Drummond, but also added criteria which are relevant for the subject of self care. Perhaps the same can be done for the obesity. That is, additional criteria could be added to the criteria of Chiou et al. (2003) which would provide a tool to evaluate the quality of economic evaluations on obesity. These additional obesity-specific criteria could also affect the appointment of the weight which could lead to more accurate reflection of the quality of the article found in the sum score. However, even though it could be valuable to create a disease-specific quality assessment tool, much effort will have to be put in before this tool is made. The question would also be to what extent a disease-specific quality assessment tool is actually preferred. Therefore it would be an addition to the current knowledge if first more information would be presented on the use of the assessment tool in other disease areas.

The previously mentioned recommendations make it possible to use of the quality assessment both before and after publications of economic evaluations.
The quality assessment tool can be used before a research is done to plan the manner in which a study will be performed. Before the article is submitted, researchers can check again whether they have included all relevant information. Afterwards the publications can be judged on their quality and this information could offer more accurate and new insights which could lead to important policy implications.
For the interventions which are aimed at decreasing obesity, the quality assessment tool gives better insight in the general quality of these economic evaluations, but also in the economic evaluations which are performed in the different categories. Insight in the quality of these publications could also hand more tools to change the policies which are aimed at decreasing obesity, because the results of especially the high quality studies can be better implemented in current interventions.

References

- American Obesity Association. 2006. 'Costs of obesity.' [Internet] [quoted 11-30-2006] attainable at www.obesity.org.
- Bhattacharya, J. Bundorf, M. 2005. 'The incidence of the healthcare costs of obesity.' *National Bureau of Economic Research*: Working Paper no. 11303.
- Centre for Reviews and Dissemination (CRD). 2006. '*About the…*' [Internet], 2006 [quoted on 05-08-2006] attainable at http://www.york.ac.uk/inst/crd/crddatabases.htm.
- Chiou, C. Hay, J. Wallace, J. 2003. 'Development and Validation of a Grading System for the Quality of Cost-Effectiveness Studies.' *Medical Care* 41(1):32-44.
- Chou, S. Grossman, M. Saffer, H. 2004. 'An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System.' *Journal of Health Economics* 23:565-587.
- Cutler, D. Glaeser, E. 2006. 'Why do Europeans smoke more than Americans?' *National Bureau of Economic Research*: Working Paper no. 12124.
- Drummond, M. Sculpher, M. Torrance, G. 2005. *'Methods for the Economic Evaluation of Health Care Programmes'*. Third edition. New York: Oxford University Press Inc.[1987]
- Gravelle, H. Richardson, G. Weatherly, H. 2003. 'An assessment of the Quality of Economic Evaluations of Self-Care Support. A report for the department of Health.' attainable at
- International Diabetes Federation (IDF). 2006. '*Obesity and type 2 diabetes*.' [Internet] , 2005
  [quoted on 04-14-2006] attainable at www.eatlas.idf.org.
- International Obesity Task Force (IOTF). 2006. '*About Obesity*.' [Internet] [quoted on 05-20-2006] attainable at http://www.iotf.org/.
- Kan, K. Tsai, D. 2004. 'Obesity and risk knowledge.' *Journal of Health Economics* 23:907-934.
- Ofman, J. Sullivan, S. Neumann, P. 2003. 'Examining the Value and Quality of Health Economic Analyses: Implications of Utilizing the QHES.' *Journal of Managed Care Pharmacy* 9(1):53-61.
- Philipson, T. 2001. 'The world-wide growth in obesity: an economic research agenda.' *Health Economics* 10:1-7.
- Pignone, M. Saha, S. Hoerger, T. 2005. 'Challenes in systematic reviews of economic analyses.' *Annals of Internal Medicine* 142(12):1073-1079.
- Pushka, P. Nishida, C., Porter, D. 2003. '*Obesity and overweight, fact sheet.*' [Internet] WHO fact sheet [quoted on 06-02-2006] attainable at www.who.int.

- Rashad, I. 2006. 'Structural estimation of caloric intake, exercise, smoking, and obesity.' *National Bureau of Economic Research*: Working Paper no. 11957.
- Spiegel, B. Targownik, L. Kanwal, F. 2004. 'The quality of publishes health economic analyses in digesitive diseases: A systematic review and quantitative appraisal. *Gastroenterology* 127(2):403-411.
- Vaidya, V. 2006. *Health and Treatment Strategies in Obesity*. *Advances in Psychomatic Medicine vol. 27*. Basel:Karger [2006].
- Winson, A. 2004. 'Bringing political economy into the debate on the obesity epidemic.' *Agriculture and Human Values* 21:299-312.
- World Health Organization (WHO). 2006. '*5.2 Recommendations for preventing excess weight gain and obesity*.' [Internet] [quoted on 05-20-2006] attainable at www.who.int.
- --. 2006. 'Obesity & overweight.' [Internet] [quoted on 10-20-2006] attainable at www.who.int.
- --. 2000. 'Obesity: Preventing and managing the Global Epidemic - Report of a WHO Consultation on Obesity, 3-5 June 1997'. Geneva, WHO/NUT/NCD/98.1
- --. 1998. 'MC4R associated with dominantly inherited obesity.' *Nat Genet* 20:111–12.

Annex 1

Criteria selected for the Grading System (Chiou et al. 2003)

1	Was the study objective presented in a clear, specific, and measurable manner?
2	Were the perspective of the analysis (societal, third-party payer, etc.) and reasons for its selection stated?
3	Were variable estimates used in the analysis from the best available source (i.e. Randomized Control Trial—Best, Expert Opinion—Worst)?
4	If estimates came from a subgroup analysis, were the groups prespecified at the beginning of the study?
5	Was uncertainty handled by: 1) statistical analysis to address random events; 2) sensitivity analysis to cover a range of assumptions?
6	Was incremental analysis performed between alternatives for resources and costs?
7	Was the methodology for data abstraction (including value health states and other benefits) stated?
8	Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3–5%) and justification given for the discount rate?
9	Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described?
10	Were the primary outcome measure(s) for the economic evaluation clearly stated and were the major short term, long term and negative outcomes included?
11	Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for the measures/scales used?
12	Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner?
13	Were the choice of economic model, main assumptions and limitations of the study stated and justified?
14	Did the author(s) explicitly discuss direction and magnitude of potential biases?
15	Were the conclusions/recommendations of the study justified and based on the study results?
16	Was there a statement disclosing the source of funding for the study?

Estimated weights for criteria (Chiou et al. 2003)

Criteria	Coefficient	Weight
1	6.89	7
2	3.91	4
3	7.36	8
4	1.20	1
5	8.86	9
6	5.79	6
7	4.46	5
8	6.53	7
9	7.31	8
10	6.02	6
11	6.54	7
12	7.11	8
13	6.23	7
14	5.48	6
15	7.05	8
16	2.77	3
Constant	−0.95	100 (Total)

Annex 2

Summaries of NHS EED

- Brolin, R. 1996. 'Prospective, randomized evaluation of midline fascial closure in gastric bariatric operations'. *American Journal of Surgery* 172(4):328-331.
- Chua, T. Mendiola, R. 1995. 'Laparoscopic vertical banded gastroplasty: the Milwaukee experience'. *Obesity Surgery* 5(1):77-80.
- Cooney, R. Bryant, P. Haluck, R. 2001. 'The impact of a clinical pathway for gastric bypass surgery on resource utilization'. *Journal of Surgical Research* 98(2):97-101.
- Craig, B. Tseng, D. 2002. 'Cost-effectiveness of gastric bypass for severe obesity'. *American Journal of Medicine* 113(6):491-498.
- DeMaria, E. Schweitzer, M. Kellum, J. 2002. 'Hand-assisted laparoscopic gastric bypass does not improve outcome and increases costs when compared to open gastric bypass for the surgical treatment of obesity'. *Surgical Endoscopy and Other Interventional Techniques* 16(10):1452-1455.
- Foxcroft, D. Milne, R. 'Orlistat for the treatment of obesity: rapid review and cost-effectiveness model'. 2000. *Obesity Reviews* 1(2):121-126.
- Fries, J. McShane, D. 1998. 'Reducing need and demand for medical services in high-risk persons: a health education approach'. *Western Journal of Medicine* 169(4):201-207.
- Gemert van, W. Adang, E. Kop, M. 1999. 'A prospective cost-effectiveness analysis of vertical banded gastroplasty for the treatment of morbid obesity'. *Obesity Surgery* 9(5):484-491.
- Ghassemian, A. MacDonald, K. Cunningham, P. 1997. 'The workup for bariatric surgery does not require a routine upper gastrointestinal series'. *Obesity Surgery* 7(1):16-18.
- Goldfield, G. Epstein, L. Kilanowski, C. 2001. 'Cost-effectiveness of group and mixed family-based treatment for childhood obesity'. *International Journal of Obesity* 25(12):1843-1849.
- Gould, J. Garren, M. Starling, J. 2004. 'Laparoscopic gastric bypass results in decreased prescription medication costs within 6 months'. *Journal of Gastrointestinal Surgery* 8(8):983-987.
- Katopodi, T. Katostaras, T. 2005. The weight reduction operation of choice: Mason vs gastric bypass. A comparative study'. *Review of Clinical Pharmacology and Pharmacokinetics. International Edition* 19(3):155-158.
- Katz, D. Chan, W. Gonzalez, M. 2002. 'Technical skills for weight loss: preliminary data from a randomized trial'. *Preventive Medicine* 34:608-615.
- Kumpusalo, E. Neittaanmaki, L. Halonen, P. 1996. 'Finnish healthy village study: impact and outcomes of a low-cost local health promotion programme'. *Health Promotion International* 11(2):105-115.

- Lacey, L. Wolf, A. O'Shea, D. 2005. 'Cost-effectiveness of orlistat for the treatment of overweight and obese patients in Ireland'. *International Journal of Obesity* 29(8):975-982.
- Lamotte, M. Annemans, L. Lefever, A. 2002. 'A health economic model to assess the long-term effects and cost-effectiveness of orlistat in obese type 2 diabetic patients'. *Diabetes Care* 25(2): 303-308.
- Maetzel, A. Ruof, J. Covington, M. 2003. 'A Economic evaluation of orlistat in overweight and obese patients with type 2 diabetes mellitus'. *Pharmacoeconomics* 21(7):501-512.
- Malone, D. Raebel, M. Porter, J. 2005. 'Cost-effectiveness of sibutramine in the LOSE Weight Study: evaluating the role of pharmacologic weight-loss therapy within a weight management program'. *Journal of Managed Care Pharmacy* 11(6):458-468.
- Martin, L. Tan, T. Horn, J. 1995. 'Comparison of the costs associated with medical and surgical treatment of obesity'. *Surgery* 118(4):599-607.
- Meyers, A. Graves, T. Whelan, J. 1996. 'An evaluation of a television-delivered behavioural weight loss program: are the ratings acceptable?' *Journal of Consulting and Clinical Psychology* 64(1):172-178.
- Olsen, J. Willaing, I. Ladelund, S. 2005. 'Cost-effectiveness of nutritional counseling for obese patients and patients at risk of ischemic heart disease'. *International Journal of Technology Assessment in Health Care* 21(2):194-202.
- Paxton, J. Matthews, J. 2005. 'The cost effectiveness of laparoscopic versus open gastric bypass surgery'. *Obesity Surgery* 15(1):24-34.
- Pritchard, D. Hyndman, J. Taba, F. 1999. 'Nutritional counselling in general practice: a cost effectiveness analysis'. *Journal of Epidemiology and Community Health* 53:311-316.
- Rowe, R. Cowx, M. Poole, C. 2005. 'The effects of orlistat in patients with diabetes: improvement in glycaemic control and weight loss'. *Current Medical Research and Opinion* 21(11):1885-1890.
- Segal, L. Dalton, A. Richardson, J. 1996. 'The cost-effectiveness of primary prevention for non-insulin dependent diabetes mellitus'. Centre for Health Program Evaluation (Research Report 8) :1-73.
- Segal, L. Dalton, A. Richardson, J. 1998. 'Cost-effectiveness of the primary prevention of non-insulin dependent diabetes mellitus'. *Health Promotion International* 13(3):197-209.
- Shearer, A. Bagust, A. Liebl, A. 2006. 'A Cost-effectiveness of rosiglitazone oral combination for the treatment of type 2 diabetes in Germany'. *Pharmacoeconomics* 24(1):35-48.
- Shearer, A. Bagust, A. Ampudia-Blasco, F. 2006. 'Paris G Lifetime health consequences and cost-effectiveness of rosiglitazone in combination with metformin for the treatment of type 2 diabetes in Spain'. *Pharmacoeconomics* 24(1):49-59.

- Shope, T. Cooney, R. McLeod, J. 2003. 'Early results after laparoscopic gastric bypass: EEA vs GIA stapled gastrojejunal anastomosis'. *Obesity Surgery* 13(3):355-359.
- Siggaard, R. Raben, A. Astrup, A. 1996. 'Weight loss during 12 week's ad libitum carbohydrate-rich diet in overweight and normal-weight subjects at a Danish work site'. *Obesity Research* 4(4):347-356.
- Sjostrom, L. Narbro, K. Sjostrom, D. 1995. 'Costs and benefits when treating obesity'. *International Journal of Obesity* 19(6):9-12.
- Wang, L. Yang, Q. Lowry, R. 2003. 'Economic analysis of a school-based obesity prevention program'. *Obesity Research* 11(11):1313-1324.
- Warren, E. Brennan, A. Akehurst, R. 2004. 'Cost-effectiveness of sibutramine in the treatment of obesity'. *Medical Decision Making* 24:9-19.
- Wylie-Rosett, J. Swencionis, C. Ginsberg, M. 2001. 'Computerized weight loss intervention optimizes staff time: the clinical and cost results of a controlled clinical trial conducted in a managed care setting'. Journal of the American Dietetic Association 101(10):1155-1166.

Author(-s) Surgery	Chua & Mendiola (1995)	Martin et al. (1995)	Sjostrom et al. (1995)	Brolin (1996)	Ghassemian et al. (1997)
Aim	Determine the cost-effectiveness of laparoscopic vertical banded gastroplasty (VBG).	Assess cost-effectiveness of surgical treatment (gastric bypass) vs. medical treatment (VLCD + behavioural therapy)	Assess cost-effectiveness of performing surgery vs. conventional treatment	Comparing the two methods of closure of the linea abla after gastric restrictive operations. Double stranded #1 PDS suture vs. #1 Ethibond suture	Determining the diagnostic and patient-management of the gastrointestinal (GI) series in the assessment of patients undergoing a gastric bypass operation.
Perspective of the analysis	-	-	-	-	-
Country	USA	USA	Sweden	USA	USA
Data	Single study	Single study	Multiple studies. Evidence from an ongoing single study, effectiveness and resource data from 1993-1995.	Single study	Single study
Method	Two case series	Cohort study of patients choosing from both treatments. Two-stage overweight progression model > number of overweight adults	Not specified. Focus on a decrease of the average number of hypertensive and diabetic patients.	RCT	The estimates for clinically significant preoperative findings by radiography and influence on subsequent patient management were derived from a single study. Effectiveness and resource data from other studies.
Statistical Analysis (SA)/ Sensitivity analysis (SY)	-	-	-	SA: Unpaired Student's t test	-
Results	For the laparoscopic VBG patient, the average hospital charges were $12,800 vs. $16,700 for the open gastric bypass.	Surgical treatment (costs = $24000) was successful in 89% vs. med. treatment (costs = $3000) successful in 21%	For surgical group (costs = SEK 16.5 million per 100 pat.) decrease from 39 to 6 per 100 hypertensive and diabetic patients opposed to 65 to 25 for the conventional treatment (costs = SEK 15,5 million per 100 pat).	The mean closure time was 13.3 minutes in the Ethibond group vs. 9.1 minutes in the PDS group ($p<0.0001$). Although the cost per suture pack for #1 PDS was $4.63 vs. $1.76 per pack of #1 Ethibond, the mean cost of sutures per patient was $4.81 for PDS vs $11.09 for Ethibond, due to the lower number of sutures.	By sparing a patient the radiologic examination,$741 in charges would be saved, which amounts to an annual figure of $111,150 at an annual rate of 150 procedures performed.
Conclusion	Laparoscopic VBG is cost-saving and feasible.	Surgical treatment is more cost-effective at producing and maintaining weight loss.	Insufficient information to make a complete calculation.	#PDS place in continuous fashion is more secure and cost-effective operation.	GI series can be omitted as a routine preoperative test in the evaluation of a gastric bypass.
Funding	-	-	-	-	Partly funded by the Faculty Practice Plan and by institutional support of the Department of Surgery of East Carolina University School of Medicine.

Author(-s) Surgery	Van Gemert et al. (1999)	Cooney (2001)	Craig & Tseng (2002)	Demaria et al. (1999)	Shope et al. (2003)
Aim	Determine the costs of illness and the cost-effectiveness of vertical banded gastroplasty (VBG) for the treatment of morbid obesity.	Assess the impact of the introduction of a new clinical pathway on resource utilisation, the costs of care and patient health.	To compare the lifetime expected costs and outcomes between gastric bypass and no treatment. The study was carried out from the perspective of the.	To perform a cost-effectiveness analysis of hand-assisted laparoscopic surgery (HALS) versus open GB in the surgical treatment of obesity.	To investigate the effectiveness of two different laparoscopic techniques, the circular end-to-end (EEA) and linear cutting (GIA) staplers.
Perspective of the analysis	Societal	-	Third-party payer	-	Single provider (hospital) perspective
Country	The Netherlands	USA	USA	USA	USA
Data	Single study for the intervention + review of studies and authors' assumptions	Single study	Review and synthesis of completed studies + expert opinion.	Single study	Single study
Method	A prospective, before-and-after study was carried	Prospective case-control	Deterministic decision analysis model	Prospective observational study	Retrospective cohort study
Statistical Analysis (SA)/ Sensitivity analysis (SY)	SA: appropriate tests SY: on the prevalence of obesity, the complication rate, and the definition of surgical failure.	SA were performed	SA: costs = deterministically SY: One- and two-way analyses	SA: cost = stochastically. Comparison and analysis; t-test and chi-squared analysis.	SA: costs were reported as descriptive statistics (mean values plus standard deviations).
Results	3.6 life years were gained with VBG. The improvement on the VAS scale was 0.25, resulting in 12 QALYs gained when a lifelong scenario was considered.	The total costs were $10,176 (+/- 788.71) in the Pre group and $8,511 (+/- 762.60) in the Post group. This difference was not statistically significant, (p=0.15).	The costs per life-year for gastric bypass in individuals (BMI=50) were $70,300 for men and $9,130,000 for women, cost per QALY for gastric bypass in individuals (BMI=50) were $10,700 for men and $5,700 for women.	The total costs were $14,725 (+/-3,089) for the HALS group and $10,281 (+/- 3,687) for the open GB.	The total costs were $9,761 (+/- 6,567) for the EEA group and $10,830 (+/- 6,588) for the GIA group, (p=0.53).
Conclusion	The treatment of morbid obesity with VBG produced more QALYs and lower costs than no treatment and should therefore be introduced or continued from a societal point of view.	The new clinic pathway for gastric bypass surgery was a safe procedure, which reduced the length of hospitalisation and a 15% reduction in the overall costs was reported.	Bypass was not cost-saving (perspective of the payer). Cost-effectiveness ratios did compare favourably with those of other accepted interventions and appear to have been robust to parameter variation.	The effectiveness findings of the study showed HALS and open GB did not differ in terms of either early or long-term postoperative outcomes.	Although operative time was significantly shorter in the linear cutting (GIA) group, other outcomes not significantly different between the study groups.
Funding	-	-	Supported by a T32 institutional training grant from the Agency for Healthcare Research and Quality, Rockville (MD), to the University of Wisconsin Program in Population Health	-	-

Author(-s) Surgery	Gould et al. (2004)	Katapodi Kastasaros (2005)	Paxton & Matthews (2005)
Aim	To assess the clinical and economic impact of LGB on co-morbid health conditions.	To evaluate the clinical effectiveness and cost of both the Mason procedure and gastric bypass.	Compare the cost-effectiveness of laparoscopic versus open GBP for the treatment of obesity. The use of a laparoscopic approach to a Roux-en-Y gastric bypass (GBP) for the treatment of obesity.
Perspective of the analysis	-	-	-
Country	USA	Greece	USA
Data	Single study	Single study	The effectiveness data were taken from studies published between 1984 and 2004. The costs were estimated from data from the years 2002 and 2004. All costs were adjusted to 2004 values.
Method	This was a within-group comparison study that was carried out at a single institution	Details given were consistent with a cohort study where groups were defined by the patients' exposure to the treatments of interest	Exploring differences between the primary studies in the discussion. They discussed the differential reporting of complications by different authors and the differences between participants
Statistical Analysis/ Sensitivity analysis	SA: Wilcoxon rank sum test.	SA: The costs were treated deterministically	SA: The resource use and cost data were treated as point estimates. SY: The generalisability of the results was investigated by calculating costs adjusted for gender.
Results	The average number of medications per patient preoperatively was 3.7 (+/- 2.5). After surgery, patients took an average of 1.7 (+/- 1.6) prescription medications. The cost of all prescription medications was $217.6 (+/- 189) before and $97.3 (+/- 107) after surgery (monthly cost-savings $120.3.	The indicator of quality of life was 1.53 (SD=0.51) in the Mason group and 2.42 (SD=0.52) in the bypass group, (p=0.001). The total cost per patient was EUR 2,596.83 in the Mason group and EUR 4,311.23 in the bypass group. The difference was EUR 1,714.40.	The total cost of laparoscopic GBP was $17,660, compared with $20,443 for open GBP.
Conclusion	LGB resulted in a significant improvement in co-morbid health conditions 6 months after surgery.	Gastric bypass achieved a bigger reduction BMI after the 6 months-follow-up, although the Mason procedure was the cheaper alternative.	Laparoscopic GBP is a more cost-effective weight loss method than open GBP.
Funding	-	-	-

Author(-s) Medicines	Foxcroft & Milne (2000)	Warren et al. (2004)	Malone et al. (2005)	Lacey et al. (2005)
Aim	Carry out a systematic review to report the potential benefits, disbenefits and costs of orlistat for the treatment of obesity	Examine the cost-effectiveness of 1 year of SIB treatment added to diet and lifestyle advice, compared with diet and lifestyle advice alone, for the treatment of obesity in the UK.	Assess the cost-effectiveness of sibutramine in conjunction with a WMP, compared with a WMP alone, for the treatment of overweight and obese patients in the USA.	Assess the cost-effectiveness of orlistat combined with a calorie-controlled diet, compared with a calorie-controlled diet alone, for the treatment of overweight and obese patients in Ireland.
Perspective of the analysis	-	National Health Service (NHS)	Managed care	Health care service
Country	UK	UK	USA	Ireland
Data	Review and synthesis of published studies	Synthesis of completed studies and authors' assumptions.	Single study	Review or synthesis of published studies.
Method	Assessment of the abstracts and ordered only randomised controlled trials or systematic reviews of randomised controlled trials.	Hypothetical cohort of obese patients, BMI ≥ 30 kg/m2 vs. healthy individuals. Decision analysis tree was constructed.	RCT	A decision analysis tree was constructed.
Statistical Analysis/ Sensitivity analysis	The costs were treated deterministically. SY: Multi-way and extreme values sensitivity analyses.	SA: The costs were treated deterministically. SY: Univariate and multivariate sensitivity analyses.	SA: Wilcoxon rank sum test, non-parametric statistical procedures, regression analysis. SY: Univariate analyses, multivariate analyses, a non-parametric bootstrap.	SA: deterministic, one-way, sensitivity analyses
Results	If 100 patients were treated with orlistat for one year they would gain 1.60 QALYs. If 100 patients were treated with orlistat for one year the base-case cost would be Â£73,436.	The incremental cost per QALY of SIB in comparison with placebo was Â£4,780 in the UK setting and $9,299 in the US setting.	In the multivariate analysis, the strongest predictor of total health care costs was the study arm, with those participants receiving sibutramine plus WMP having an estimated annual cost of $474 more than those in the control group, (p<0.001).	The ICER for orlistat was EUR 16,954 per QALY gained.
Conclusion	Orlistat may be effective in reducing weight for some obese people.	One year of sibutramine (SIB) combined with diet and lifestyle advice was a cost-effective treatment for obesity in the UK and USA.	Sibutramine plus a weight management programme (WMP) led to a significant greater reduction in weight than the WMP alone.	Orlistat is effective and cost-effective in obese patients if, after 3 months of treatment, only treatment responders continue treatment.
Funding	Supported by the Wessex Development and Evaluation Service.	Supported by a grant from Abbott.	Funded by the Knoll Pharmaceuticals Company.	-

Author(-s) Diabetes type 2	Segal et al. (1996)	Segal et al. (1998)	Lamotte et al. (2002)
Aim	Assess the cost-effectiveness of six different strategies for the primary prevention of non-insulin dependent diabetes mellitus (NIDDM)	Assess the cost-effectiveness of alternative programmes for the primary prevention of non-insulin dependent diabetes mellitus (NIDDM).	To assess the long-term clinical consequences of weight loss, and the associated costs of treating obese Type 2 diabetic patients with orlistat.
Perspective of the analysis	-	-	Health care consumer
Country	Australia	Australia	Belgium
Data	Effectiveness data were derived from two clinical studies, a database from the Melbourne University/Mercy unit, literature reviews, expert opinions, and a Markov model.	The opinion of the research team and their clinical advisors. The final health outcomes were derived from a Markov model.	Review of completed studies
Method	Modelling	Modelling: A Markov model was constructed to estimate the cost-effectiveness of the programmes using transition probabilities, resource utilisation and cost data.	Modelling: A Markov state transition model was used to predict the complication rates and mortality over a 10-year period, with and without treatment with orlistat for 2 years.
Statistical Analysis/ Sensitivity analysis	SY: values for programme cost, programme effectiveness, and the discount rate.	SY: A set of one-way sensitivity analyses was performed on all key parameters.	SY were performed
Results	The workplace programme was the most cost-effective strategy with a cost-effectiveness ratio of Aus$300 and Aus$500, respectively. The least cost-effective programme in terms of these two criteria was the GP advice, with a cost-effectiveness ratio of Aus$10,600 and Aus$3,200, for a mixed population of NGT and IGT. The corresponding values for the bariatric surgery were Aus$3,200 and Aus$12,100, for the IGT target population.	Type I for IGT only, type IV for all groups of participants and type V programmes had net savings (the values were not given). The remainder of the programmes had a range of cost-effectiveness ratios from Aus$1,000 for type VI to Aus$12,300 for type III.	For obese diabetic patients without hypercholesterolaemia and arterial hypertension at the beginning of the study, the cost per LYG when treated with orlistat in comparison with placebo was 19,986 euros.
Conclusion	There is some evidence that preventive primary programmes for NIDDM based on workplace, community and media interventions are both effective and efficient.	Research suggests that there are alternatives - that NIDDM is a disease for which prevention options exist and that interventions may be cost-saving or highly cost-effective relative to other possible uses of health care resources".	Orlistat was cost-effective in the treatment of obese diabetic patients, especially in the presence of hypercholesterolaemia and/or hypertension.
Funding	The Health Economics Unit of the CHPE receives core funding from the Public Health Research and Development Committee of the National health and Medical Research Council, Monash University and the Victorian Health Promotion Foundation. The Program Evaluation Unit of the CHPE receives core funding from the Victorian Health Promotion Foundation and The University of Melbourne. Addition funding for the research project was provided by the Public Health Division of the Department of Human Services, Victoria.	Financial support from the Department of Human Services, Victoria. Core funding for the Centre for Health Program Evaluation from the National Health and Medical Research Council, Monash University and the Victorian Health Promotion Foundation.	Funded by the Medical Product Agency, Uppsala, Sweden.

Author(-s) Diabetes type 2	Maetzel et al. (2003)	Rowe et al. (2005)	Shearer et al. (2006)	Shearer et al. (2006)
Aim	Evaluate the cost-effectiveness of orlistat in addition to standard type-2 diabetes treatment for the treatment of overweight and obese patients.	Determine the overall use of drugs, the related costs, and the effectiveness of treating diabetic patients with orlistat.	Assess the cost-effectiveness of conventional therapies, represented by MET monotherapy or MET+SUL, in comparison with combination therapy based on ROS in patients with Type 2 diabetes.	Assess the cost-effectiveness of ROS+MET, compared with MET in combination with either sulfonylureas (SUL) or bedtime insulin (INS), in obese or overweight patients.
Perspective of the analysis	Health care provider	-	Sickness funds	Spanish National Health System (SNS)
Country	USA	UK	Germany	Spain
Data	Effectiveness data were derived from a review of the literature, supplemented with authors' assumptions.	Single study	Multiple studies: Studies published between 2001 and 2006. Most of the resource use data and some cost data were derived from a database published in 2002. Other resources and costs were obtained from studies published in 2001 and 2002. The price year was 2001.	A synthesis of published studies.
Method	Markov state transition model	This was a within-group comparison study with data collected in a single centre, with a follow-up period of 24 months.	Interconnected Markov models	The DiDACT is a long-term economic model series of interconnected Markov models
Statistical Analysis/ Sensitivity analysis	SY: Two-way sensitivity analyses, first and second order Monte Carlo > probabilistic sensitivity analysis. The cost-effectiveness acceptability curve was constructed to represent uncertainty in the model estimates.	SA: Statistical comparisons between the mean costs at baseline and at 6 months were reported. Statistical comparisons between the mean costs at baseline and at 6 months were reported.	SA: The costs were treated deterministically. SY: Univariate sensitivity analyses were carried out	SY: A series of univariate sensitivity analyses
Results	When raw HbA1c values and 1-year persistence were used the incremental cost-effectiveness ratio was reported to be $23,574 per event-free life-year gained.	The average cost of the diabetic treatments was Â£1.16 at baseline versus Â£0.83 at 6 months. The cost of insulin for those on insulin therapy was Â£1.92 per day at baseline versus Â£1.33 after 6 months on orlistat.	In the cohort of obese patients, the incremental cost per LY gained was EUR 33,787 with ROS+MET over MET+SUL and EUR 22,004 with ROS+MET over MET alone. The incremental cost per QALY gained was EUR 17,523 with ROS+MET over MET+SUL and EUR 8,669 with ROS+MET over MET alone.	In the cohort of obese patients, the incremental cost per LY gained with ROS+MET was EUR 27,225 in comparison with MET+SUL and EUR 29,860 in comparison with MET+INS. The incremental cost per QALY gained with ROS+MET was EUR 23,514 in comparison with MET+SUL and EUR 11,174 in comparison with MET+INS.
Conclusion	Orlistat is a cost-effective therapy in the management of overweight and obese patients with type-2 diabetes in the US.	The use of orlistat improved both weight and glycaemic control. It also reduced the intake of antidiabetic treatments and, consequently, their associated costs.	The cost-effectiveness ratios of rosiglitazone (ROS) in combination with other oral agents for the treatment of Type 2 diabetes in Germany fell below international thresholds for the choice of efficient health technologies.	Rosiglitazone (ROS) in combination with metformin (MET) was a cost-effective treatment for both obese and overweight patients who failed MET monotherapy in Spain.
Funding	Supported by a grant from Roche Pharmaceuticals, Switzerland.		Supported by GlaxoSmithKline	Supported by GlaxoSmithKline.

Author(-s) Diets	Siggaard et al. (1996)	Olsen et al. (2005)
Aim	Assessing the cost-effectiveness of using twelve weeks' intensive instruction versus no instruction to achieve and maintain an ad libitum carbohydrate-rich diet in overweight and normal-weight subjects.	Examining nutritional counselling by a general practitioner (GP) or by a dietician to patients with obesity and dyslipidaemia, both risk factors for ischaemic heart disease (IHD).
Perspective of the analysis	-	Societal
Country	Denmark	Denmark
Data	Single study	Single study
Method	Nonrandomised study with concurrent controls conducted in a free-living, normal-weight and overweight employees at a work-site.	This was a prospective, randomised controlled trial.
Statistical Analysis/ Sensitivity analysis	-	SA:The costs were treated deterministically in the base-case. SY: Univariate sensitivity analyses GPs). Finally, a probabilistic sensitivity analysis was carried out using the non-parametric bootstrap method (10,000 iterations).
Results	The cost of losing 1 kg body weight was $14.7 per person, while the cost of a 1% reduction in percentage over-weight was $11.9 per person. The cost of a gram change in daily carbohydrate intake was $1.6 per person.	In comparison with no intervention, the additional cost per extra LYG Dkr 8,213 (95% CI: 5,910 - 12,850) with GP counselling and Dkr 59,987 (95% CI: 30,545 - 996,368) with dietician counselling. The intervention was most cost-effective among men counselled by a GP. In comparison with no intervention, the additional cost per extra LYG without IHD was Dkr 4,670 (95% CI: 3,480 - 6,905) with GP counselling and Dkr 23,469 (95% CI: 16,223 - 41,912) with dietician counselling.
Conclusion	The study showed that the ad libitum dietary principle was successfully applied at a work site with low cost and relatively high effectiveness.	Nutritional counselling by a general practitioner (GP) was the most cost-effective strategy for the treatment of patients with obesity and a high risk of ischaemic heart disease (IHD). However, it was stated that nutritional counselling by dieticians was also cost-effective.
Funding	-	-

Author(-s) Behaviour	Meyers et al. (1996)	Pritchard et al. (1999)	Wylie-Rosett et al. (2001)	Katz et al. (2002)
Aim	Evaluating the efficacy of television delivery of a behavioural weight reduction programme.	Examining the cost-effectiveness of nutritional counselling in general practice for overweight individuals and patients with hypertension and type II diabetes. The two interventions compared were a doctor and dietician providing the counselling or a dietician only.	Aimed to evaluate weight-loss outcomes and the effect on cardiovascular disease risk factors, and the resources required for various approaches, a do-it-yourself workbook approach, an expert computer system together with the workbook (intermediate intervention); and staff consultation in addition to the computer system and workbook (intensive intervention), to weight loss.	Examining in the study was a skill-based intervention (SBI) developed to provide skills for overweight and obese subjects. Subjects in the SBI group had telephone and e-mail access to the dietician for one year.
Perspective of the analysis	-	-	Health Maintenance Organisation (HMO)	-
Country	USA	Australia	USA	USA
Data	Single study	Single study	Single study	Single study
Method	RCT	RCT	RCT	Prospective, randomised, clinical trial.
Statistical Analysis/ Sensitivity analysis	-	-	-	SA: Costs were treated deterministically
Results	The mean body weights were: videotaped group: 82.06kg before and 77.93kg after treatment, television delivered group: 88.77kg before and 84.55kg after, live contact group: 86.55kg before and 82.06kg after, waiting list control group: 91.72kg before and 90.86kg after.	Compared with the control group, the cost for an extra kilogram of weight loss was Aus$9.76 for the doctor/dietician group and Aus$7.30 for the dietician group.	The mean cost per lb lost was $6.23 for the workbook alone approach, $8.57 for the intermediate intervention and $18.78 for the intensive intervention.	The average cost per lb lost was $21 (range: $12 - $205) with CBI and $46 (range: $18 - $516) with SBI.
Conclusion	In this study, a behavioural weight reduction programme delivered by television was as effective as live-contact programmes. There were significant weight losses for the three treatment groups but not for the waiting-list control group.	General practice can provide health promotion to reduce weight and hypertension in patients, at a reasonable cost over a 12-month period, with a combination of a doctor and a dietician. Although the GP contribution increases the cost of health promotion, it has fewer patients dropping out and better outcomes than those achieved by the dietician alone.	The computerised tailoring of weight-control programmes may facilitate the implementation of weight-loss programmes in HMOs. The authors argued that the cost per participant was quite modest in each of the three interventions studied.	The innovative skill-based intervention, delivered in group sessions, was not as effective as the standard individualised approach for weight loss in obese and overweight women in the short-term. The skill-based intervention also led to an increase in treatment costs.
Funding	Supported by a Tennessee Centres of Excellence grant to the Department of Psychology of the University of Memphis.	Research funded by a grant from the Western Australian Health Promotion Foundation.	Supported by the National Heart, Lung and Blood Institute and the Diabetes Research and Training Centre.	Supported by Grant from the Centres for Disease Control and Prevention.

Author(-s) Mixed Intervention Programmes	Goldfield et al. (2001)	Wang et al. (2003)	Kumpusalo et al. (1996)	Fries & Mcshane (1998)
Aim	Focus on childhood obesity and the effect of a 13-session programme (diet, activity, behavioural change techniques, parenting and coping, to determine the cost-effectiveness of two protocols for the delivery of family-based behavioural treatment of childhood obesity)	Assess cost-effectiveness and cost-benefit of Planet Health, a school-based intervention to reduce obesity in youth of middel-school age.	Investigate the cost-effectiveness of a health promotion programme versus no programme.	Assess the cost-effectiveness of implementing a set of heath education programmes.
Perspective of the analysis	Health service	Social perspective	-	-
Country	USA	USA	Finland	USA
Data	Single study of 24 families divided in 2 groups. Therapist vs. Group treatment.	Multiple studies reporting efficacy of Planet Health and the NHANES I EFS.	Single study of four villages in 1986 en 6 villages in 1989 (524 people).	Single study of unscreened individuals vs. high – risk individuals for multiple risk factors.
Method	RCT	Decision model > cost-effectiveness Two-stage overweight progression model > number of overweight adults.	Non-randomized quasi –experimental study with concurrent controls.	Prospective cohort using questionnaires.
Statistical Analysis/ Sensitivity analysis	SY: One Way Analysis of Variance	SA: Data= deterministic, Costs= point estimates SY: One- and Multi-way sensitivity analyses > data. Separate univariate analyses > intervention	-	SA: Two-tailed t-tests
Results	Cost-effectiveness at 12 months: a decrease of 0.014 percentage overweight units per dollar and 0.01 Z- BMI units per dollar for the group treatment compared with a decrease of 0.005 percentage overweight units per dollar and 0.0004 Z- BMI units per dollar for the mixed group.	Planet Health was associated with savings of $7,313. The incremental cost per QALY gained was $4,305 when Planet Health was compared with no intervention.	There were no decreases in diastolic blood pressure or body mass indices. Increase of the HDL-cholesterol, HDL-cholesterol, plasma vitamin C concentrations in both intervention and control villages. Decrease of systolic blood pressure higher in intervention village than control villages.	The percentage reduction (absolute change) in total costs over six months compared to baseline was 25% ($484) in the overall high-risk group, 13% ($87) in the comparison employee group, and 9% ($120) in the comparison senior group for differences between all three groups.
Conclusion	Family-based behavioural treatment for childhood obesity was more cost-effective than in a combined or individual- based approach	The Planet Health programme was cost-effective and cost-saving. The authors also concluded that school-based prevention programmes of this type were likely to be cost-effective uses of public funds.	The programme was cost-effective without any synthesis of costs and benefits.	Effective health education programmes can result in larger changes in use and costs in high-risk persons than in unscreened persons, justifying more intensive educational interventions in high-risk groups.
Funding	In part by a grant from the National Institutes of Diabetes and Digestive Diseases, National Institute of Health	-	National board of Health and National Board of Education	-

Annex 3

Quality scores of economic evaluations

Surgery

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Chua Mendiola 1995	6.98	0	7.36	1.2	0	0	4.46	0	7.31	3.01	6.54	7.11	3.115	0	7.05	0	-0.95	54.14
Martin et al. 1995	6.98	3.91	7.36	1.2	0	2.895	4.46	0	3.655	3.01	3.27	3.555	3.115	0	3.525	0	-0.95	45.99
Sjostrom et al. 1995	3.49	0	3.68	1.2	0	2.895	2.23	0	3.655	0	0	3.555	3.115	0	3.525	0	-0.95	26.4
Brolin 1996	6.98	1.955	7.36	1.2	4.43	0	4.46	3.265	7.31	3.01	3.27	3.555	6.23	2.74	3.525	0	-0.95	58.34
Ghassemian et al. 1997	3.49	1.955	3.68	0.6	0	2.895	2.23	3.265	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	40.07
Gemert van et al. 1999	6.98	3.91	3.68	1.2	4.43	2.895	4.46	6.53	3.655	3.01	3.27	3.555	3.115	2.74	3.525	0	-0.95	56.01
Cooney et al. 2001	6.98	1.955	3.68	1.2	4.43	2.895	4.46	0	3.655	3.01	3.27	3.555	3.115	0	3.525	0	-0.95	44.78
Demaria et al. 2002	6.98	1.955	3.68	0.6	4.43	2.895	2.23	3.265	3.655	3.01	3.27	3.555	3.115	0	3.525	0	-0.95	45.22
Craig Tseng 2002	3.49	3.91	3.68	0.6	4.43	2.895	2.23	3.265	3.655	3.01	3.27	3.555	3.115	2.74	3.525	2.77	-0.95	49.19
Shope et al. 2003	6.98	3.91	3.68	1.2	0	0	2.23	0	7.31	3.01	3.27	3.555	3.115	0	3.525	0	-0.95	40.84
Paxton et al. 2005	6.98	3.91	3.68	1.2	4.43	2.895	4.46	0	7.31	3.01	3.27	7.11	6.23	5.48	7.05	0	-0.95	66.07
Gould et al. 2004	6.98	1.955	3.68	1.2	4.43	5.79	4.46	3.265	3.655	3.01	3.27	7.11	3.115	2.74	3.525	0	-0.95	57.24
Katopodi Katostaras 2005	3.49	1.955	3.68	0.6	0	0	2.23	0	3.655	3.01	3.27	3.555	3.115	0	3.525	0	-0.95	31.14

Medicines

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Warren et al. 2004	6.98	1.955	3.68	1.2	4.43	2.895	4.46	6.53	7.31	3.01	3.27	3.555	3.115	2.74	7.05	2.77	-0.95	64
Lacey et al. 2005	6.98	1.955	3.68	1.2	4.43	5.79	4.46	6.53	7.31	3.01	6.54	7.11	3.115	2.74	7.05	0	-0.95	70.95
Malone et al. 2005	6.98	3.91	3.68	1.2	8.86	5.79	4.46	3.265	3.655	6.02	3.27	7.11	6.23	2.74	7.05	2.77	-0.95	76.04
Foxcroft Milne 2000	6.98	0	3.68	1.2	4.43	2.895	4.46	3.265	3.655	3.01	3.27	3.555	3.115	2.74	3.525	2.77	-0.95	51.6

Type 2 diabetes

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Maetzel et al. 2003	6.98	3.91	3.68	1.2	4.43	2.895	4.46	6.53	7.31	3.01	6.54	7.11	3.115	5.48	7.05	2.77	-0.95	75.52
Rowe et al. 2005	3.49	0	3.68	0.6	8.86	0	4.46	0	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	43.04
Shearer et al. 2006 sp	6.98	0	3.68	1.2	4.43	2.895	4.46	0	7.31	3.01	3.27	7.11	3.115	0	3.525	2.77	-0.95	52.81
Shearer et al. 2006 gr	3.49	3.91	3.68	0.6	4.43	5.79	2.23	3.265	7.31	3.01	3.27	3.555	3.115	2.74	3.525	2.77	-0.95	55.74
Segal et al. 1996	6.98	1.955	3.68	0.6	4.43	2.895	4.46	6.53	7.31	3.01	3.27	7.11	3.115	0	3.525	2.77	-0.95	60.69
Segal et al. 1998	3.49	1.955	3.68	1.2	4.43	0	2.23	6.53	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	45.47
Lamotte et al. 2002	3.49	3.91	3.68	1.2	4.43	2.895	4.46	3.265	3.655	3.01	3.27	3.555	3.115	2.74	3.525	2.77	-0.95	52.02

Diets

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Olsen et al. 2005	6.98	3.91	7.36	1.2	4.43	0	4.46	6.53	3.655	3.01	3.27	3.555	3.115	2.74	3.525	0	-0.95	56.79
Siggaard et al. 1996	3.49	1.955	3.68	0.6	0	2.895	2.23	0	0	0	0	3.555	3.115	0	3.525	0	-0.95	24.1

Behaviour

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Meyers et al. 1996	3.49	1.955	3.68	1.2	0	0	4.46	3.265	0	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	36.35
Pritchard et al. 1999	6.98	1.955	3.68	0.6	0	2.895	2.23	0	3.655	3.01	3.27	3.555	3.115	2.74	3.525	2.77	-0.95	43.03
Wylie-Rosett et al. 2001	6.98	3.91	7.36	1.2	0	2.895	4.46	3.265	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	52.02
Katz et al. 2002	6.98	0	7.36	1.2	4.43	2.895	4.46	3.265	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	52.54

Mixed intervention programs

Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	C	
Weights	6.98	3.91	7.36	1.2	8.86	5.79	4.46	6.53	7.31	6.02	6.54	7.11	6.23	5.48	7.05	2.77	-0.95	Total
Kumpusalo et al. 1996	3.49	3.91	3.68	1.2	0	0	2.23	0	3.655	3.01	3.27	3.555	3.115	0	3.525	2.77	-0.95	36.46
Fries & Mcshane 1998	6.98	1.955	3.68	1.2	4.43	2.895	4.46	0	0	3.01	3.27	3.555	3.115	2.74	3.525	0	-0.95	43.87
Goldfield et al. 2001	6.98	3.91	7.36	1.2	4.43	2.895	2.23	0	7.31	3.01	6.54	7.11	3.115	2.74	7.05	2.77	-0.95	67.88
Wang et al. 2003	6.98	3.91	7.36	1.2	4.43	5.79	4.46	6.53	7.31	3.01	6.54	7.11	3.115	5.48	3.525	0	-0.95	75.8