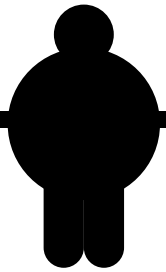


Systematic evaluation, replication and validation of structural health economic modelling approaches: lessons learned in the field of obesity

by Björn Schwander

ISBN: 978-94-6419-670-2

Printing: Gildeprint Enschede, gildeprint.nl
Layout and design: Eduard Boxem, persoonlijkproefschrift.nl

Systematic evaluation, replication and validation of structural health economic modelling approaches: lessons learned in the field of obesity

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. Pamela Habibović;
in accordance with the decision of the Board of Deans,
to be defended in public on
Monday, 6th March 2023, at 10:00 hours

by

Björn Schwander

Supervisors:
Prof. Dr. Silvia Evers
Dr. Mickaël Hiligsmann

Co-supervisor:
Dr. Mark Nuijten, A2M

Assessment Committee
Prof. Dr. Aggie Paulus (chair)
Prof. Dr. Carmen Dirksen
Prof. Dr. Stef Kremers
Dr. Karen Freijer, PON Rotterdam
Dr. Ken Redekop, Erasmus Univesity Rotterdam

Table of Contents

CHAPTER 1

General Introduction

1. General Introduction

This introductory chapter provides the general background of the research presented in this dissertation, entitled "Systematic evaluation, replication and validation of structural health economic modelling approaches: lessons learned in the field of obesity". The introduction begins with a basic description of the research field and the underlying research approaches and methods. These include cost-effectiveness analyses, decision analyses and modelling, systematic reviews, followed by research integrity, model transparency and validation. The chapter concludes with a description of the aims of the dissertation, an outline of this document and the positioning of the research in terms of novelty in the international context.

1.1 Economics & Cost-Effectiveness Analyses

Economics is the study of how people allocate scarce resources for production, distribution, and consumption, both individually and collectively [1]. Consequently, health economics is a branch of economics concerned with issues related to efficiency, effectiveness, values, and behavior in the production and consumption of health and health care [2].

Cost-effectiveness analysis (CEA) of health interventions is the comparative analyses of alternative courses of action in terms of their costs and outcomes [3, 4]. The results of such CEAs are usually expressed as incremental cost-effectiveness ratios (ICER), reflecting the ratio of incremental costs and incremental health effects. The incremental health effects are expressed typically in quality-adjusted life years (QALYs), comparing a new intervention (therapy or prevention option) to the current standard of care.

In many healthcare systems, such as the Netherlands, Canada, Australia or the United Kingdom, CEA and the ICER are used (alongside other assessments) to decide whether a healthcare intervention is to be funded by health care payers. Such decisions are usually based on specific willingness-to-pay (WTP) thresholds per QALY gained; this means an intervention is regarded as cost-effective if the ICER per QALY gained comes below this WTP threshold.

1.2 Decision Analysis and Modelling

As such CEAs are usually performed for new / innovative interventions, the evidence body for decision making is often limited. Therefore, such adoption decisions need to be made under a specific uncertainty. Decision models are frequently used to simulate and describe the cost and health effects of different alternatives, consider and investigate the given uncertainty, and to reflect the long-term and lifetime consequences, that are often not reflected adequately in clinical studies.

According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), a decision model synthesizes evidence on health consequences and costs from many different sources, including data from clinical trials, observational studies, insurance claim databases, case registries, public health statistics, and preference surveys [5]. A model is furthermore defined by the ISPOR as a logical mathematical framework that permits the integration of facts and values and which links these data to outcomes that are of interest to health-care decision makers [5].

Hence, when the available data alone does not provide sufficient information, CEA decision models and the related analysis support the clinical and the economic decision making.

1.3 Systematic Reviews

A dramatic increase in published health economic studies, more specifically cost and CEA studies, has resulted in the consequent proliferation of systematic reviews with cost and cost-effectiveness outcomes [6].

Systematic reviews aim to identify, evaluate, and summarize the findings of all relevant individual studies over a health-related issue, thereby making the available evidence more accessible to decision makers [7]. Such systematic reviews were introduced as the centerpiece of evidence-based medicine and policy making [8].

Systematic reviews in the context of health economic modelling help to identify strengths and weaknesses in CEA studies, modelling methodologies, and data for modelling inputs [6].

1.4 Research Integrity, Model Transparency and Validation

Research integrity is an important driver of reliable and trustworthy research, and includes issues such as reproducibility and replicability [9]. Both aspects serve as proof that an established and documented work can be verified, repeated, and reproduced [10].

The purpose of health economic models is to provide decision makers with quantitative information about the consequences of the options being considered; for a model to be useful for this purpose, decision makers need confidence in the model's results [11].

Modelers can impart such confidence and enhance model credibility in two main ways. 1) Transparency: clearly describing the model structure, equations, parameter values, and assumptions to enable interested parties to understand the model. 2) Validation: subjecting the model to tests such as comparing the

model's results with events observed in reality [11]. We can test whether a model is reported transparently by investigating its replicability. Method replicability and reproduction of results, which in other disciplines are common criteria of adequate research reporting to assure research integrity, are gaining importance in the field of health economic modelling, and have been the subject of recent studies [12, 13]. In order to investigate and proof the validity of a health economic model, an external validation (comparing model results with real-world results) needs to be performed [11]. By definition an external validation compares a model's results with actual event data; and involves simulating events that have occurred, such as those in a clinical trial, and examining how well the results correspond [11].

1.5 Obesity

The concepts of the research approaches and methods described above are applied to the field of obesity. Obesity is a multifactorial, chronic disorder that is usually defined as a body mass index (BMI) >30 kg/m² [14]. According to the World Health Organization, obesity has reached epidemic proportions globally and is a major contributor to the global burden of chronic disease and disability [15]. A recently published systematic review has determined that this clinical burden of obesity is also associated with a substantial and increasing economic burden, and that there is an urgent need for public health measures in order to save societal resources [16]. Given this clinical and economic burden, it is of major interest for healthcare decision makers to identify effective and cost-effective programs or interventions for obesity. However, assessing the long-term clinical and economic impact of such programs or interventions on obesity is difficult, as associated risk factors (e.g. high blood pressure, hyperlipidemia, etc.) and diseases (e.g. type 2 diabetes, cardiovascular events, etc.) develop over a long period of time, ideally requiring long-term studies, which are usually not available for new / innovative interventions. Consequently, health economic modelling is particularly relevant for obesity due to the chronic nature of the obesity-associated risk factors, morbidities and related mortality; hence several health economic models have been used to inform medical decision making in the context of obesity. We selected obesity to illustrate our concepts, due to the high complexity and heterogeneity of methods used for modelling obesity. Hence the field of obesity provides an excellent area for testing replicability and assessing validity.

1.6 Outline of Dissertation & Positioning of the Research

This dissertation studies the systematic evaluation, replication and validation of structural health economic modelling approaches in the field of obesity. In particular it evaluates, replicates and validates the current structural modelling landscape in obesity with an emphasis on commonly applied obesity-associated event simulation approaches. This research aims to increase trust and confidence in the selection

and the interpretation of the results related to a specific methodological approach used as a basis for decision analytic models in obesity.

The dissertation is divided into five complementary and interconnected research steps (chapters 2 to chapters 6), which are visualized in Figure 1-1.

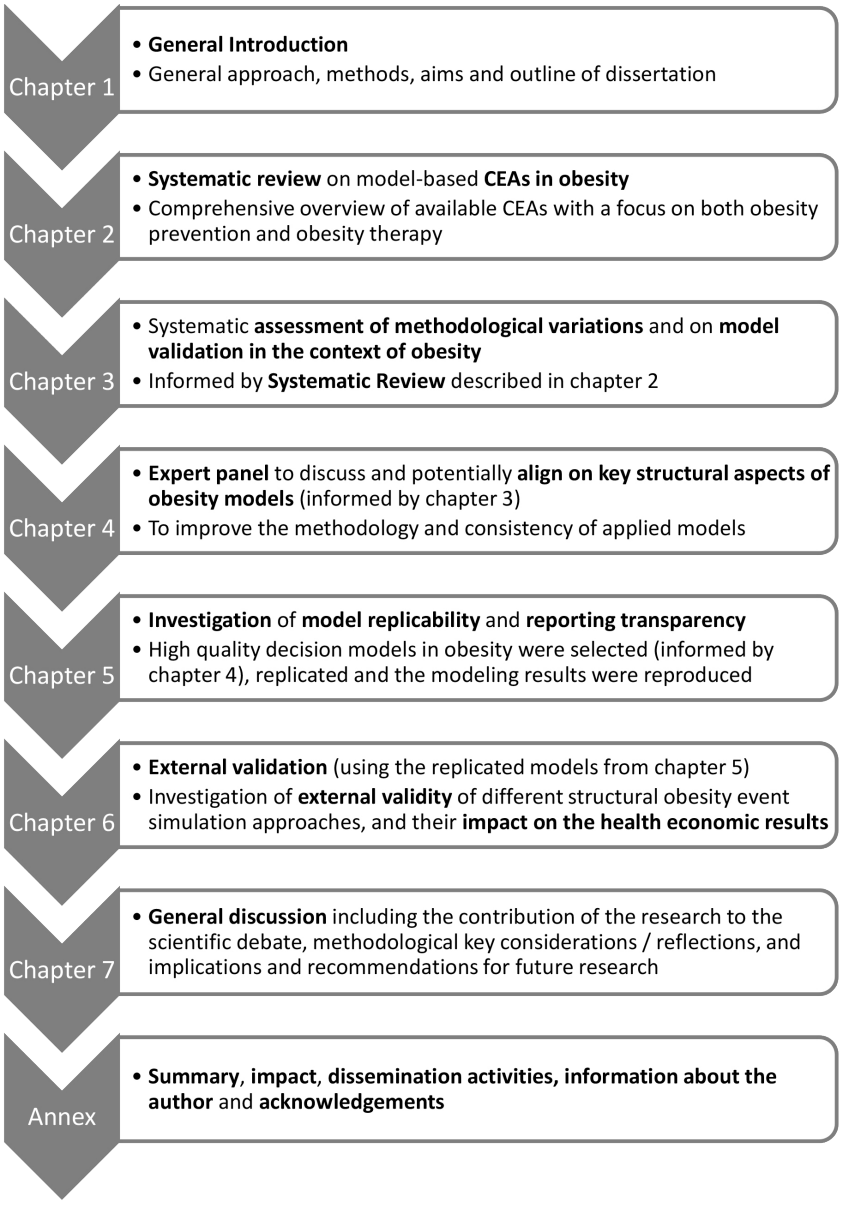


Figure 1-1. Visualization of the Outline of the Dissertation

A brief description of the content of each chapter, and the positioning of the research in relation to other published work, is provided below.

Due to the increasing burden of obesity alongside an increasing need for efficient allocation of resources, several model-based CEAs in obesity were performed, which were evaluated in published systematic reviews [17-26]. However, a comprehensive and systematic overview of such models that focuses on both obesity prevention and obesity therapy was lacking. In order to close this research gap, the aim of the research presented in **chapter 2** was to systematically review existing decision models for full health economic assessments in obesity, focusing on both obesity prevention and therapy, in order to summarize and compare their key characteristics and to identify and inform future research in this area.

One core element of each decision model is the clinical model structure. The details of the specific obesity-associated event simulation approaches are of fundamental influence, as these have a central impact on all clinical, economic and quality of life outcomes simulated by a decision model. Accordingly, the objective of our research described in **chapter 3**, was to systematically determine and describe the methodological variations in the event simulation approaches of published health economic decision models in obesity. This had not been performed at this level of detail before. In addition, the attempts of validating different event simulation approaches, by means of external validation analysis, were investigated in chapter 3.

The research described in chapter 2 and chapter 3 highlighted the need for an expert consensus on key structural aspects of obesity models, as huge variations were identified in the structural modelling approaches. Previously no consensus meeting on the structural aspects of obesity models has been performed and published in the international literature. This makes it difficult for researchers to select an appropriate approach when designing a model, and subsequently for policy makers and stakeholders to assess the quality of an applied model, intended to inform political or medical decision making. The aim of the research presented in **chapter 4** was therefore to assess and measure expert group consensus for key structural modelling approaches of obesity models, and to provide guidance to improve the methodology and consistency of applied models.

Using the results of the expert consensus, high quality decision models in obesity were selected, replicated, and the modeling results were reproduced as described in **chapter 5**. Method replicability and reproduction of results, which in other disciplines are common criteria of adequate research reporting to assure scientific rigor, are gaining importance in the field of CEA modelling, and have been the subject of recent studies [12, 13]. The research presented in chapter 5 goes beyond currently published approaches, and investigates model replication and result

reproduction in complex obesity models. We especially focused on a systematic assessment of results reproduction success, and on identifying solutions for improving current reporting standards, to enhance model transparency and replicability.

Using the successful replicated models as a basis, **chapter 6** investigated the impact of the most commonly applied structural obesity-associated event simulation approaches, on the validity of event prediction and on health economic results. This was performed in the context of obesity for the first time. The objective of the research presented in chapter 6 was to assess the external validity (in terms of clinical event prediction) of different structural obesity event simulation approaches, and to investigate their impact on the health economic results. This research could help offer better guidance for outcome researchers, health economists and decision makers when choosing and rating the structural approaches applied in health economic obesity models.

Afterwards the methodology and results of the thesis are extensively discussed and a conclusion is drawn in **chapter 7**. This includes a presentation of the main objectives and results of the research, the contribution to health economic research and to scientific debate, methodological key considerations / reflections, and implications and recommendations for future research.

Finally, additional sections provide background information on the research including a general summary, general research impact, research dissemination activities, information about the author and acknowledgements.

References

1. Hayes A. Guide to Economics. Investopedia. 2022. https://www.investopedia.com/terms/e/economics.asp. Accessed 03.06.2022.
2. Krabbe PFM. The Measurement of Health and Health Status. Elsevier Inc.; 2017.
3. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2022;25(1):10-31. doi:10.1016/j.jval.2021.10.008.
4. Noyes K, Holloway RG. Evidence from cost-effectiveness research. NeuroRx. 2004;1(3):348-55. doi:10.1602/neurorx.1.3.348.
5. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2003;6(1):9-17.
6. Mandrik OL, Severens JLH, Bardach A, Ghabri S, Hamel C, Mathes T et al. Critical Appraisal of Systematic Reviews With Costs and Cost-Effectiveness Outcomes: An ISPOR Good Practices Task Force Report. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2021;24(4):463-72. doi:10.1016/j.jval.2021.01.002.
7. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. J Family Med Prim Care. 2013;2(1):9-14. doi:10.4103/2249-4863.109934.
8. Anderson R. Systematic reviews of economic evaluations: utility or futility? Health economics. 2010;19(3):350-64. doi:10.1002/hec.1486.
9. Munafò M. Reproducibility and research integrity. In: BMC Research Notes. BioMedCentral. 2022. https://www.biomedcentral.com/collections/reproducibility. Accessed 27.09.2022.
10. Diaba-Nuhoho P, Amponsah-Offeh M. Reproducibility and research integrity: the role of scientists and institutions. BMC Res Notes. 2021;14(1):451. doi:10.1186/s13104-021-05875-3.
11. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50. doi:10.1016/j.jval.2012.04.012.
12. Bermejo I, Tappenden P, Youn JH. Replicating Health Economic Models: Firm Foundations or a House of Cards? PharmacoEconomics. 2017;35(11):1113-21. doi:10.1007/s40273-017-0553-x.
13. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and Facilitators to Model Replication Within Health Economics. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2019;22(9):1018-25. doi:10.1016/j.jval.2019.04.1928.
14. World Health Organization. Fact Sheet on Obesity. 2003. https://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf. Accessed 09.02.2019.
15. World Health Organization. Updated Fact Sheet on Obesity. 2018. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed 09.02.2019.

16. Tremmel M, Gerdtham UG, Nilsson PM, Saha S. Economic Burden of Obesity: A Systematic Literature Review. International journal of environmental research and public health. 2017;14(4). doi:10.3390/ijerph14040435.

17. Boyers D, Avenell A, Stewart F, Robertson C, Archibald D, Douglas F et al. A systematic review of the cost-effectiveness of non-surgical obesity interventions in men. Obes Res Clin Pract. 2015;9(4):310-27. doi:10.1016/j.orcp.2015.03.001.

18. Flego A, Keating C, Moodie M. Cost-effectiveness of whole-of-community obesity prevention programs: an overview of the evidence. Expert review of pharmacoeconomics & outcomes research. 2014;14(5):719-27. doi:10.1586/14737167.2014.953933.

19. Wang BC, Furnback W. Modelling the long-term outcomes of bariatric surgery: A review of cost-effectiveness studies. Best Pract Res Clin Gastroenterol. 2013;27(6):987-95. doi:10.1016/j.bpg.2013.08.022.

20. Henteleff HJ, Birch DW, Hallowell PT. Cost-effectiveness of bariatric surgery for severely obese adults with diabetes. Can J Surg. 2013;56(5):353-5. doi:10.1503/cjs.020713.

21. Griffiths UK, Anigbogu B, Nanchahal K. Economic evaluations of adult weight management interventions: a systematic literature review focusing on methods used for determining health impacts. Appl Health Econ Health Policy. 2012;10(3):145-62. doi:10.2165/11599250-000000000-00000.

22. Gandjour A. Cost-effectiveness of preventing weight gain and obesity: what we know and what we need to know. Expert review of pharmacoeconomics & outcomes research. 2012;12(3):297-305. doi:10.1586/erp.12.13.

23. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. Nutrition. 2012;28(9):829-39. doi:10.1016/j.nut.2011.11.016.

24. Lehnert T, Sonntag D, Konnopka A, Riedel-Heller S, König HH. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. Obes Rev. 2012;13(6):537-53. doi:10.1111/j.1467-789X.2011.00980.x.

25. Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess. 2011;15(2):1-182. doi:10.3310/hta15020.

26. Levy DT, Mabry PL, Wang YC, Gortmaker S, Huang TT, Marsh T et al. Simulation models of obesity: a review of the literature and implications for research and policy. Obes Rev. 2011;12(5):378-94. doi:10.1111/j.1467-789X.2010.00804.x.

CHAPTER 2

Systematic Review and Overview of Health Economic Evaluation Models in Obesity Prevention and Therapy

Chapter 2 was informed by Schwander B, Hiligsmann M, Nuijten M, Evers S: Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert Rev Pharmacoecon Outcomes Res. 2016 Sep 9:1-10. https://doi.org/10.1080/14737167.2016.1230497

2.1 Abstract

Introduction: Given the increasing clinical and economic burden of obesity, it is of major importance to identify cost-effective approaches for obesity management.

Areas Covered: This study aims to systematically review and compile an overview of published decision models for health economic assessments (HEA) in obesity, in order to summarize and compare their key characteristics as well as to identify, inform and guide future research.

Key Results: Of the 4,293 abstracts identified, 87 papers met our inclusion criteria. A wide range of different methodological approaches have been identified. Of the 87 papers, 69 (79%) applied unique / distinctive modelling approaches.

Conclusions: This wide range of approaches suggests the need to develop recommendations / minimal requirements for model-based HEA of obesity. In order to reach this long-term goal, further research is required. Valuable future research steps would be to investigate the predictiveness, validity and quality of the identified modelling approaches.

2.2 Introduction / Background

Obesity is a multifactorial, chronic disorder that has, according to the World Health Organization (WHO), reached epidemic proportions globally and is a major contributor to the global burden of chronic disease and disability [1]. Obesity is defined as abnormal or excessive fat accumulation that may impact health. The body mass index (BMI) is a simple index of weight-for-height that is commonly used to classify obesity in adults. It is defined as a person's weight in kilograms divided by the square of his height in meters (kg/m2). According to the WHO definition, a BMI ≥ 25 and < 30 is overweight; a BMI ≥30 is obesity [1]. In 2014, worldwide, more than 1.9 billion adults (≈39%), 18 years and older, were overweight. Of these, over 600 million adults (≈13%) were obese [2].

Overweight and obesity are leading risks for global deaths. In 2010, worldwide, it has been estimated that around 3.4 million adults died (≈ 6% of total deaths per year) as a result of being overweight or obese [3]. In addition, 44% of the diabetes cases, 23% of the ischemic heart disease cases and between 7% and 41% of certain cancer cases are attributable to overweight and obesity [4].

Given this clinical and its associated economic burden, it is of major interest for healthcare decision makers to identify cost-effective programs or interventions for obesity prevention and therapy. Due to the potentially high cost of such programs or interventions, an increasing number of economic evaluations have been conducted to assess their value for money. Economic evaluations are defined as the comparative analysis of alternative interventions in terms of both their costs and consequences [5]. The results of such evaluations can help public health policymakers and central HTA bodies make informed decisions, ensuring that limited resources are allocated as efficiently as possible to improve overall population health.

Assessing the long-term clinical and economic impact of such programs or interventions on obesity is difficult, as associated risk factors and diseases, such as high blood pressure, hyperlipidemia, Type 2 diabetes and cardiovascular events, develop over a long period of time, requiring long-term interventional and observational studies. Consequently, decision analytic modelling is particularly relevant in the field of obesity, due to the chronic nature of obesity-associated morbidities and related mortality. Furthermore, purely empirical evaluations (e.g. randomized controlled trials or natural experiments) often examine the effect of a limited number of programs or interventions and often only for a selected population over a limited time horizon. In these cases, modelling enables the comparison of various programs and interventions and the possibility to extrapolate short-term randomized controlled trial (RCT) data to long-term outcomes.

To date, several decision-analytic models for interventions and lifestyle changes in patients with obesity have been published, but a comprehensive and systematic overview of such models that focuses on both obesity prevention and obesity therapy is lacking. Accordingly, the aim of this study was to systematically review existing decision models for full health economic assessments in obesity, focusing on both obesity prevention and therapy, in order to summarize and compare their key characteristics and to identify and inform future research in this area.

2.3 Methods

This systematic review was conducted according to the PRISMA guidelines [6].

2.3.1 Literature Search

We have conducted a systematic review using the Pubmed Database and the NHS Economic Evaluation Database (which includes MEDLINE, EMBASE, CINAHL, PsycINFO and PubMed) to identify full health economic assessments in the context of obesity that have been published before the end of May 2015. This dataset selection reflects the strategy recommended by Alton et al. [7] and Sassi et al. [8] for identifying economic evaluation studies.

In order to identify relevant publications three different searches were performed and combined: one for health economic evaluations, one for decision models and one for obesity. For identifying health economic evaluations in Pubmed, the most sensitive search strategy proposed by the Canadian Agency for Drugs and Technologies in Health was used (NHS EED strategy OR Royle and Waugh OR Wilczynski) [9]. For identifying decision models, we combined the key words used by Goehler et al. [10] and Van Haalen et al. [11]) in previous similar studies in other diseases areas. To identify obesity-related publications (LeBlanc et al. [12] OR Briant et al. [13]), search strategies of previously published systematic reviews were applied in Pubmed and in the NHS EED database, respectively. Further details on the search strategies are provided in the Appendix.

The only limitations applied were related to the publication type as we have excluded letters, editorials and historical articles. The reference lists of retrieved papers and of relevant reviews [14,15,16,17,18,19,20,21,22,23,24,25] were also checked manually to identify additional relevant studies. Search results were exported to Reference Manager Version 12 Software (Thomson Reuters, New York, NY, USA), and duplicate articles and non-original research were removed.

2.3.2 Eligibility Criteria

Eligible studies were original research articles on decision models for full health economic assessment in the context of obesity. Full health economic assessments

(HEAs) were defined as *"the comparative analysis of alternative courses of action in terms of both their costs and consequences",* according to Drummond et al. [5]. We therefore included cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis. Decision models were defined as *"an analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources, and whose purpose is to estimate the effects of an intervention on valued health consequences and costs"* according to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Good Research Practices – Modeling Studies [26]; therefore, health economic evaluations performed alongside a clinical trial were excluded if events were not simulated/ extrapolated over time – and hence only reflect the observed study results over the follow-up period of the clinical trial. In case the study focused on obesity therapy, obesity was defined as "a BMI greater than or equal to 30", according to the WHO criteria [1]. When the study focused on the prevention of obesity, a clear statement on the obesity prevention focus was the precondition for inclusion.

2.3.3 Literature Selection

The primary screening of titles and abstracts according to the eligibility criteria was performed by one reviewer. In this process it was documented for each abstract whether the different inclusion criteria were fulfilled or rejected. In all "unclear" cases, where an inclusion criterion could not clearly be categorized as fulfilled or rejected, as well as in cases that "might have potentially met" or "might have met" the inclusion criteria, the full text article was ordered and all relevant information on the inclusion / exclusion criteria was extracted. The information on all "unclear" cases was then shared, reviewed by and discussed within the author team. In case of disagreements those were discussed in the author team in order to reach a consensus decision.

2.3.4 Data Extraction

For data extraction, a predefined template was developed and used in order to summarize information on eligibility criteria, modelling approach, primary outcome, time horizon, perspective, setting (prevention or therapy), type of intervention, target population, country and information on the obesity-related event simulation of all studies that met the eligibility criteria. Furthermore, the models were classified using the taxonomy of model structures for the economic evaluation of health technologies proposed by Brennan et al. [27]. These characteristics were categorized into specific characteristics of modelling (modelling approach, primary outcome, time horizon, perspective) and specific characteristics of the simulation setting, which defines in which context/setting the model was applied (setting, type of intervention, target population, country) and were presented accordingly.

2.4 Results

2.4.1 Literature Search

In total 4,293 studies were identified via the database searches, and 4,304 abstracts were reviewed (database search plus n=11 hand search). From these, 142 articles were selected for full-text review, and 87 papers met our inclusion criteria. 55 full text articles were excluded for the following reasons: no decision model (n=20), no full health economic assessment (n=17), not original research (n=16), not about obesity (n=2). The flow chart of study selection is shown in Figure 2-1. A detailed list of the included studies is provided in Table 2-1 in the appendix.

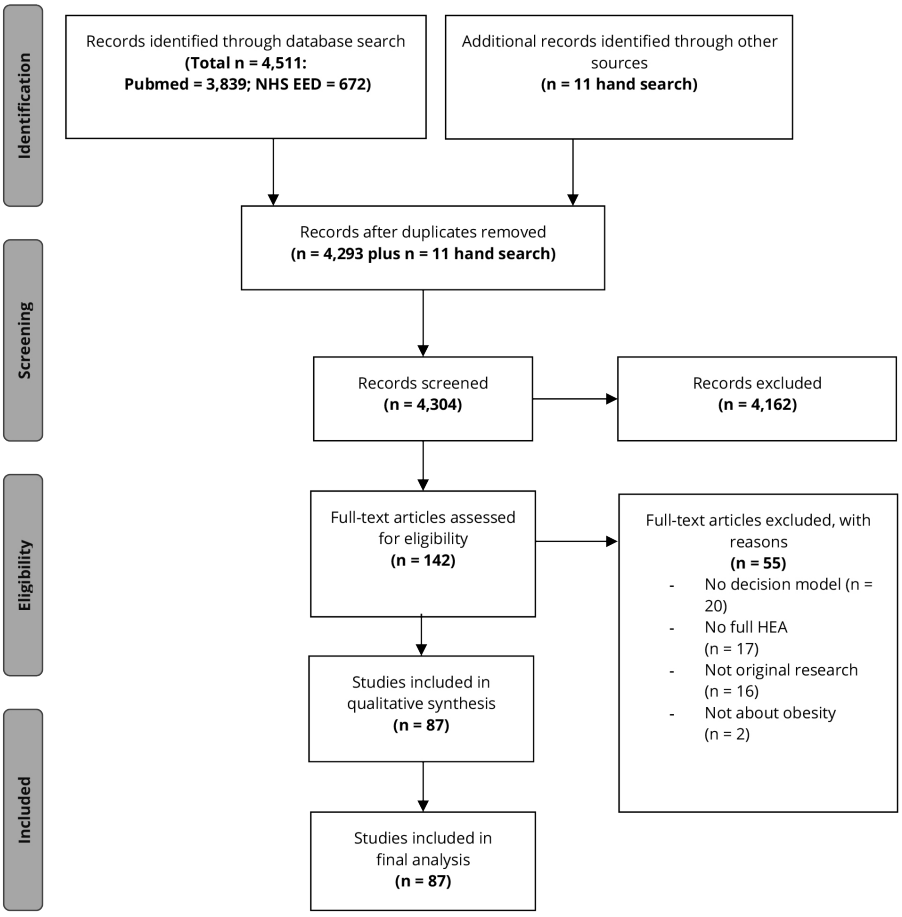


Figure 2-1. Flow Diagram of the systematic review process
NHS EED = National Health Service Economic Evaluation Database; HEA = Health Economic Assessment

2.4.2 Key Characteristics of Decision Models

The modelling-specific key characteristics are presented in Figure 2-2. The majority of models used a Markov approach (85%). This major group of Markov models (85%) could be subcategorized as "stochastic cohort models" (54%), "deterministic cohort models" (24%), "deterministic patient-level models" (6%), and "stochastic patient-level models" (1%).

Quality-adjusted life years (QALYs) were applied as the primary outcome in 69% of all studies, followed by disability-adjusted life years (DALYs) (18%), life years (LYs) (8%) and others (5%). In the models that simulated QALYs or DALYs (n=76) different approaches have been applied in order to estimate the impact of an intervention: in most cases a positive impact of a BMI reduction on the quality of life (QoL) was combined with a positive impact on the QoL by avoiding obesity associated events (38%); followed by considering a positive impact of a BMI reduction only (35%), by considering the positive impact of avoiding obesity associated events only (24%), and very rarely it was not described how an intervention effect on QoL was simulated (3%). In most cases the included QoL data was based on published sources (88%) and only rarely the QoL data was based on an own survey (9%) or on a database analysis (2%). Details on the QoL instrument were only provided in 37% of the publications and the EQ-5D (20%) was the most frequently applied method.

In most cases a lifetime horizon was applied (69%). A time horizon of >20 years but < lifetime was less frequently applied (14%), and a time horizon of <20 years was more common (23%). Most models adopted the perspective of the healthcare payer (66%), whereas a societal perspective was applied less frequently (25%). In 9% of cases the perspective was not clearly stated.

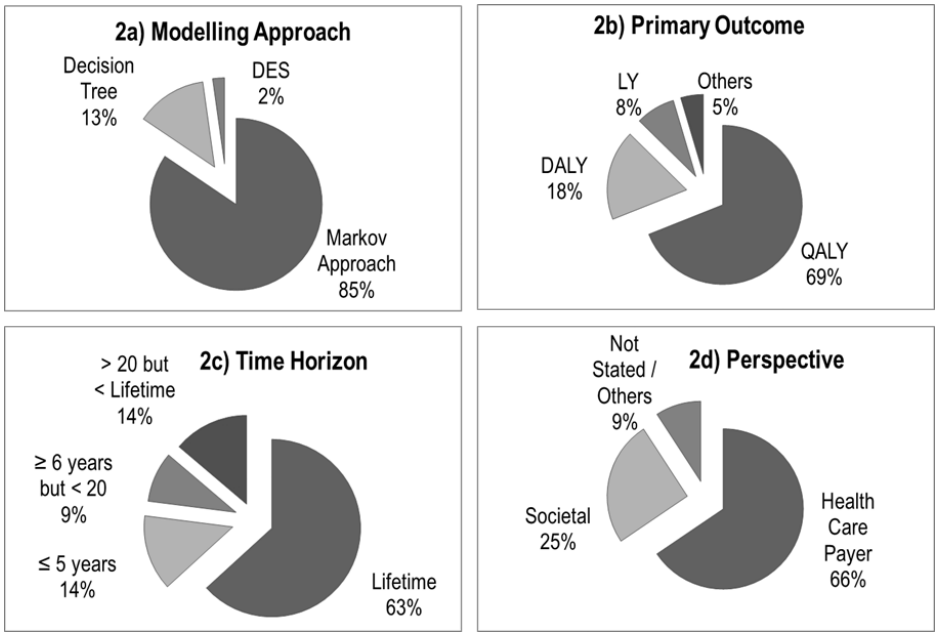


Figure 2-2. Overview of modelling specific key characteristics (modelling approach 2a, primary outcome 2b, time horizon 3c; perspective 2d) of decision models for obesity prevention and therapy
DES = Discrete Event Simulation; LY = Life Year; DALY = Disability Adjusted Life Year; QALY = Quality Adjusted Life Year

The simulation setting-specific key characteristics are presented in Figure 2-3.

Most models focused on therapy for obesity (77%). The prevention of obesity was simulated less frequently (20%) and only rarely were both aspects "therapy & prevention" simulated (3%). The type of intervention was most commonly a behavioral or public health approach (47%), more seldom but quite frequently a surgical approach (30%), less frequently a pharmacological approach (20%), and various other interventions (more than one intervention group analyzed) were investigated very rarely (3%). Adults were the most common target population (80%), whereas children were much less frequently the focus (15%) and very rarely (5%) both adults & children were simulated. Most models were simulating a European country setting (47%); less commonly simulations focused on North America (27%) or Australia (20%) and only rarely was an Asian setting (6%) simulated.

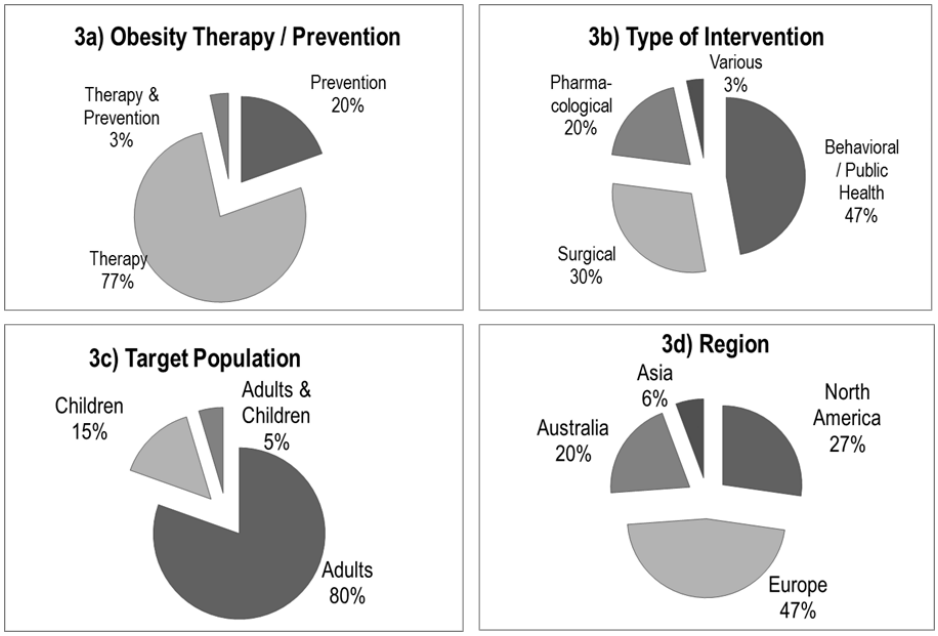


Figure 2-3. Overview of key characteristics and specifications (obesity therapy/prevention 3a, type of intervention 3b, target population 3c; region 3d) of decision models for obesity prevention and therapy

2.4.3 Obesity-associated Event Simulation

Most but not all included models simulated obesity-associated events: of the 87 decision models identified, 72 simulated obesity-associated events; in the other models the change in BMI was directly transferred into costs and effects.

As shown in Figure 2-4, most of these models simulated cardiovascular diseases (83%), mortality (81%), Type 2 diabetes (74%), and stroke (67%). A minority of the models simulated cancer (35%), osteoarthritis (24%), hypertension (11%), hyperlipidemia (11%) and peripheral arterial disease (10%). Information on an event validation was only provided in 21% of cases: most frequently an internal and external validation was performed (13%), followed by a cross-validation (3%), only internal (3%) and only external validation (3%).

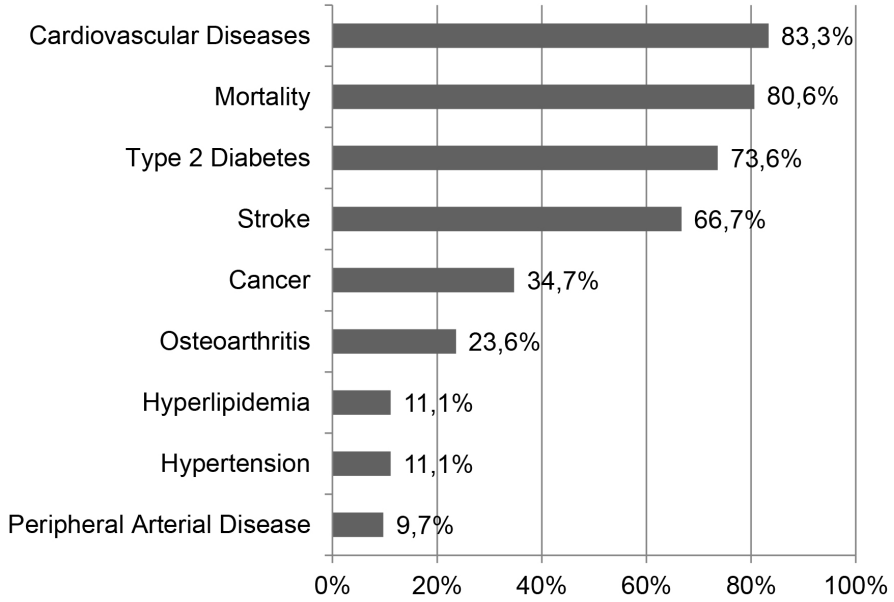


Figure 2-4. Proportion of decision models simulating specific obesity associated events
Comment: The percentages presented above are calculated on the basis of the 72 decision models that simulate obesity-associated events; the 15 remaining decision models that were excluded simulated no obesity-associated events.

In the 87 publications identified for our study, 69 unique / distinctive modelling approaches were identified. In contrast 27 (of 87) publications were based on 9 unique / distinctive modelling approaches that have been applied and published several times. The model that was applied most frequently was the Australian ACE-obesity model [29]; this was adapted and published in nine different studies.

One further key difference identified related to the approach of simulating the impact of obesity and of obesity interventions/prevention measures on costs and effects. Roughly four different approaches were identified: (1) no events were simulated, and the change in BMI was transferred directly into costs and effects; (2) events were simulated by BMI-related functions, and the change in BMI was transferred into events that subsequently impact the costs and effects; (3) events were simulated by risk equations, and the change in BMI was transferred into BMI-specific relative risks that impact events and subsequently the costs and effects; (4) the events were simulated by risk equations and the change in BMI was transferred into a change in risk factors that impact the risk equations that impact the events and subsequently the costs and effects.

2.5 Discussion

This systematic review identified 87 papers, on decision models for full health economic assessment in obesity, that were published before the end of May 2015. Of these 87 publications, 69 applied unique / distinctive modelling approaches and accordingly we have identified a broad range of unique methodological frameworks. Previous systematic reviews of economic evaluations in the context of obesity were limited to specific populations, interventions or settings. Our review provides a comprehensive overview on full HEA decision models in obesity without being limited to specific populations, interventions or settings. In comparison with a systematic review in the context of Type 2 diabetes (Yi et al. 2002) [30], that obtained a comparable number of included publications (n=78), but only 20 unique / distinctive modelling approaches (26%), the diversity of approaches is much stronger in obesity (79% are unique modeling / distinctive approaches). In the context of obesity, it seems to be the case that each research team builds its own obesity model; this is reflected by the obtained diversity of obesity modelling approaches. This makes it difficult to compare model outcomes, as the structural and methodological differences could have a major impact on the modelling results. Therefore, our review informs the need for developing recommendations and/or minimal requirements for model-based HEA of obesity in order to promote a better comparability and interpretation of modelling results.

Due to our focus on full HEAs (eligibility criteria) we might have excluded simulation approaches that were applied in epidemiologic or clinical obesity models. However, as we are aiming at informing and supporting future HEAs of interventions for the prevention and treatment of obesity, we decided to focus on available full HEAs, as those usually consider the specifications and requirements of comparative assessments and take into account the economic component of the disease. Furthermore, we have identified that there are several full HEA decision models published that focus on the prevention of obesity-associated diseases - such as Type 2 diabetes, coronary heart disease, stroke etc. – but without any clear connection to/definition of obesity. In these situations, we selected only prevention studies that have clearly stated that the intervention/public health measure focuses on the prevention of obesity or overweight; this might lead to the possibility that we might have excluded some relevant decision models. However, due to the fact that in the case of obesity prevention, the BMI-related eligibility criteria (BMI $\geq$ 30 kg/m^2) was not applicable, there was a need to define a suitable selection criteria in order to keep our research focused on the obesity topic.

Although we have identified a large number of unique modelling approaches, there are some key characteristics that are applied quite commonly throughout the models. For example, most models applied a Markov approach and simulated a

lifetime horizon. This appears logical, as a Markov model is appropriate to simulate disease with a continuous risk over time and recurrence of events [31], and is therefore an appropriate design for long-term evaluations. Furthermore, most of the decision models simulated cardiovascular diseases and Type 2 diabetes, and cohort studies have demonstrated that these diseases are two of the most important consequences of obesity [32]. In this context, it might be valuable to determine whether the addition of further obesity-associated events to decision models (e.g. cancers) has a major impact on the outcomes of a full HEA.

Furthermore, it was possible to identify roughly 4 different approaches for simulating the impact of obesity and of obesity interventions/prevention measures on costs and effects. These approaches range from very simple (change in BMI directly transferred into costs and effects) to very complex (BMI impacts risk factors of risk equations, subsequently events and subsequently costs and effects) and hence one key question for future research might be to determine how complex an analyses approach for a full HEA needs to be in order to provide valid results.

Considering the diversity of methods used, it would be important to define minimal requirement for model in obesity. Given the chronic nature of obesity as well as considering the key characteristics that are applied quite commonly by the reviewed models a first suggestion of minimal requirements for an full HEA obesity model might include the following components: Markov approach, lifetime horizon, simulates at least cardiovascular diseases, type 2 diabetes and mortality.

There are some potential limitations to our study. First, we did not perform a quality assessment of included studies and second, we did not evaluate how the modelling method could affect cost-effectiveness results. Applying a quality checklist have been considered and discussed within the author team. Due to the strong variation of (clinical) simulation approaches we have decided that investigating the impact of those different event simulation approaches and their validation is the most valuable next research step (research is currently ongoing). Rationale for this decision is that the clinical event simulation forms the foundation of the model; hence we decided to investigate this matter of predictiveness and validity first in detail, before we start assessing the qualitative aspects of the model/health economic assessment that is built/based on this clinical foundation.

Accordingly, further research is required in order to investigate the predictiveness, validity and quality of the identified modelling approaches.

2.6 Conclusions

Despite these limitations, our review provides a comprehensive overview of the model types and simulation approaches used in obesity models. On the basis of this comprehensive overview we were able to identify and inform future research in this area. These findings could be very interesting for researchers, modelers and also for policy makers, and could be a step further on the road to developing recommendations and/or minimal requirements for the model-based HEA of obesity.

2.7 References

1. World Health Organization (WHO) 2003. Fact Sheet on Obesity: Available at: http://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf. Accessed on 23rd February 2016.
2. WHO 2015. Updated Fact Sheet N 311 on Obesity and Overweight: Available at: http://www.who.int/mediacentre/factsheets/fs311/en/. Accessed on 23rd February 2016.**
3. Ng M, Fleming T, Robinson M et al.: Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384(9945):766-81.
4. WHO 2014. Updated Fact Sheet N 311 on Obesity and Overweight: Available at: http://www.wpro.who.int/mediacentre/factsheets/obesity/en/. Accessed on 23rd February 2016.
5. Drummond MF, O'Brien B, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1997.
6. Moher D, Liberati A, Tetzlaff J, et al. for the PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009;339:332-339.
7. Alton V1, Eckerlund I, Norlund A. Health economic evaluations: how to find them. Int J Technol Assess Health Care. 2006 Fall;22(4):512-7.
8. Sassi F, Archard L, McDaid D. Searching literature databases for health care economic evaluations: how systematic can we afford to be? Med Care. 2002 May;40(5):387-94.
9. Glanville J, Fleetwood K, Yellowlees A, Kaunelis D, Mensinkai S. Development and Testing of Search Filters to Identify Economic Evaluations in MEDLINE and EMBASE. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.
10. Goehler A, Geisler BP, Manne JM, Jahn B, Conrads-Frank A, Schnell-Inderst P, Gazelle GS, Siebert U. Decision-analytic models to simulate health outcomes and costs in heart failure: a systematic review. Pharmacoeconomics. 2011 Sep;29(9):753-69.
11. van Haalen HG, Severens JL, Tran-Duy A, Boonen A. How to select the right cost-effectiveness model? : A systematic review and stepwise approach for selecting a transferable health economic evaluation model for rheumatoid arthritis. Pharmacoeconomics. 2014 May;32(5):429-42.
12. LeBlanc E, O'Connor E, Whitlock EP, Patnode C, Kapka T. Screening for and Management of Obesity and Overweight in Adults [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011 Oct. Report No.: 11-05159-EF-1.
13. Bryant M, Ashton L, Brown J, Jebb S, Wright J, Roberts K, Nixon J. Systematic review to identify and appraise outcome measures used to evaluate childhood obesity treatment interventions (CoOR): evidence of purpose, application, validity, reliability and sensitivity. Health Technol Assess. 2014 Aug;18(51):1-380.
14. Boyers D, Avenell A, Stewart F, Robertson C, Archibald D, Douglas F, Hoddinott P, van Teijlingen E. A systematic review of the cost-effectiveness of non-surgical obesity interventions in men. Obes Res Clin Pract. 2015 Jul-Aug;9(4):310-27.
15. Flego A, Keating C, Moodie M. Cost-effectiveness of whole-of-community obesity prevention programs: an overview of the evidence. Expert Rev Pharmacoecon Outcomes Res. 2014 Oct;14(5):719-27.
16. Wang BC, Furnback W. Modelling the long-term outcomes of bariatric surgery: A review of cost-effectiveness studies. Best Pract Res Clin Gastroenterol. 2013 Dec;27(6):987-95.

17. Henteleff HJ, Birch DW, Hallowell PT; CAGS/ACS Evidence Based Reviews in Surgery Group. Cost-effectiveness of bariatric surgery for severely obese adults with diabetes. Can J Surg. 2013 Oct;56(5):353-5.

18. Griffiths UK, Anigbogu B, Nanchahal K. Economic evaluations of adult weight management interventions: a systematic literature review focusing on methods used for determining health impacts. Appl Health Econ Health Policy. 2012 May 1;10(3):145-62.

19. Gandjour A. Cost-effectiveness of preventing weight gain and obesity: what we know and what we need to know. Expert Rev Pharmacoecon Outcomes Res. 2012 Jun;12(3):297-305.

20. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. Nutrition. 2012 Sep;28(9):829-39.

21. Lehnert T, Sonntag D, Konnopka A, Riedel-Heller S, König HH. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. Obes Rev. 2012 Jun;13(6):537-53.

22. Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J, Welch K, Clegg A. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess. 2011 Jan;15(2):1-182.

23. Levy DT, Mabry PL, Wang YC, Gortmaker S, Huang TT, Marsh T, Moodie M, Swinburn B. Simulation models of obesity: a review of the literature and implications for research and policy. Obes Rev. 2011 May;12(5):378-94.

24. Saha S, Gerdtham UG, Johansson P. Economic evaluation of lifestyle interventions for preventing diabetes and cardiovascular diseases. Int J Environ Res Public Health. 2010 Aug;7(8):3150-95.

25. Neovius M, Narbro K. Cost-effectiveness of pharmacological anti-obesity treatments: a systematic review. Int J Obes (Lond). 2008 Dec;32(12):1752-63.

26. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. Value Health. 2003 Jan-Feb;6(1):9-17.

27. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006 Dec;15(12):1295-310.

28. Haby MM, Vos T, Carter R et al.: A new approach to assessing the health benefit from obesity interventions in children and adolescents: the assessing cost-effectiveness in obesity project. Int J Obes (Lond). 2006 Oct;30(10):1463-75.*

29. Carter R, Moodie M, Markwick A, et al.: Assessing cost-effectiveness in obesity (ACE-obesity): an overview of the ACE approach, economic methods and cost results. BMC Public Health. 2009 Nov 18;9:419.

30. Yi Y, Philips Z, Bergman G, Burslem K. Economic models in type 2 diabetes. Curr Med Res Opin. 2010 Sep;26(9):2105-18.*

31. Sonnenberg FA1, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making. 1993 Oct-Dec;13(4):322-38.

32. Field AE, Coakley EH, Must A, Spadano JL, Laird N, Dietz WH, Rimm E, Colditz GA. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. Arch Intern Med. 2001 Jul 9;161(13):1581-6.

2.8 Appendix

2.8.1 Search Strategy

NHS EED (((((Economics) OR (exp costs and cost analysis) OR (Economics, Dental) OR (exp economics, hospital) OR (Economics, Medical) OR (Economics, Nursing) OR (Economics, Pharmaceutical) OR (economic*[Title/Abstract] OR cost[Title/Abstract] OR costs[Title/Abstract] OR costly[Title/Abstract] OR costing[Title/Abstract] OR price[Title/Abstract] OR prices[Title/Abstract] OR pricing[Title/Abstract] OR pharmacoeconomic*[Title/Abstract]) OR (expenditure* NOT energy[Title/Abstract]) OR (value for money[Title/Abstract]) OR (budget[Title/Abstract])) NOT ((energy OR oxygen AND cost[Title/Abstract]) OR (metabolic AND cost[Title/Abstract]) OR (energy OR oxygen AND expenditure[Title/Abstract]))) OR **Royle and Waugh** (cost* OR economic* OR (quality adj* AND life)) OR **Wilczynski** (cost effective OR sensitivity analys* OR cost effectiveness)) AND **Goehler** (((Decision Support Techniques[MeSH Major Topic]) OR (Models, Statistical[MeSH Major Topic]) OR (Markov Chains[MeSH Major Topic]) OR (Monte Carlo Method[MeSH Major Topic])) OR **Van Haalen** (simulation OR model OR models OR modeling OR modelling OR (decision AND (analys* OR analytic)))) AND **LeBlanc** (((“Obesity”[Majr:noexp] OR “Obesity, Morbid”[Majr] OR “Overweight”[Majr:noexp]) OR (“Anti-Obesity Agents”[Majr:noexp] OR “Appetite Depressants”[Majr] OR “Anti-Obesity Agents “[Pharmacological Action] OR “Appetite Depressants “[Pharmacological Action] OR “sibutramine “[Substance] OR “orlistat “[Substance]) OR (“Bariatric Surgery”[Majr:noexp] OR “Gastric Bypass”[Majr] OR “Gastroplasty”[Majr]) OR (“Body Mass Index”[Majr] OR “Weight Loss”[Majr:noexp])) OR **Bryant** ((obesity) OR (obesity hypoventilation syndrome) OR (obesity, abdominal) OR (obesity, morbid) OR (prader-willi syndrome) OR (Weight Gain) OR (weight loss) OR (body weight changes) OR (Ideal Body Weight) OR (adiposity) OR (Overweight)))) NOT **Limits** ((Letter[Publication Type]) OR (Editorial[Publication Type]) OR (Historical Article[Publication Type]))

Table 2-1. Overview of included studies

Author	Year	Title	CVD	T2D	Stroke
Ackroyd, R.	2006	Cost-effectiveness and budget impact of obesity surgery in patients with type-2 diabetes in three European countries	0	0	0
Ananthapavan, J.	2010	Assessing cost-effectiveness in obesity: laparoscopic adjustable gastric banding for severely obese adolescents	0	0	0
Annemans, L.	2007	Health economic evaluation of controlled and maintained physical exercise in the prevention of cardiovascular and other prosperity diseases	1	1	1
Anselmino, M.	2009	Cost-effectiveness and budget impact of obesity surgery in patients with type 2 diabetes in three European countries(II)	0	0	0
Ara, R.	2007	The cost-effectiveness of sibutramine in non-diabetic obese patients: evidence from four Western countries	1	1	0
Ara, R.	2012	What is the clinical effectiveness and cost-effectiveness of using drugs in treating obese patients in primary care? A systematic review	1	1	1
Au, N.	2013	The cost-effectiveness of shopping to a predetermined grocery list to reduce overweight and obesity	1	1	1
Avenell, A.	2004	Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement	0	1	0
Bemelmans, W.	2008	The costs, effects and cost-effectiveness of counteracting overweight on a population level. A scientific base for policy targets for the Dutch national plan for action	1	1	1
Benarroch-Gampel, J.	2012	Cost-effectiveness analysis of cholecystectomy during Roux-en-Y gastric bypass for morbid obesity	0	0	0
Borisenko, O.	2015	Bariatric Surgery can Lead to Net Cost Savings to Health Care Systems: Results from a Comprehensive European Decision Analytic Model	1	1	1
Brennan, A.	2006	Assessment of clinical and economic benefits of weight management with sibutramine in general practice in Germany	1	1	0
Brown, H. S., III	2007	The cost-effectiveness of a school-based overweight program	0	0	0
Burch, J.	2009	Rimonabant for the treatment of overweight and obese people	0	0	0
Campbell, J.	2010	Cost-effectiveness of laparoscopic gastric banding and bypass for morbid obesity	0	0	0
Caro, J. J.	2007	Cost effectiveness of rimonabant use in patients at increased cardiometabolic risk: estimates from a Markov model	1	1	1
Carter	2009	Assessing Cost-Effectiveness in Obesity (ACE-Obesity): an overview of the ACE approach, economic methods and cost results	1	1	1

Table 2-1. Continued

Author	Year	Title	CVD	T2D	Stroke
Castilla, I.	2014	Cost-utility analysis of gastric bypass for severely obese patients in Spain	1	1	1
Cecchini, M.	2010	Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness	1	1	1
Chang, S. H.	2011	Cost-effectiveness of bariatric surgery: should it be universally available?	0	0	0
Clegg	2003	Clinical and cost effectiveness of surgery for morbid obesity: a systematic review and economic evaluation	0	1	0
Cobiac, L.	2010	Cost-effectiveness of Weight Watchers and the Lighten Up to a Healthy Lifestyle program	1	1	1
Craig, B. M.	2002	Cost-effectiveness of gastric bypass for severe obesity	0	0	0
Dalziel, K.	2007	Time to give nutrition interventions a higher profile: cost-effectiveness of 10 nutrition interventions	1	1	1
Faria, G. R.	2013	Gastric bypass is a cost-saving procedure: results from a comprehensive Markov model	1	1	0
Finkelstein,E. A.	2014	Meta- and cost-effectiveness analysis of commercial weight loss strategies	0	0	0
Forster, M.	2011	Cost-effectiveness of diet and exercise interventions to reduce overweight and obesity	1	1	1
Fuller, N. R.	2013	Cost-effectiveness of primary care referral to a commercial provider for weight loss treatment, relative to standard care-a modelled lifetime analysis	0	0	0
Fuller, N. R.	2014	Cost effectiveness of primary care referral to a commercial provider for weight loss treatment, relative to standard care: a modelled lifetime analysis	0	1	0
Galani, C.	2007	Modelling the lifetime costs and health effects of lifestyle intervention in the prevention and treatment of obesity in Switzerland	1	1	1
Galani, C.	2008	Uncertainty in decision-making: value of additional information in the cost-effectiveness of lifestyle intervention in overweight and obese people	1	1	1
Ginsberg, G. M.	2012	Economic effects of interventions to reduce obesity in Israel	0	0	0
Gustafson, A.	2009	Cost-effectiveness of a behavioral weight loss intervention for low-income women: The Weight-Wise Program	1	0	0
Haby, M. M.	2006	A new approach to assessing the health benefit from obesity interventions in children and adolescents: the assessing cost-effectiveness in obesity project	1	1	1
Hampp, C.	2008	Cost-utility analysis of rimonabant in the treatment of obesity	1	1	0

Table 2-1. Continued

Author	Year	Title	CVD	T2D	Stroke
Hertzman, P.	2005	The cost effectiveness of orlistat in a 1-year weight-management programme for treating overweight and obese patients in Sweden : a treatment responder approach	0	1	0
Hoerger	2010	Cost-Effectiveness of Bariatric Surgery for Severely Obese Adults With Diabetes	1	0	1
Hoerger, T. J.	2010	Cost-effectiveness of bariatric surgery for severely obese adults with diabetes	1	0	1
Hoerger, T. J.	2015	Medicare's intensive behavioral therapy for obesity: an exploratory cost-effectiveness analysis	1	0	1
Hollingworth, W.	2012	Economic evaluation of lifestyle interventions to treat overweight or obesity in children	1	1	1
HTA New Zealand	2010	COST EFFECTIVENESS REPORT OF PUBLIC HEALTH INTERVENTIONS TO PREVENT OBESITY	1	1	1
Iannazzo, S.	2008	Economic evaluation of treatment with orlistat in Italian obese patients	1	1	1
Ikramuddin, S.	2009	Cost-effectiveness of Roux-en-Y gastric bypass in type 2 diabetes patients	1	0	1
Jensen, C.	2005	The costs of nonsurgical and surgical weight loss interventions: is an ounce of prevention really worth a pound of cure?	0	0	0
Johannesson, M.	1992	A health-economic comparison of diet and drug treatment in obese men with mild hypertension	1	0	0
Kahn, R.	2008	The impact of prevention on reducing the burden of cardiovascular disease	1	1	1
Keating, C. L.	2009	Cost-effectiveness of surgically induced weight loss for the management of type 2 diabetes: modeled lifetime analysis	0	0	0
Konchak, C.	2012	Incorporating social network effects into cost-effectiveness analysis: a methodological contribution with application to obesity prevention	0	0	0
Lacey, L. A.	2005	Cost-effectiveness of orlistat for the treatment of overweight and obese patients in Ireland	0	1	0
Lamotte, M.	2002	A health economic model to assess the long-term effects and cost-effectiveness of orlistat in obese type 2 diabetic patients	1	0	1
Lee, Y. Y.	2013	The cost-effectiveness of laparoscopic adjustable gastric banding in the morbidly obese adult population of Australia	1	1	1
Lewis, L.	2014	The cost-effectiveness of the LighterLife weight management programme as an intervention for obesity in England	1	1	0
Maetzel, A.	2003	Economic evaluation of orlistat in overweight and obese patients with type 2 diabetes mellitus	1	0	1
Magnus, A.	2009	The cost-effectiveness of removing television advertising of high-fat and/or high-sugar food and beverages to Australian children	0	0	0

Table 2-1. Continued

Author	Year	Title	CVD	T2D	Stroke
Maklin, S.	2011	Cost-utility of bariatric surgery for morbid obesity in Finland	0	0	0
McEwan	2010	The Cost, Quality of Life Impact, and Cost–Utility of Bariatric Surgery in a Managed Care Population	0	0	0
McPherson	2007	Tackling Obesities: Future Choices – Modelling Future Trends in Obesity & Their Impact on Health	1	1	1
Meads, D. M.	2014	The cost-effectiveness of primary care referral to a UK commercial weight loss programme	1	1	1
Miners, A.	2012	An economic evaluation of adaptive e-learning devices to promote weight loss via dietary change for people with obesity	1	1	0
Moodie, M.	2008	Cost-effectiveness of a family-based GP-mediated intervention targeting overweight and moderately obese children	1	1	1
Moodie, M.	2009	Cost-effectiveness of active transport for primary school children - Walking School Bus program	1	1	1
Moodie, M. L.	2010	The cost-effectiveness of Australia's Active After-School Communities program	1	1	1
Moodie, M.	2011	Assessing cost-effectiveness in obesity: active transport program for primary school children--TravelSMART Schools Curriculum program	1	1	1
Moodie, M. L.	2013	The cost-effectiveness of a successful community-based obesity prevention program The Be Active Eat Well Program	1	1	1
Olsen, J.	2005	Cost-effectiveness of nutritional counseling for obese patients and patients at risk of ischemic heart disease	1	0	0
Persson, U.	2010	A case study of ex ante, value-based price and reimbursement decision-making: TLV and rimonabant in Sweden	1	1	1
Picot, J.	2009	The clinical effectiveness and cost-effectiveness of bariatric (weight loss) surgery for obesity: a systematic review and economic evaluation	1	1	1
Picot, J.	2012	Weight loss surgery for mild to moderate obesity: a systematic review and economic evaluation	1	1	1
Pil, L.	2014	Establishing a method to estimate the cost-effectiveness of a kindergarten-based, family-involved intervention to prevent obesity in early childhood. The ToyBox-study	1	1	1
Pollock, R. F.	2013	Evaluating the cost-effectiveness of laparoscopic adjustable gastric banding versus standard medical management in obese patients with type 2 diabetes in the UK	1	0	1
Roux, L.	2006	Economic evaluation of weight loss interventions in overweight and obese women	1	1	0

Table 2-1. Continued

Author	Year	Title	CVD	T2D	Stroke
Ruof, J.	2005	Orlistat in responding obese type 2 diabetic patients: meta-analysis findings and cost-effectiveness as rationales for reimbursement in Sweden and Switzerland	1	0	1
Rush, E.	2014	Lifetime cost effectiveness of a through-school nutrition and physical programme: Project Energize	1	1	1
Sacks, G.	2011	Traffic-light' nutrition labelling and 'junk-food' tax: a modelled comparison of cost-effectiveness for obesity prevention	1	1	1
Salem, L.	2008	Cost-effectiveness analysis of laparoscopic gastric bypass, adjustable gastric banding, and nonoperative weight loss interventions	0	0	0
Sassi	2009	Improving Lifestyles, Tackling Obesity: The Health and Economic Impact of Prevention Strategies	1	1	1
Siddiqui, A.	2006	A comparison of open and laparoscopic Roux-en-Y gastric bypass surgery for morbid and super obesity: a decision-analysis model	0	0	0
Song, H. J.	2013	Bariatric surgery for the treatment of severely obese patients in South Korea--is it cost effective?	1	1	1
Spyra, A.	2014	[Cost-effectiveness of different programs for weight reduction in obese patients with diabetes]	1	0	1
Trueman, P.	2010	Long-term cost-effectiveness of weight management in primary care	1	1	0
van Baal, P. H.	2008	Cost-effectiveness of a low-calorie diet and orlistat for obese persons: modeling long-term health gains through prevention of obesity-related chronic diseases	1	1	1
Veerman, J. L.	2011	Cost-effectiveness of pharmacotherapy to reduce obesity	1	1	1
Verhaeghe, N.	2014	Cost-effectiveness of health promotion targeting physical activity and healthy eating in mental health care	1	1	1
Wang, B. C.	2014	Cost-effectiveness of bariatric surgical procedures for the treatment of severe obesity	0	0	0
Wang, L. Y.	2003	Economic analysis of a school-based obesity prevention program	0	0	0
Warren, E.	2004	Cost-effectiveness of sibutramine in the treatment of obesity	1	1	0
Wilson, K. J.	2014	Cost-effectiveness of a community-based weight control intervention targeting a low-socioeconomic-status Mexican-origin population	1	1	1

CVD=Cardiovascular Disease; T2D=Type 2 Diabetes; 0=event not simulated; 1=event simulated

CHAPTER 3

Event Simulation and External Validation applied in published Health Economic Models for Obesity: a Systematic Review

Chapter 3 was informed by Schwander B., Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert Review of Pharmacoeconomics & Outcomes Research, 2018, DOI: 10.1080/14737167.2018.1501680; https://doi.org/10.1080/14737167.2018.1501680

3.1 Abstract

Introduction: This study aims to determine methodological variations in the event simulation approaches of published health economic decision models, in the field of obesity, and to investigate whether their predictiveness and validity were investigated via external event validation techniques, which investigate how well the model reproduces reality.

Areas covered: A systematic review identified a total of 87 relevant papers, of which 72 that simulated obesity-associated events were included. Most frequently simulated events were coronary heart disease (≈83%), type 2 diabetes (≈74%), and stroke (≈66%). Only for ten published model-based health economic assessments in obesity an external event validation was performed (14%; 10 of 72), and only for one the predictiveness and validity of the event simulation was investigated in a cohort of obese subjects.

Conclusions: We identified a wide range of obesity related event simulation approaches. Published obesity models lack information on the predictive quality and validity of the applied event simulation approaches. Further work on comparing and validating these event simulation approaches is required to investigate their predictiveness and validity, which will offer guidance future modelling in the field of obesity.

3.2 Introduction

Obesity is a multifactorial, chronic disorder that has, according to the World Health Organization (WHO), reached epidemic proportions globally and is a major contributor to the global burden of chronic disease and disability [1]. Obesity is defined as abnormal or excessive fat accumulation that may impact health. The body mass index (BMI) is a simple index of weight-for-height that is commonly used to classify obesity in adults. It is defined as a person's weight in kilograms divided by the square of his height in meters (kg/m2). According to the WHO definition, a BMI ≥ 25 and < 30 is overweight; a BMI ≥30 is obesity [1]. In 2014, worldwide, more than 1.9 billion adults (≈39%), 18 years and older, were overweight or obese. Of these, over 600 million adults (≈13%) were obese [2]. Overweight and obesity are leading risks for global deaths. In 2010, worldwide, it has been estimated that around 3.4 million adults died (≈6% of total deaths per year) as a result of being overweight or obese [3]. In addition, 44% of diabetes cases, 23% of coronary heart disease cases and 7% to 41% of certain cancer cases are attributable to overweight and obesity [4]. A recently published systematic review has determined that this clinical burden of obesity is associated also with a substantial economic burden, and that there is an urgent need for public health measures in order to save societal resources [5].

Given this clinical and economic burden, it is of major interest for healthcare decision makers to identify effective and cost-effective programs or interventions for obesity prevention and therapy. However, assessing the long-term clinical and economic impact of such programs or interventions on obesity is difficult, as associated risk factors (e.g. high blood pressure, hyperlipidemia, etc.) and diseases (e.g. type 2 diabetes, cardiovascular events, etc.) develop over a long period of time, requiring, ideally, long-term observational studies. Consequently, decision analytic modelling is particularly relevant for obesity due to the chronic nature of the obesity-associated risk factors, morbidities and related mortality; this requires long-term observations that are often not provided by purely empirical evaluations with only a limited follow-up period - for example, the randomized clinical trial. Hence several decision analytic models have been used to inform medical decision making in the context of obesity.

Previously, it was shown that there are huge variations in the modelling approaches focusing on the prevention and therapy of obesity, making it difficult for researchers and modelers to select an appropriate approach when designing a model, and subsequently for policy makers and physicians to assess the quality of an applied model, intended to inform political or medical decision making; this highlighted the need for ongoing in-depth research on this matter [6]. The core of each decision model is the clinical model structure; accordingly, the details on the specific event simulation approaches are of fundamental influence, as these have

a central impact on all clinical, economic and quality of life outcomes simulated by a decision model. According to an up to date modelling guideline the models' accuracy of making relevant predictions should be investigated by performing specific validation procedures [7]. In our study we focused on the external event validation procedures, that determine how good the modelling results do compare to external populations (e.g. long-term studies and/or real-world observations), as those are of major interest for investigating the predictiveness and validity of the event simulation approach.

Accordingly, the objective of our research was to determine and describe the methodological variations in the event simulation approaches and the related external validations of published health economic decision models in the context of obesity. We set the focus on coronary heart disease (CHD), type 2 diabetes (T2D) and stroke. The rationale for this selection was that we have previously found that these events are most frequently simulated by published decision models [6], and that cohort studies have demonstrated that these diseases are three of the most important consequences of obesity [8].

3.3 Methods

Following the PRISMA guidelines [9], we performed a systematic review using the Pubmed Database and the NHS Economic Evaluation Database (which includes MEDLINE, EMBASE, CINAHL, PsycINFO and PubMed) to identify full health economic assessments in the context of obesity that were published before the end of May 2015. In order to identify relevant publications, we performed and combined three different searches: one for health economic evaluations, one for decision models and one for obesity. The only restrictions which we applied were related to the publication type, as we excluded letters, editorials and historical articles. In addition, we manually checked the reference lists of retrieved papers and of relevant reviews [10,11,12,13,14,15,16,17,18,19,20,21] to identify additional relevant studies.

Eligible studies were original research articles on decision models for full health economic assessment in the context of obesity; we applied definitions from Drummond et al.[22] (health economic assessments), from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force [23] (health economic decision models), and the WHO criteria [1] (obesity) in order to define eligible studies.

Literature selection was performed in a two-step approach. First, all titles and abstracts were screened and rated according to the eligibility criteria. In all "unclear" cases, the full text article was ordered and all relevant information on the inclusion/

exclusion criteria was extracted, shared, reviewed by and discussed within the team of authors to reach a consensus decision. For more details on the literature search, the eligibility criteria and the literature selection we refer to our previous publication [6].

3.3.1 Data Extraction

For the data extraction we developed a predefined extraction form in order to summarize information on the obesity-associated events and simulation approaches. This included the following information: overview of obesity-associated events simulated; CHD/T2D/stroke incidence simulation approach, CHD/T2D/stroke simulation of the intervention effect and the event-specific mortality simulation. After the data were extracted we formed groups for the different incidence simulation approaches and for the different intervention effect simulation approaches identified, as those two approaches form the two fundamental parts of the obesity-associated event simulation investigated in this paper. These groups have been built by describing, counting and grouping the applied methodological differentiations for the simulation of each investigated obesity-associated key event (CHD, T2D, stroke). The various categories obtained are described below.

3.3.2 Categorization of event simulation approaches

We have categorized the incidence simulation approaches (base risk simulation approaches) into the following groups: a) established risk functions (e.g. Framingham, UKPDS or other risk functions; here population-specific risk factors are used to estimate the base risk), b) potential impact fraction (estimates the obesity-related incidence of an event in the investigated population), c) BMI-based incidence estimation (BMI-specific incidence rates) are used to estimate the base risk), d) BMI group-based incidence estimation (e.g. BMI<25, 25-30, >30 etc.; BMI group-specific incidence rates are used to estimate the base risk), e) age and gender-based incidence estimation (base risk is estimated on the basis of age and gender-specific incidence rates), and f) others (e.g. incidence based on waist circumference functions).

Looking at the intervention effect simulation approaches (influence of therapy or prevention approach on the base risk) we formed the following categories: a) effect on primary risk factors (in case of risk function-based incidence estimation; e.g. blood pressure change impacts the base risk for cardiovascular disease), b) BMI-related relative risk [RR] (base risk is changed by a BMI-specific RR), c) BMI group-related RR (base risk is changed by a BMI-specific RR), d) obesity-related RR (base risk is changed by an obesity status-related RR), e) change in BMI (in the case of BMI-based incidence estimation, a change in BMI has a direct impact on the base risk), f) change in BMI group (in the case of BMI group-based incidence,

a change in BMI has a direct impact on the base risk), and g) others (e.g. RR based on a % weight change).

A specific obesity-associated event simulation approach always consists of the combination of an incidence simulation approach and an intervention effect simulation approach. Accordingly, we investigated which combinations were most frequently applied to simulate a specific obesity-associated event, and presented the results in figure format (in order of frequency of application). Furthermore we grouped the different approaches in three major general event simulation methodologies, identified as: 1) established risk functions/equations (e.g. Framingham [24,25] or UKPDS [26], or the combination of both and others) were used to estimate the base risk of an event; in these cases the intervention effect was estimated by simulating the intervention's impact on the risk equation's risk factors (such as systolic blood pressure, age, diabetes status etc.); 2) the base risk of the events was estimated using different incidence estimation approaches (potential impact fraction, age & gender etc.) and a BMI or BMI group-specific relative risk (RR) was applied in order to estimate the intervention effect on the frequency of obesity-associated events; 3) the base risk was estimated on the basis of the BMI or a BMI group (BMI is the central part of the risk equation applied) and hence the intervention effects on the BMI or the BMI group were directly impacting the base risk. In general methodology 3, the BMI is used as a direct risk factor implemented in the incidence equation – hence the intervention effect on the BMI or the BMI group directly impacts the base risk (in contrast, method 2 applies a BMI-related RR).

3.3.3 Model Validation

For each included study, we extracted information on the external event validation approach, using the best practice recommendations of the report on "Model Transparency and Validation" issued by the ISPOR-SMDM Modeling Good Research Practices Task Force (ISPOR = International Society For Pharmacoeconomics and Outcomes Research; SMDM = Society for Medical Decision Making) [7]. This included information on the suitability of the external validation cohort (obesity cohort or other), on the systematic identification of suitable data sources, on the specification of dependence / independence of the used data sources (versus those used in the model simulations) and on the justification of the data source selection (due to predefined criteria). Furthermore, we extracted information on whether the external validation results were provided for each source separately, whether presentations of discrepancies (model vs. external validation) were provided, and whether quantitative measures on fit were provided.

3.4 Results

3.4.1 Literature Search

In total we identified 4,293 studies via the database searches, and we reviewed 4,304 abstracts (resulting from the database search plus n=11 hand search). From these we selected 142 articles for full-text review; of these, 87 papers met our inclusion criteria. We excluded 55 full text articles for the following reasons: no decision model (n=20), no full health economic assessment (n=17), review article (n=14), not about obesity (n=2), comment or protocol (n=2). Furthermore, we have identified models that have not simulated obesity associated events (n = 15)], so finally 72 papers were selected that simulated obesity associated events.

The flow chart of study selection is shown in Figure 3-1.

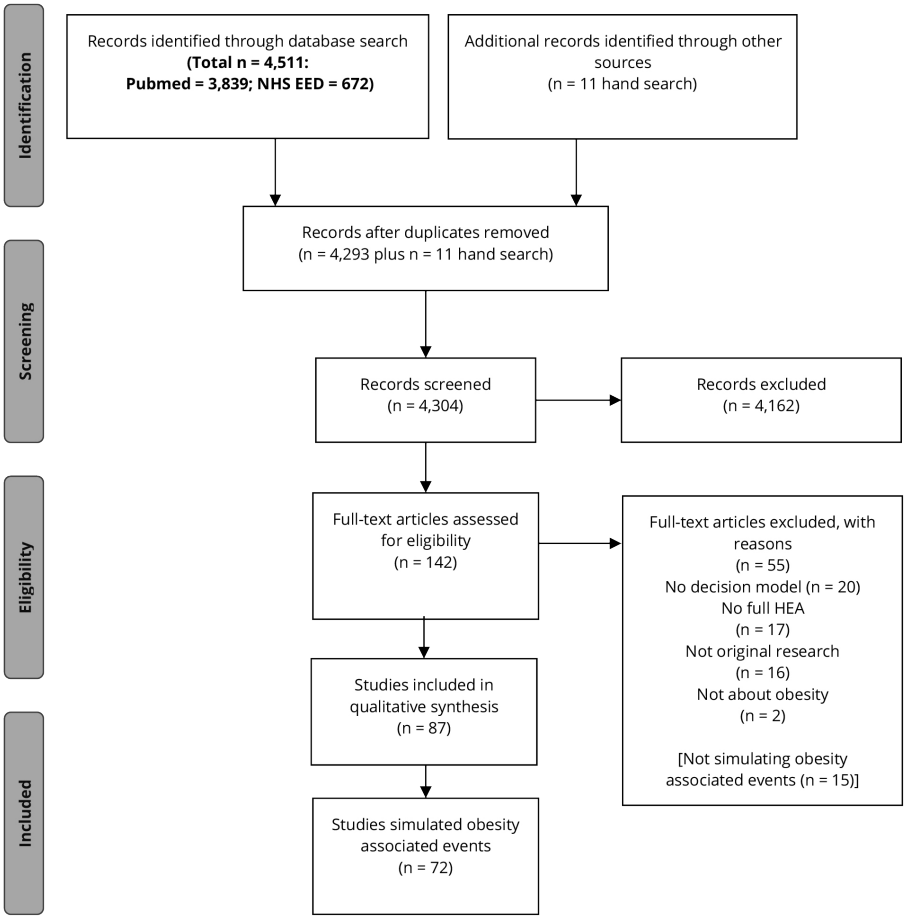


Figure 3-1. Flow Diagram of the systematic review process
NHS EED=National Health Service Economic Evaluation Database; HEA=Health Economic Assessment

3.4.2 Obesity-associated Event Simulation

Most but not all included models simulated obesity-associated events; of the 87 decision models identified, 72 simulated obesity-associated events, and in the other models the change in BMI was directly transferred into costs and effects (e.g. expressed as quality-adjusted life years).

As shown in Figure 3-2, most of these models simulated coronary heart disease (≈83%; 60 of 72), type 2 diabetes (≈74%; 53 of 72), and stroke (≈67%; 48 of 72). A minority of the models simulated cancer (≈35%), osteoarthritis (≈24%), hyperlipidemia (≈11%), hypertension (≈11%), and peripheral arterial disease (≈10%). The majority of models simulate more than one event: ≈36% simulate five or more events, ≈25% simulate four events, ≈17% three events, ≈10% two events and ≈12% only one event.

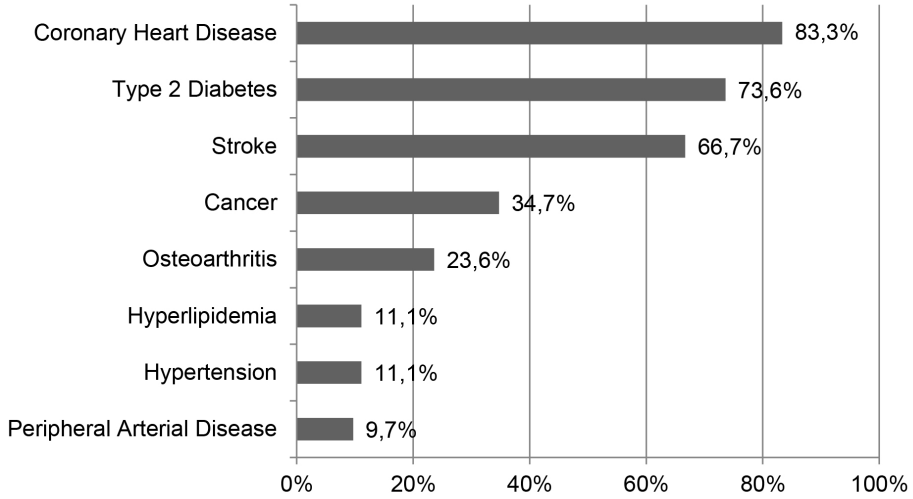


Figure 3-2. Proportion of decision models simulating specific obesity-associated events
The percentages presented above are calculated on the basis of the 72 decision models that simulate obesity-associated events; the 15 remaining decision models that were excluded simulated no events.

We have found that 65 of the identified 87 decision models simulated at least one obesity-associated key event (CHD, T2D and/or stroke). When looking at each single disease we obtained the following numbers: 60 models simulated CHD events; 53 models simulated T2D and 48 models simulated stroke. All three obesity-associated key events were simulated by 39 models.

3.4.3 Modelling approaches for coronary heart disease

With a combined frequency of more than 40%, risk equations were the most frequently applied general methodology for estimating the base risk for CHD (combined frequency of bar #1, bar #4, bar #8 and bar #9, shown in Figure 3-3).

In the second most applied general methodology (≈37% of approaches; combined frequency of bar #2, bar #3 and bar #7), the base risk was estimated using various incidence estimation approaches (potential impact fraction, age & gender etc.) and a BMI or BMI group-specific RR was applied in order to estimate the intervention effect on the frequency of CHD events.

In the third most applied general approach (≈18% of approaches; combined frequency of bar #5, and bar #6), the CHD incidence was estimated on the basis of the BMI or a BMI group and hence the intervention effect on the BMI or the BMI group was directly impacting the CHD base risk. Details on the frequency of each single approach identified for the CHD estimation are shown in Figure 3-3.

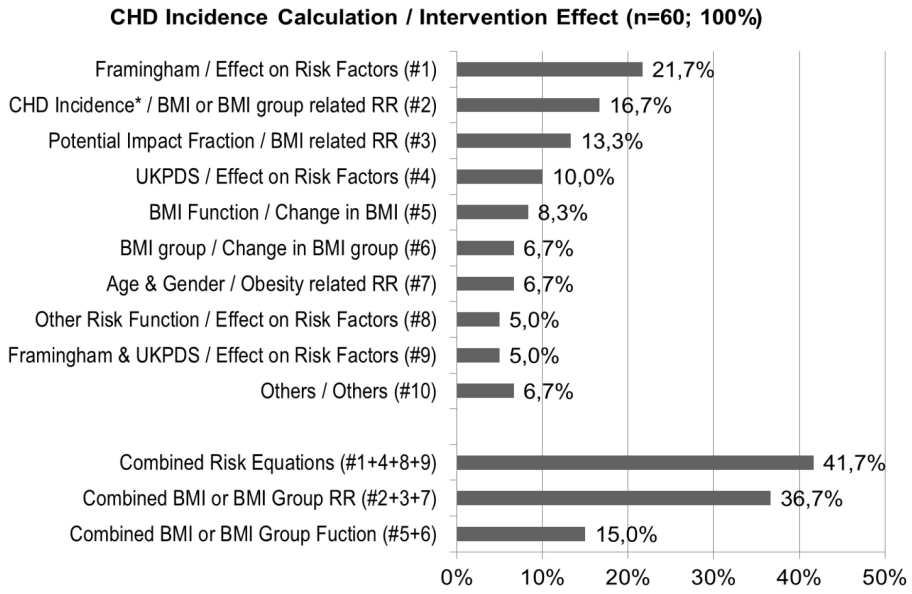


Figure 3-3. Overview of modelling approaches for coronary heart disease
*Incidence calculation based on different factors: CHD=Coronary Heart Diseases; BMI=Body Mass Index; RR=Relative Risk

3.4.4 Modelling approaches for type 2 diabetes

With a combined frequency of more than 40% (combined frequency of bar #2, bar #3 and bar #4, counted from above, shown in Figure 3-4) the T2D base risk was estimated using various incidence estimation approaches (potential impact

fraction, age & gender etc.) and a BMI-, BMI group- or obesity status-specific RR was applied in order to estimate the intervention effect on the frequency of T2D. As the second most applied general methodology (≈36% of approaches; combined frequency of bar #1, bar #5 and bar #7) the T2D incidence was estimated on the basis of a BMI or a BMI group function and hence the intervention effect on the BMI or the BMI group directly impacted the T2D base risk.

As the third most applied general approach (≈13% of approaches; combined frequency of bar #6 and bar #9), different risk equations were applied in order to estimate the base risk and the intervention effect on the basis of underlying risk factors. Details on the frequency of each single approach identified for the CVD estimation are shown in Figure 3-4.

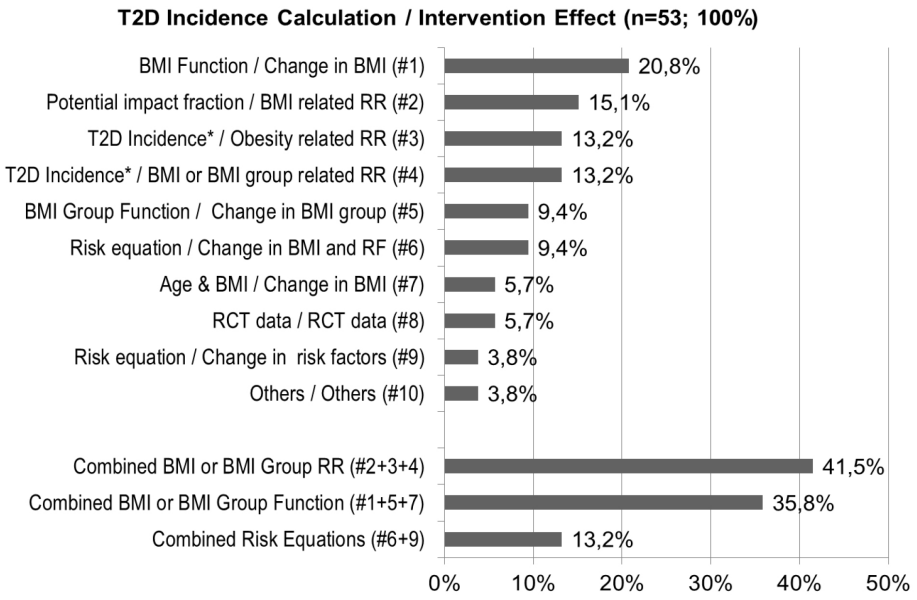


Figure 3-4. Overview of modelling approaches for Type 2 Diabetes
* Incidence calculation based on different factors: T2D = Type 2 Diabetes; BMI = Body Mass Index; RR = Relative Risk

3.4.5 Modelling approaches for stroke

The approaches for modelling stroke event simulation, shown in Figure 3-5, are largely comparable to the approaches identified for the CVD risk calculation. However, looking at the frequency, there are two general event simulation approaches that clearly dominate; these two share the top position with 41.7% each: namely, the risk equation-based general methodology (the combined frequency of bar #2, bar #4, bar #7 and bar #8, counted from above, shown in Figure 3-5)

and the application of (BMI-, BMI group- or obesity status-specific) relative risks on the base risk (the combined frequency of bar #1, bar #3 and bar #5). A direct connection between the BMI or BMI group and the base risk was applied only in 12.5% of cases (the combined frequency of bar #6 and bar #9).

Details on the frequency of each single approach identified for the stroke event estimation are shown in Figure 3-5.

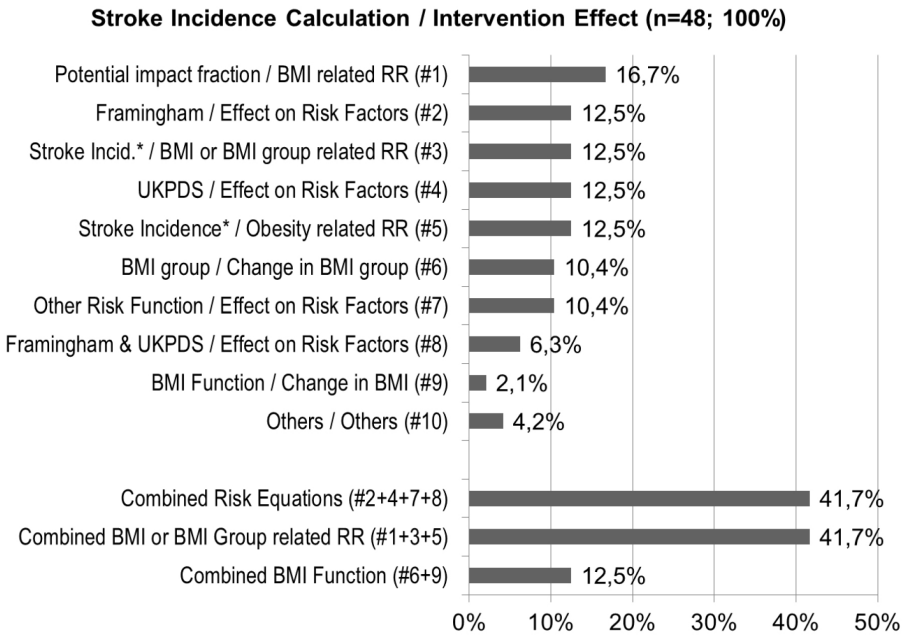


Figure 3-5. Overview of modelling approaches for stroke
* Incidence calculation based on different factors; CVD = Cardiovascular Diseases; BMI = Body Mass Index; RR = Relative Risk

3.4.6 Simulation of mortality

As mortality is one major consequence of obesity associated events and/or a consequence of their complications, we had a look on the mortality simulation related to the three key events CHD, T2D and/or stroke. We found that 91% (59 of 65) of the decision models that simulated CHD, T2D and/or stroke simulated any event-specific mortality (so 9% did not simulate either CHD, T2D or stroke mortality). Looking at the event-specific situation we found that 93% (56 of 60) of the models that simulated CHD events simulated CHD–specific mortality (so 7% did not simulate a CHD-specific mortality), 44% (23 of 53) of the models that simulated T2D simulated a T2D-specific mortality rate (so 56% did not simulate a T2D-specific

mortality), and 98% (47 of 48) of the models that simulated stroke simulated a stroke-specific mortality rate (so 2% did not simulate a stroke-specific mortality).

3.4.7 Focus only on type 2 diabetes patients

In some cases, type 2 diabetes models are used in order to investigate the health economic impact of interventions or prevention measures in the context of obesity. As a consequence, those models focus on an isolated cohort of T2D patients (non-diabetics are not considered). Such an approach was used in ≈17% (11 of 65) of the decision models that simulated CHD, T2D and/or stroke. In seven of those cases (≈64%; 7 of 11) sophisticated T2D models, namely the CDC-RTI Diabetes Model [27] (n=3) the Core Diabetes Model [28] (n=2), and the Archimedes Diabetes Model (n=2) [29] have been applied.

3.4.8 Model validation

Only ten models (≈15%; 10 of 65) that simulated CHD, T2D and/or stroke, performed an external validation procedure [30,31,32,33,34,35,36,37,38,39], and only for one of those models [39] the predictiveness of the event simulation was investigated in a cohort of obese subjects. All other models focused mainly on the external event validation in cohorts of type 2 diabetes patients; which is related to the fact that most of these external validation procedures were performed for the large diabetes models (7 of 10 models), namely the CDC-RTI Diabetes Model [27] (n=3), the Core Diabetes Model [28] (n=2), and the Archimedes Diabetes Model [29] (n=2). The events investigated by the external validation are varying from model to model but in most cases CHD (n=9), stroke (n=9) and mortality (n=8) were investigated; type 2 diabetes was only validated in one case. An overview of the ten model based health economic assessments in obesity that performed an external validation is shown in Table 3-1.

For the external validation of the large obesity models there are validation papers available (CDC-RTI Diabetes Model – Hoerger et al. 2009 [40]; Core Diabetes Model – Palmer et al. 2004 [41] and Archimedes Diabetes Model – Eddy et al. [42]) but in this context no obesity cohort was used as basis for the external validation (mainly diabetes cohorts and in some cases general populations). Excluding these large diabetes models, there were only three obesity models for which results of an external validation were provided, which are outlined in the following.

For a Markov model, developed by Caro et al. [30] for the UK healthcare setting, the authors used five different studies as validation basis, namely: WOSCOPS (West of Scotland Coronary Prevention Study Group)[43], ALLHAT (Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial)[44], TNT (Treating to New Targets)[45], CARDS (Collaborative Atorvastatin Diabetes Study)[46], and PPP (Primary Prevention Project)[47]. The underlying events simulation approaches

were "Framingham & UKPDS / Change in Risk Factors" for CHD and stroke, as well as a "Risk Equation / Change in Risk Factors" approach to simulate T2D. Results by events were not reported but the authors reported a poor fit comparing the modelling results with the outcomes of the TNT subsequent event study. The TNT study result had an R^2 value of 0.53. Excluding the TNT study, the R^2 value was reported as 0.80.

Table 3-1. Overview of model-based health economic assessments in obesity for that an external validation was performed

Author (Year)	Only T2D Cohort?	T2D	Stroke	CHD	Validation Paper†	Model Name	Intervention
Caro, J.J. (2007) [30]	No	1	1	1	no	NA	Drug Therapy
Kahn, R. (2008) [31]	No	1	1	1	yes	Archimedes Model	Behavioral Therapy
Ikramuddin, S. (2009) [32]	Yes	0	1	1	yes	Core Diabetes Model	Surgery
Hoerger, T.J. (2010a) [33]	yes	0	1	1	yes	CDC-RTI Diabetes Model	Surgery
Hoerger, T.J. (2010b) [34]	yes	0	1	1	yes	CDC-RTI Diabetes Model	Surgery
Pollock, R. F. (2013) [35]	yes	0	1	1	yes	Core Diabetes Model	Surgery
Castilla, I. (2014) [36]	no	1	1	1	no	NA	Surgery
Wilson, K.J. (2015) [37]	no	1	1	1	yes	Archimedes model	Behavioral Therapy
Hoerger, T.J. (2016) [38]	yes	0	1	1	yes	CDC-RTI Diabetes Model	Behavioral Therapy
Borisenko, O. (2015) [39]	no	1	1	1	yes (appendix)	NA	Surgery

† Additional / standalone publication focusing on the results of the external validation; NA = not available

For a discrete event simulation model, developed by Castilla et al. [36] for the Spanish healthcare setting, the authors used two studies as validation basis for these three key events, namely: the Di@bet.es Study [48 for T2D], the Framingham study [49 for CHD, 50 for Stroke]; hence only one study by event has been investigated. The underlying events simulation approaches were "BMI Group Function / Change in BMI Group" for CHD, T2D and stroke. In the external validation the lifetime risks obtained in the model were compared to those of the validation studies. For CHD this resulted in 33,6% (model) vs. 40,2% (validation study), for T2D in 24,2% vs. 23,2% and 25,2 vs. 19,0%; related goodness of fit estimates were not provided.

For a Markov model, developed by Borisenko et al. [39] for the Swedish healthcare setting, the authors used three studies as validation basis, namely ASCOT-BPLA (Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm) [51], AHEAD (Action for Health in Diabetes Study)[52] and ACCORD (Action to Control Cardiovascular Risk in Diabetes Study) [53], as well as the interventional quality registry SOREG (Scandinavian Obesity Surgery Registry) [54].The underlying events simulation approaches were "BMI Function / Change in BMI" for T3D and "Framingham / Change in Risk Factors" for CHD and stroke. According to the authors, the external validation showed that the model predicts the majority of clinical events with a high degree of precision, although there was a tendency to overestimate all-cause mortality and combined (fatal and non-fatal) myocardial infarction. Related goodness of fit estimates were not provided. Notably for this model the only pure obesity cohort was used as validation basis (Scandinavian Obesity Surgery Registry)[54], whereas all other cohorts used as basis for the external validation used pure diabetic or general populations; however this obesity cohort was only used to validate remission and incidence of T2D, as it was used to inform the all other endpoints simulated by the model, hich disqualifies using it for external validation.

For none of these three models a cross validation to other published obesity models was performed.

Considering the ISPOR best practice criteria, proposed by Eddy et al.2012 [7], we found that for none of these external validation cases a systematic identification of suitable data sources was performed, that for none of these external validation cases a specification of dependence / independence of the used data sources (versus those used in the model simulations) was performed, and that a justification of the data source selection, due to predefined criteria, was identified only in three [32,35,38] (of ten) cases. However, an adequate result presentation was provided for most external validation cases. In all cases the external validation results were provided for each source separately, presentations of discrepancies (model vs. external validation) were provided in nine cases, and quantitative measures on fit were provided in eight (of ten) cases. We found that only a limited number of published decision models for full HEAs in obesity have applied an external event validation. In addition, when external validation was conducted, there were major limitations including the data source selection process, as only in one case, obesity cohorts were used as basis for the validation procedure. An overview on the proportion of applied ISPOR best practice criteria for external validation is provided in Figure 3-6.

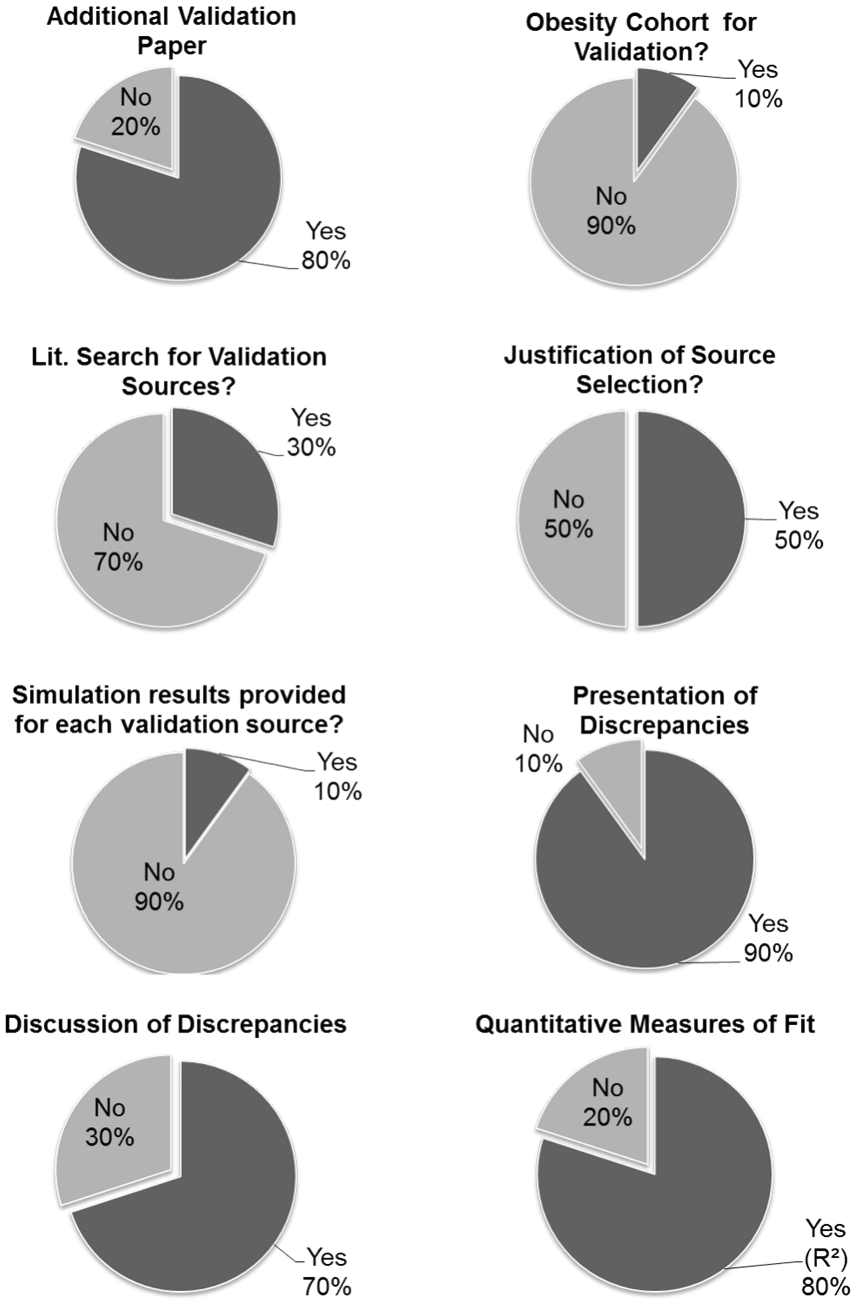


Figure 3-6. Overview of selected ISPOR best practice criteria for performing an external event validation

3.5 Discussion

As cohort studies have demonstrated that CHD, T2D and stroke are three of the most important consequences of obesity [8], we have focused our research on the investigation of these three events. Irrespective of the type of event we have identified three general event simulation approaches: 1) simulation is based on published risk equations, and the intervention effect is simulated by the change in equation-specific risk factors (key pros and cons: risk equations represent high clinical and structural validity but have high input data requirements that are not always available for the related decision problem and population of interest); 2) simulation is based on an event incidence estimate and the intervention effect is simulated by a BMI, BMI group or obesity status-specific relative risk (key pros and cons: represents a population-specific incidence estimate, but the relative risk approach, applied to the population base risk, might over- or underestimate the size of the intervention effect; consequently, it's difficult to apply or adapt this methodology to populations with a different base risk); 3) the incidence is estimated on the basis of the BMI or a BMI group and the intervention effect is simulated via a change in BMI/BMI group (this direct estimation of the event risk on the basis of the BMI requires valid population-specific data; accordingly, this approach might produce valid results only for the population the data is based on).

We found that published risk equations were more frequently applied in CHD and stroke but less commonly used in T2D. This might be based on the fact that with Framingham and UKPDS there are two large population-based studies that have produced widely accepted risk equations for CHD and stroke. For predicting T2D there are also some risk equations available (e.g. Stern et al. equation [55]) but none has acceptance comparable with Framingham or UKPDS, which might be why other approaches have been used more frequently.

Most of the publications identified (69 of 87) applied unique or distinctive modelling approaches; accordingly, we have identified a broad range of unique methodological frameworks. Currently it seems to be the case that each research team builds its own obesity model, and this is reflected in the diversity of obesity modelling approaches, although we have allowed potential double counting of republished modelling approaches (27 (of 87) publications were based on 9 unique / distinctive modelling approaches that have been applied and published several times) as those were not excluded from our review; when excluding these double counts the situation may even look more diverse. One key limitation of these models is the lack of published external validation results which could provide valuable information on the predictiveness, and hence on the quality, of their event simulation approaches. Only ten models performed an external validation and the predictiveness of the event simulation was investigated in a cohort of obese

subjects for only one. All others used diabetic or general populations as basis for the validation; consequently, no insights on the predictiveness of the applied event simulation approaches in obese subjects could be provided, which is a major limitation of all (but one) presented external validation results for published obesity models.

Accordingly, one limitation of our (current) findings is that the assessed frequency of use of the event simulation approaches does not provide sufficient insights into the quality of the event simulation approach. In order to investigate the quality of the different methodologies we need to perform some additional research steps. This means that it will be necessary to rebuild/reprogram the key event simulation approaches, feeding the rebuilt models with comparable patient population and intervention effect data and performing an external validation that compares the model-based event simulations to long-term epidemiological observations in the field of obesity.

The question in this context is whether it is worth including each approach in an in-depth investigation (reprogramming and validation) or whether there are specific qualitative aspects that might help to narrow down the number of different methodologies to be investigated in more detail. Important issues for this investigation are, for example, whether the model structure includes all aspects of reality that are considered important by clinical and health economic experts, and whether the model structure is consistent with medical science. For example, known cross-event dependencies could be a potential quality marker to identify the most valuable modelling approaches for the in-depth investigation. T2D is a known risk factor for CHD and stroke; accordingly, the model structure needs to take into account an increased risk of CHD and stroke in T2D patients. This structural quality aspect could be rated as fulfilled for models that use the Framingham algorithms, as T2D is already included as a risk factor within the related equation. A further aspect in this context could be whether the model structure accounts for an increased risk of subsequent CHD or stroke events; again the Framingham algorithms already account for this aspect. Looking at those two structural quality aspects, the approach of using published risk equations (such as the Framingham equations) have the advantage that they already account for key clinical aspects that need to be considered in the clinical model structure. In order to define such key structural quality criteria for the simulation of the-obesity associated events (CHD, T2D and stroke) an expert panel/advisory board is highly indicated and is planned as a future research step.

An additional topic to be discussed by this expert panel/advisory board is which data are the most valuable to be used for the external validation of the obesity-related event simulation. Here, in the best case, long-term epidemiological studies/

databases in obesity are required in order to compare the modelling results to real-world event data and to subsequently rate and rank the predictiveness of the modelling approaches.

As this review is focusing on the key structural aspects related to the clinical event simulation, other important aspects that may have a major impact on the modelling results, such as assumptions on the persistence of observed reductions in weight and or BMI, are not covered by this paper.

3.6 Conclusions

We have identified a wide range of health economic simulation approaches to model obesity-associated events, and published obesity models lack information on the predictive quality of the applied event simulation approaches. Therefore, further work on comparing and validating these event simulation approaches is required to investigate their predictiveness and validity, which will offer guidance on future modelling in the field of obesity.

3.7 References

1. World Health Organization (WHO) 2003. Fact Sheet on Obesity: Available from: http://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf. Accessed July 24, 2017.
2. WHO 2016. Updated Fact Sheet N 311 on Obesity and Overweight: Available from: http://www.who.int/mediacentre/factsheets/fs311/en/. Accessed July 24, 2017.
3. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384(9945):766-81.
4. WHO 2014. Updated Fact Sheet N 311 on Obesity and Overweight: Available from: http://www.wpro.who.int/mediacentre/factsheets/obesity/en/. Accessed July 24, 2017.
5. Tremmel M, Gerdtham UG, Nilsson PM, et al. Economic Burden of Obesity: A Systematic Literature Review. Int J Environ Res Public Health. 2017;14(4).
6. Schwander B, Hiligsmann M, Nuijten M, et al. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert Rev Pharmacoecon Outcomes Res. 2016;16(5):561-570.
7. Eddy DM, Hollingworth W, Caro JJ, et al. ISPOR–SMDM Modeling Good Research Practices Task Force. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value Health. 2012;15(6):843-50.
8. Field AE, Coakley EH, Must A, et al. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. Arch Intern Med. 2001;161(13):1581-6.
9. Moher D, Liberati A, Tetzlaff J, et al.for the PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009;339:332-339.
10. Boyers D, Avenell A, Stewart F, et al. A systematic review of the cost-effectiveness of non-surgical obesity interventions in men. Obes Res Clin Pract. 2015;9(4):310-27.
11. Flego A, Keating C, Moodie M. Cost-effectiveness of whole-of-community obesity prevention programs: an overview of the evidence. Expert Rev Pharmacoecon Outcomes Res. 2014;14(5):719-27.
12. Wang BC, Furnback W. Modelling the long-term outcomes of bariatric surgery: A review of cost-effectiveness studies. Best Pract Res Clin Gastroenterol. 2013;27(6):987-95.
13. Henteleff HJ, Birch DW, Hallowell PT; CAGS/ACS Evidence Based Reviews in Surgery Group. Cost-effectiveness of bariatric surgery for severely obese adults with diabetes. Can J Surg. 2013;56(5):353-5.
14. Griffiths UK, Anigbogu B, Nanchahal K. Economic evaluations of adult weight management interventions: a systematic literature review focusing on methods used for determining health impacts. Appl Health Econ Health Policy. 2012;10(3):145-62.
15. Gandjour A. Cost-effectiveness of preventing weight gain and obesity: what we know and what we need to know. Expert Rev Pharmacoecon Outcomes Res. 2012;12(3):297-305.
16. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. Nutrition. 2012;28(9):829-39.
17. Lehnert T, Sonntag D, Konnopka A, et al. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. Obes Rev. 2012;13(6):537-53.
18. Loveman E, Frampton GK, Shepherd J, et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess. 2011;15(2):1-182.

19. Levy DT, Mabry PL, Wang YC, et al. Simulation models of obesity: a review of the literature and implications for research and policy. Obes Rev. 2011;12(5):378-94.

20. Saha S, Gerdtham UG, Johansson P. Economic evaluation of lifestyle interventions for preventing diabetes and cardiovascular diseases. Int J Environ Res Public Health. 2010;7(8):3150-95.

21. Neovius M, Narbro K. Cost-effectiveness of pharmacological anti-obesity treatments: a systematic review. Int J Obes (Lond). 2008;32(12):1752-63.

22. Drummond MF, O'Brien B, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1997.

23. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. Value Health. 2003;6(1):9-17.

24. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.

25. D'Agostino RB, Russell MW, Huse DM, et al. Primary and subsequent coronary risk appraisal: new results from the Framingham study. Am Heart J. 2000;139(2 Pt 1):272-81.

26. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ. 2000;321(7258):405-12.

27. Hoerger TJ, Harris R, Hicks KA, et al. Screening for type 2 diabetes mellitus: a cost-effectiveness analysis. Ann Intern Med. 2004;140(9):689-99.

28. Palmer AJ, Roze S, Valentine WJ, et al. The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. Curr Med Res Opin. 2004;20 Suppl 1:S5-26.

29. Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. Diabetes Care. 2003 Nov;26(11):3093-101..

30. Caro JJ, Ozer Stillman IO, Danel A, et al. Cost effectiveness of rimonabant use in patients at increased cardiometabolic risk: estimates from a Markov model. JME. 2007;10(3):239-54.

31. Kahn R, Robertson RM, Smith R, et al. The impact of prevention on reducing the burden of cardiovascular disease. Circulation. 2008;118(5):576-85.

32. Ikramuddin S, Klingman D, Swan T, et al. Cost-effectiveness of Roux-en-Y gastric bypass in type 2 diabetes patients. Am J Manag Care. 2009;15(9):607-15.

33. Hoerger TJ, Segel JE, Zhang P, et al. Cost-effectiveness of bariatric surgery for persons with diabetes. Diabetes. 2010;59(suppl. 1): Congress of the American Diabetes Association. 70th Scientific Sessions: Abstract Number 201-OR.

34. Hoerger TJ, Zhang P, Segel JE, et al. Cost-effectiveness of bariatric surgery for severely obese adults with diabetes. Diabetes Care. 2010;33(9):1933-9.

35. Pollock RF, Muduma G, Valentine WJ. Evaluating the cost-effectiveness of laparoscopic adjustable gastric banding versus standard medical management in obese patients with type 2 diabetes in the UK. Diabetes Obes Metab. 2013;15(2):121-9.

36. Castilla I, Mar J, Valcárcel-Nazco C, et al. Cost-utility analysis of gastric bypass for severely obese patients in Spain. Obes Surg. 2014;24(12):2061-8.

37. Wilson KJ, Brown HS, Bastida E. Cost-effectiveness of a community-based weight control intervention targeting a low-socioeconomic-status Mexican-origin population. Health Promot Pract. 2015;16(1):101-8.

38. Hoerger TJ, Crouse WL, Zhuo X, et al. Medicare's intensive behavioral therapy for obesity: an exploratory cost-effectiveness analysis. Am J Prev Med. 2015;48(4):419-25.

39. Borisenko O, Adam D, Funch-Jensen P, et al. Bariatric Surgery can Lead to Net Cost Savings to Health Care Systems: Results from a Comprehensive European Decision Analytic Model. Obes Surg. 2015;25(9):1559-68.

40. Hoerger TJ, Segel JE, Zhang P and Sorensen SW. (2009). Validation of the CDC-RTI Diabetes Cost-Effectiveness Model. RTI Press publication No. MR-0013-0909. Research Triangle Park, NC: RTI International. Retrieved date. from http://www.rti.org/rtipress.

41. Palmer AJ, Roze S, Valentine WJ, et al. Validation of the CORE Diabetes Model against epidemiological and clinical studies. Curr Med Res Opin. 2004 Aug;20 Suppl 1:S27-40..

42. Eddy DM, Schlessinger L. Validation of the archimedes diabetes model. Diabetes Care. 2003 Nov;26(11):3102-10.

43. West of Scotland Coronary Prevention Study: identification of high-risk groups and comparison with other cardiovascular intervention trials. Lancet. 1996 Nov 16;348(9038):1339-42.

44. ALLHAT Officers and Coordinators. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002 Dec 18;288(23):2981-97.

45. Deedwania P, Barter P, Carmena R, et al. Reduction of low-density lipoprotein cholesterol in patients with coronary heart disease and metabolic syndrome: analysis of the Treating to New Targets study. Lancet. 2006 Sep 9;368(9539):919-28.

46. Colhoun HM, Betteridge DJ, Durrington PN, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet. 2004 Aug 21-27;364(9435):685-96.

47. Sacco M, Pellegrini F, Roncaglioni MC, et al. Primary prevention of cardiovascular events with low-dose aspirin and vitamin E in type 2 diabetic patients: results of the Primary Prevention Project (PPP) trial. Diabetes Care. 2003 Dec;26(12):3264-72.

48. Soriguer F, Goday A, Bosch-Comas A, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. Diabetologia. 2012 Jan;55(1):88-93.

49. Lloyd-Jones DM, Larson MG, Beiser A, et al. Lifetime risk of developing coronary heart disease. Lancet. 1999 Jan 9;353(9147):89-92.

50. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. Lancet Neurol. 2007 Dec;6(12):1106-14.

51. Dahlöf B, Sever PS, Poulter NR, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. Lancet. 2005 Sep 10-16;366(9489):895-906.

52. Look AHEAD Research Group. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. N Engl J Med. 2013 Jul 11;369(2):145-54.

53. ACCORD Study Group. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med. 2010 Apr 29;362(17):1575-85.

54. Hedenbro JL, Näslund E, Boman L, et al. Formation of the Scandinavian Obesity Surgery Registry, SOReg. Obes Surg. 2015 Oct;25(10):1893-900.

55. Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? Ann Intern Med. 2002;136(8):575-81.

CHAPTER 4

Identification and Expert Panel Rating of Key Structural Approaches applied in Health Economic Obesity Models

Chapter 4 was informed by Schwander B, Nuijten M, Hiligsmann M, Queally Q, Leidl R, Joore M; Oosterhoff M, Frew E, van Wilder P, Postma M, Evers S. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy Technol. 2020 Sep;9(3):314-322. https://doi.org/10.1016/j.hlpt.2020.03.005

4.1 Abstract

Background: Decision analytic modelling has increasingly been used to assess the long-term economic impact of obesity management measures. However, variability in quality and heterogeneity of underlying modelling methods could limit the use of these evaluations by decision makers. This study aims to assess the key structural modelling approaches applied in published obesity models, and to provide an expert consensus to improve the methodology and consistency of the application of decision-analytic modelling in obesity research.

Methods: Using a previously published systematic literature search as basis, ten individual interviews, and a face-to-face expert panel meeting were conducted. Within the expert panel meeting, the interview findings were presented and discussed, rated and where possible consensus statements were obtained. During the meeting, ten health economics experts assessed and made potential recommendations regarding the key structural approaches applied in published obesity models. In particular, five topics of interest were assessed: time horizon, model type, obesity-related clinical events simulated, event simulation approaches and external event validation.

Results: In addition to generic modelling standards, several obesity-specific recommendations were generated: Simulating a lifetime horizon was regarded as optimal (100% agreement); Ideally, both short and long-term results should be presented (100%); Using a risk equation approach for simulating the clinical events was the most preferred approach (60%) followed by applying a body mass index (BMI) related relative risk to a base risk estimate (30%); Continuous BMI approaches were preferred (relative to categorical ones) (100%); An individual patient/microsimulation state transition model was regarded as preferred modelling approach (90%); Discrete event simulation (DES) was regarded as the most flexible approach for building an obesity model but it was recognized as complex, and more difficult to build, populate and to disseminate (to stakeholders); Performing an external validation was rated as important (100%).

Conclusions: While the expert panel acknowledged some challenges and sometimes difficulties to achieve consensus, several recommendations for the key structural approaches for an economic obesity model were developed. The obtained insights, discussion and consensus can provide valuable guidance for developing decision-analytic models to generate high-quality and transparent economic evidence for obesity interventions that will be of use to decision makers.

4.2 Background

Obesity is a multifactorial, chronic disorder that has, according to the World Health Organization (WHO), reached epidemic proportions globally and is a major contributor to the global burden of chronic disease and disability [1]. Obesity is defined as abnormal or excessive fat accumulation that may impact health. The body mass index (BMI), defined as a person's weight in kilograms divided by the square of his height in meters (kg/m^2), is a simple index of weight-for-height commonly used to classify obesity in adults. According to the WHO definition, a BMI ≥ 25 and < 30 is overweight; a BMI ≥30 is obesity [1].

In 2016, worldwide, more than 1.9 billion adults (≈39%), 18 years and older, were overweight or obese. Of these, over 650 million adults (≈13%) were obese [2]. Overweight and obesity are leading risks for global deaths. In 2010, it has been estimated that around 3.4 million adults worldwide died (≈6% of total deaths per year) as a result of being overweight or obese [3]. In addition, 44% of diabetes cases, 23% of coronary heart disease cases and 7% to 41% of certain cancer cases were attributable to overweight and obesity [2].

A recently published systematic review, on the economic burden of obesity, has determined that this clinical burden of obesity is associated also with a substantial economic burden, and that there is an urgent need for public health measures in order to save societal resources [4]. Economic evaluations assess what the additional benefits of funding an intervention are relative to its additional costs [5]. Hence, results from economic evaluations allow decision makers to make an informed judgement on the economic impact of an intervention. To assess the long-term economic impact of prevention and therapy for obesity measures, decision analytic modelling has commonly been used [6, 7]. At the core of each decision model is the model structure [8]; accordingly, the key structural aspects of a decision model are of fundamental influence, as these have a central impact on all clinical, economic and health cost and outcomes simulated by a decision model.

Previously, it was shown that there are huge variations in the structural modelling approaches focusing on the prevention and therapy of obesity [6, 7] and up to now no consensus meeting on the structural aspects of obesity models has been performed. This makes it difficult for researchers to select an appropriate approach when designing a model, and subsequently for policy makers and stakeholders to assess the quality of an applied model, intended to inform political or medical decision making.

The aim of this study is therefore to assess and measure expert group consensus for key structural modelling approaches of obesity models, and to provide guidance to improve the methodology and consistency of applied models.

4.3 Methods

On the basis of a previously published systematic literature review [6, 7], the key structural approaches applied in published obesity models were identified.

In particular, five inter-related topics of interest were assessed: time horizon, model type, obesity-related clinical events simulated, event simulation approaches and external event validation. These features represent the structural aspects of models listed within the Phillips reporting checklist [9] which are not related on the quality of research reporting (as e.g. statement of the decision problem or statement of scope / perspective etc.). Additionally, these features showed a huge variation in published obesity models [6, 7].

The findings from the systematic literature review were then used to guide the topic content of the subsequent ten individual interviews. Data from the combined interviews were then presented and discussed at a face-to-face group meeting in order to derive consensus statements with respect to the key structural approaches applied in published obesity models.

4.3.1 Systematic Literature Search

The interviews and the group meeting were informed by a previously published systematic review [6, 7] that was performed in the PubMed Database and the NHS Economic Evaluation Database, following the PRISMA guidelines [10]. Three different searches were combined: one for health economic evaluations, one for decision models and one for obesity. Eligible studies were original research articles on decision models for full health economic assessments in the context of obesity; the definitions from Drummond et al. [11] (health economic assessments), from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force [12] (decision models), and from the WHO criteria [1] (obesity) were applied in order to define eligible studies. In total 4,293 studies were identified via the database searches, and were reviewed. From these 142 articles were selected for full-text review; of which 87 papers met the inclusion criteria. Of those, 72 models simulated obesity associated events. For more details on the literature search, the eligibility criteria and the literature selection please refer to chapters 2 & 3 of this book.

4.3.2 Individual interviews

Several health economic experts, with in-depth experience in decision analytic modelling and/or economics of obesity (using a convenience sampling), were requested to participate in an Expert panel meeting during the European Health Economic Association (EuHEA) conference 2018 in Maastricht, and ten (of twenty-two contacted) agreed to participate the meeting and to perform a 60-minute individual preparation interview beforehand. Within this interview the outcomes of the previously published systematic review, related to the key structural aspects (time horizon, model type, obesity-related clinical events simulated, event simulation approaches and external event validation) were presented via a web-based platform, and related to each of the key structural aspects of a model specific questions were asked.

With respect to the choice of a specific event simulation approach, different definitions were first obtained from the systematic review [7] The interview questions and the definitions of event simulation approaches are presented in box 4.1 and in box 4.2, respectively.

Box 4-1. Interview Questions

- Which time horizon would you rate as the minimum acceptable for a health economic obesity model?
- Which time horizon would you rate as optimal for a health economic obesity model?
- Which (obesity associated) events would you rate as the minimum acceptable to be included into a health economic obesity model?
- Which (obesity associated) events would you rate as optimal to be included into a health economic obesity model?
- Which model type would you prefer for a health economic obesity model?
- Why would you prefer this model type?
- Which event simulation approach would you prefer for a health economic obesity model? Please rank the top 3 approaches that you would prefer (1 = most preferred one to 3 = least preferred but still preferred one)
- Why would you prefer the top rated (#1) event simulation approach?
- Would you suggest to use different approaches for different events (consider coronary heart disease, type 2 diabetes, stroke)? If yes – why?
- How important do you rate an external validation for a health economic obesity model?

Box 4-2. Definitions of Event Simulation Approaches

- Risk Equation / Change in Risk Factors: E.g. Framingham / UKPDS equations – the base risk is calculated as an equation of risk factors and the intervention effect is simulated by the change of risk factors
- Disease Incidence Estimate / BMI related relative risk (RR): Any kind of incidence estimate (e.g. age-specific; gender-specific incidence etc.) is used as base risk and the intervention effect is simulated by applying a BMI related relative risk to the base risk
- BMI Function / Change in BMI: Base risk is calculated as function of the BMI which is directly influenced by the intervention effect on the BMI
- Disease Incidence Estimate / Obesity related RR: Any kind of incidence estimate (e.g. age-specific; gender-specific incidence etc.) is used as base risk and the intervention effect is simulated by applying an obesity status related relative risk (e.g. BMI <30 non-obese; BMI ≥30 obese) to the base risk
- BMI Group Function / Change in BMI Group: Base risk is calculated as function of specific BMI groups (e.g. < 25 normal weight; 25-30 overweight; 30-35 moderate obese; ≥ 35 severe obese; etc.) which is directly influenced by the intervention effect on the BMI group
- Disease Incidence Estimate / BMI Group related RR: Any kind of incidence estimate (e.g. age-specific; gender-specific etc.) is used as base risk and the intervention effect is simulated is simulated by applying a BMI group related relative risk to the base risk

The individual interview data were then analyzed quantitatively in MS Excel and summarized in a MS PPT presentation in order to serve as basis for the discussions at the expert panel meeting.

4.3.3 Expert Panel Meeting

The face-to-face expert panel meeting was performed as satellite event of the EuHEA conference in Maastricht, on July 13th 2018. Within this meeting, the interview results relating to each question were presented and discussed, with the aim of reaching a group consensus or to capture the variance in opinion for each item. Within this meeting the key structural aspects, were discussed in detail with a specific focus on obesity-specific criteria. After the meeting the results were summarized and sent to the expert panel members for further comment and approval.

The results from this expert panel meeting are presented below, together with the results of the individual interviews and the key results from the systematic literature review.

4.4 Results

4.4.1 Time Horizon

Table 4-1 presents the outcomes linked to the choice of time horizon for all published models identified in the review, and for the expert group opinion.

In the expert panel meeting, it was agreed that a lifetime horizon is optimal for a health economic obesity model (100% agreement)) and it was further agreed that both short- and long-term results should be presented (100% agreement). Short-term / trial period simulations may indeed also be interesting for practitioners / physicians, and are less susceptible to assumptions such as the sustainability of the intervention effect size and the natural course / development of BMI over time, including potential weight-regain post intervention.

Table 4-1. Time Horizon – Systematic Literature Search and Expert Interview Outcomes

Time Horizon	Literature Review (n=87 models)	Expert Interviews (n=10 experts)	
		Minimum	Optimal
< 20 years	23%	20%	10%*
≥ 20 and < lifetime	14%	20%	10%*
Lifetime	63%	60%	100%*

* 2 experts provided 2 different answers: ≥ 20 years in adults / lifetime in younger subjects; ≥ 10 years / lifetime optimal

4.4.2 Obesity Associated Events

Table 4-2 illustrates the findings from the literature review with respect to obesity-associated events (based on the 72 studies that have simulated obesity-associated events) alongside the findings from the expert interviews. Most of the published models simulated coronary heart disease (CHD) (≈83%; 60 of 72), type 2 diabetes (T2D) (≈74%), and stroke (≈67%). A minority of the models simulated cancer (≈35%), osteoarthritis (≈24%), hyperlipidaemia (≈11%), hypertension (≈11%), and peripheral arterial disease (≈10%).

From the expert interviews, with regard to the question on the minimum acceptable events to be included in a health economic obesity model (presented in Table 4-2), in 50% of cases only CHD, T2D and stroke were named as "minimum acceptable events" in 20% of cases accompanied by cancer and in 10% accompanied by hypertension; whereas in two cases no definite answer was given due to the rationale that "in general those events with strongest association / causal relationship to obesity should be included". Related to the question on the events to be included in a health economic obesity model in the optimal world (presented in Figure 4-1) the picture was more diverse.

Table 4-2. Obesity Associated Events – Systematic Literature Search and Expert Interview Outcomes

Obesity Associated Events	Literature Review* Outcomes (n=72 models)	Expert Interviews Outcomes (n=10 experts) (Minimum acceptable events)*		
		ChD, T2D and, Stroke	ChD, T2D, Stroke and Cancer	ChD, T2D, Stroke, Cancer and HT
Coronary heart disease (ChD)	83%	50%*	20%*	10%*
Type 2 Diabetes (T2D)	74%			
Stroke	67%			
Cancer	35%			
Hypertension (HT)	11%			
Osteoarthritis	27%			
Hyperlipidaemia	11%			
Peripheral arterial disease	10%			

*no definite answer was provided by 2 experts (n=20%) - in general those events with strongest association / causal relationship to obesity should be included

In 40% of cases it was stated that all events with a clear association with obesity should be included. One expert stated that this clear association should be combined with the severity of event consequences. In 50% of cases, CHD, T2D and stroke were named (alone or in combination with other diseases), whereas by one expert no definite answer was given as it was claimed that it depends on the goal of the model and on the available evidence.

During the expert panel, several discussions around these obesity associated events took place (please refer to discussion part), but it was not possible to achieve consensus on the whole. However, finally there was general agreement that those events with a strong statistical association to obesity combined with a clear clinical causal relationship to obesity should be included in the optimal case.

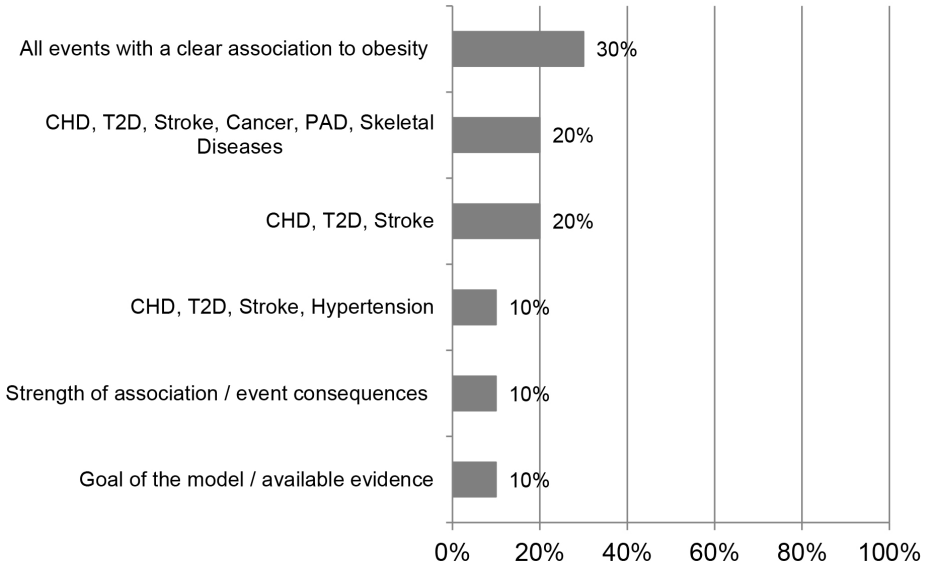


Figure 4-1. Which (obesity associated) events would you rate as optimal to be included into a health economic obesity model?

4.4.3 Model Type

Table 4-3 presents the results concerning the appropriate model type.

Table 4-3. Model Type – Systematic Literature Search and Expert Interview Outcomes

Model Type	Literature Review (n=87 models)	Expert Interviews (n=10 experts)
State Transition Model (STM)	85%	60%
Disease Event Simulation (DES)	2%	10%
Decision Tree Model	13%	--
STM or DES (expert rating)	--	30%

* 3 experts rated both STM and DES as suitable - depending on the data availability (for the DES model)

In the expert interviews, in 90% of cases a state transition model was named as the preferable approach, and, within these responses - 60% suggested a state transition model alone, and 30% also recommended a DES as an alternative model type to consider. Only one expert (10%) recommended DES alone.

On the question "why a specific model type was preferred?" the following rationales were provided by the experts:

- *"STM is adequate to simulate the three major health impacts (T2D, CHD and stroke);*
- *STM is most practicable for event-based simulation;*
- *STM is the most familiar approach (for health economists and stakeholders);*
- *STM is the most familiar approach - and individual patient simulation enables; building in specific memory;*
- *An individual patient simulation STM is preferred as it is possible to include a kind of memory".*

In three cases both the DES and the STM were preferred by the experts, for the following reasons:

- *"Memory is an important factor (as time with obesity / related morbidity impacts event risk) - therefore a DES would be preferred or a STM on a patient level with included memory states;*
- *Due to competing risks a DES / STM using a microsimulation approach will be preferred (for DES not all data might be available);*
- *DES might be scientifically the best approach but difficult to build, inform and to explain. STM might be the most accepted approach".*

For one participant the DES alone was preferred as

- *"DES allows considering timing of events which is important due to the inter-event dependencies".*

Within the expert panel, a consensus was reached in the form of the following two statements:

- An individual patient / microsimulation STM is regarded as preferred approach for an obesity model;
- DES is regarded as the most flexible approach however DES is complex, difficult to build, to inform and to explain (to stakeholders).

4.4.4 Event Simulation Approach

Within the expert interviews the experts were asked to rank a list of potential modelling approaches identified from the systematic review. The results are presented in Table 4-4 and in Figure 4-2, respectively. The risk equation approach was the most preferred approach (60% rated this as number one, followed by BMI-related RR (30% rated this as number one) and one expert felt it difficult to rank the approaches.

Table 4-4. Event Simulation Approach – Systematic Literature Search and Expert Interview Outcomes

Event Simulation Approach	Literature Review (n=72 models)	Expert Interviews (n=10 experts) – Ranking (#1, #2, #3)
Risk Equation / Change in Risk Factors	32%	#1 (60%): #2 (10%); #3 (20%)
Disease Incidence Estimate / BMI related relative risk (RR)	21%	#1 (30%): #2 (40%); #3 (0%)
BMI Function / Change in BMI	12%	#1 (0%): #2 (20%); #3 (20%)
Disease Incidence Estimate / Obesity related RR	12%	
BMI Group Function / Change in BMI Group	9%	
Disease Incidence Estimate / BMI Group related RR	7%	
Others / Others	7%	

* 3 experts rated both STM and DES as suitable - depending on the data availability (for the DES model)

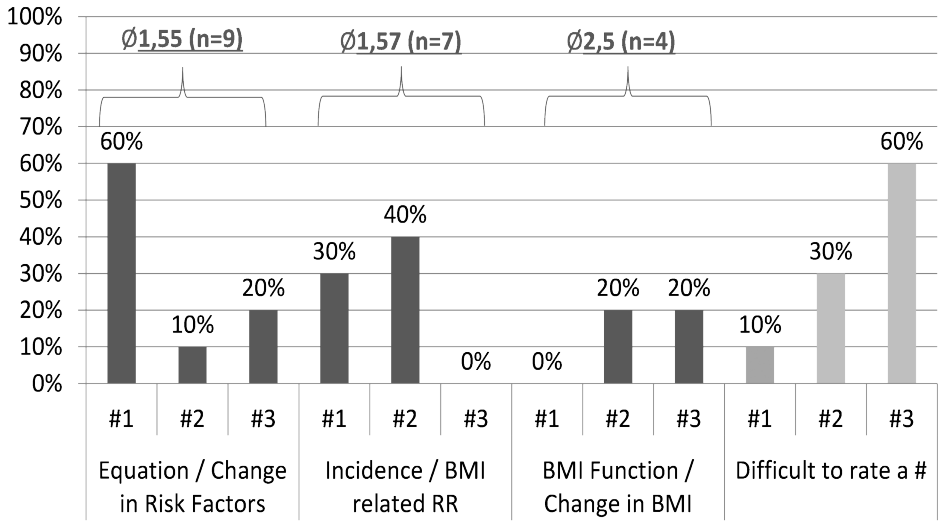


Figure 4-2. Outcomes of the interview question: *Which event simulation approach would you prefer for a health economic obesity model? (Rank 1-3)*

The reasons for the number one rating for the Equation / Change in Risk Factors were:

- *"Method is quite robust, widely validated and widely used;*
- *Quite valid (accepted) approach and most commonly used;*
- *Not everything might be explainable by change in BMI and therefore it may be important to consider further risk factors;*
- *Risk equation approach describes the whole nature of a chronic disease;*
- *Risk equation approach takes into account inter-event dependencies;*
- *Risk equation approach is widely applied and health economists are most familiar with this;*
- *Familiar approach, well know, risk equations are also used in clinical guidelines; for the others it is the key question how strong the association between BMI and risk is".*

The reasons for the number one rating for the Incidence / BMI related RR were:

- *"Most valuable / simple to set up events driven models for obesity;*
- *BMI related RR is preferred as always small changes are taken into account;*
- *Continuous BMI approaches are preferred against categorical approaches (there was 100% agreement on this statement in the expert panel)".*

Furthermore, in the interviews, the experts were asked whether they would suggest using different approaches for different events if considering CHD, T2D, and stroke. With regard to this question, 90% answered with "no"; whereas 40% mentioned that not necessarily different approaches need to be applied and 50% answered that consistent approaches (if applicable) are preferred. One expert found it difficult to rate this topic and gave no answer.

4.4.5 External Validation

External validation was defined as comparing the model's results with actual event data [13]. External validation involves simulating events that have occurred, such as those in clinical trials or epidemiologic studies, and examining how well the model results compare.

According to the systematic review, only ten published model-based health economic assessments in obesity included an external event validation (14%; 10 of 72).

Within the individual interviews the experts were asked how important they rate an external validation with possible answers being: "essential", "very important", "important", "less important", "not important" or "other" (please specify). All experts (100%) rated the external validation as "important"; 60% "very important" and 20% as "essential". These findings were confirmed during the expert panel.

4.4.6 Summary of Key Recommendations

A summary of key recommendations generated as a result of the expert interviews combined with the expert panel meeting are presented in Table 4-5.

Table 4-5. Overview of key expert recommendations by key structural aspect

Key Structural Aspect	Expert panel recommendations
Time Horizon	Simulating a lifetime horizon was regarded as optimal for an obesity model (100% agreement)
	Ideally, both short and long-term results should be presented (100% agreement)
Obesity Associated Events	No consensus was possible on which clinical events to be included in a health economic obesity model
	There was general alignment that those events with a strong association to obesity combined with a clear causal relationship to obesity should be included in the optimal case
Model Type	An individual patient/microsimulation state transition model was regarded as preferred modelling approach (90% agreement)
	Discrete event simulation (DES) was regarded as the most flexible approach for building an obesity model but DES was recognised as complex, as more difficult to build, populate and to disseminate (to stakeholders)
Event Simulation Approach	Using a risk equation approach for simulating the clinical events was the most preferred approach (60%) followed by applying a body mass index (BMI) related relative risk to a base risk estimate (30%)
	Continuous BMI approaches were preferred (relative to categorical ones) (100% agreement)
External Validation	100% of experts rated the external validation at least important

4.5 Discussion

Focusing on the key structural aspects outlined in the Philips checklist [9], this paper presents the main findings relevant to obesity models that have been identified (systematic literature search), rated (expert interviews) and discussed (expert panel). The expert panel meeting resulted in specific modelling recommendations that go beyond the findings from the systematic literature research, which is also representing the novelty of this research. The main findings by key structural aspect are discussed in detail below; each topic starts with a summary of outcomes of the expert panel meeting and these outcomes are then discussed and set into perspective by reflecting the complex circumstances and considerations related to each aspect. The latter discussion points are mainly driven by statements obtained during the expert panel meeting, which were accompanied and completed on the basis of related literature.

4.5.1 Time Horizon

With regard to the time horizon of a health economic obesity model, it was possible to obtain clear expert recommendations. However, there were some interesting viewpoints expressed during the expert panel mostly around the question of whether or not a short term (e.g. trial period) simulation should be performed and presented. One key consideration in this context was that practitioners, physicians and stakeholders might be (additionally) interested in short term results and it is recommended that health economists also consider the information needs of the health care personnel involved and also the requests / preferences of policy makers and other stakeholders. From a scientific point of view the key reasons for presenting short term / trial period outcomes (in addition to lifetime) were to present the impact of lifetime extrapolations as well as the practical need to determine whether the model adequately replicates the underlying study/trial results (internal validation). The key issues of extrapolation named in the context of obesity were the sustainability of the effect size (e.g. weight or BMI reduction and the related regain over time) and the natural course/development of weight / BMI over time, which is often based on a limited time-horizon, which again requires extrapolation to lifetime. These key issues of extrapolation were the key drivers for recommending an additional presentation of short term / trial period results.

4.5.2 Obesity Associated Events

The discussions around obesity-associated events to be modelled reflected some divergent views but there was general alignment among the experts that those events with a strong association to obesity combined with a clear causal relationship to obesity should be included in the optimal case. In contrast to the causal relationship of a specific event the strength of association could be more easily assessed, as the odds ratio or relative risk based on the best case could be extracted from prospective cohort studies. In a systematic review and meta-analysis of Guh et al. 2009 [14] the relative risk of various obesity associated events was presented and results by prospective cohort study and pooled results were provided, by gender and weight status (overweight / obese). According to the pooled results for obesity the strongest RR based associations in females (defined as RR≥2 in subjects with a BMI ≥ 30) were obtained for T2D (RR=12.41), CHD (RR=3.10), Gallbladder Disease (RR=3.08), Endometrial Cancer (RR=2.86), Kidney Cancer (RR=2.64), Hypertension (RR=2.42), osteoarthritis (RR=2.19) and congestive heart failure (RR=2.06) [14]. For males the strongest RR based associations (defined as RR≥2 in subjects with a BMI ≥ 30) were obtained for T2D (RR=6.74), osteoarthritis (RR=4.20), pancreatic cancer (RR=2.29) and asthma (RR= 2.19); the association to CHD in males (RR=1.75) was not that pronounced as in females (RR=3.10) [14]. Furthermore the association of obesity and stroke was not that pronounced with a RR of 1.50 in females and a RR of 1.68 in males [14]. Hence looking at the results of the systematic review (T2D, CHD and stroke are the most frequently included

events within health economic obesity models) it is clear that not only the strength of association is important but also the severity and consequences of the specific events need to be considered, which was also discussed and determined as a selection criteria during the expert panel meeting, and might explain the brought inclusion of CHD and stroke into the health economic obesity models, as both events are potentially leading to mortality or disability. Furthermore, from a health economic perspective the absolute incidence of events plays a role, as a strong obesity-association that is observed only in a very small number of patients, might have less impact on the cost-effectiveness than an event with a weak obesity-association that is observed in many patients.

The answer on the strength of statistical association, the severity and the absolute incidence of events are much easier to be answered than the question on the causal clinical relationship to obesity. The passage from obesity to T2D is caused by a progressive defect in insulin secretion coupled with a progressive rise in insulin resistance. Both insulin resistance and defective insulin secretion appear very prematurely in patients with obesity, and both worsen similarly towards diabetes [15], therefore the causal relationship is well understood. Also, there is good evidence on the causal relationship between obesity and CHD, and obesity and stroke and insulin resistance has been identified as the primary mechanism driving the progression of cardio-metabolic diseases (such as CHD and stroke) [16]. For different types of cancer the causal relationship is more challenging to capture and it remains unclear how obesity impacts the etiology of cancer, which itself is not fully understood [17]. Hence, many researchers might have not included cancer as an obesity associated event within the model. If including only those events, for which there is clear evidence of a causal relationship, T2D, CHD and stroke would be an adequate minimum selection to be simulated within a health economic model. In this context it is recommended that the inclusion of events for which the causal relationship to obesity is not yet fully understood is investigated within scenario analyses.

4.5.3 Model Type

The model types recommended for a health economic obesity model were either an individual patient / microsimulation STM or alternatively a DES. DES is clearly understood as the most flexible approach for building an obesity model, but it was also recognized as complex, as more difficult to build, populate and to disseminate (to stakeholders). Many shortcomings of (cohort) state transition models can be compensated by an individual patient / microsimulation approaches which enables patient history to be tracked using tunnel states and therefore overcome the Markovian assumption; this is important for obesity as time with obesity and/or obesity associated morbidities impacts the event risk. However, there is still some functionality of DES models that cannot be reproduced by a STM [18]. The

DES can simulate interactions amongst individuals or between individuals and the environment [19, 20], which might be interesting in obesity prevention models in which the positive effect of an intervention could have a positive effect on the whole community (e.g. on a whole school class or the whole school setting). Furthermore, DES is well suited to modelling situations where patients are subject to multiple or competing risks [20, 21]. A DES manages the competing and the sequencing of events by generating a future events list, then, for example, selecting the next closest time-to-event to ascertain which event occurs next in the process. This is relevant for obesity as there are several obesity associated events to be simulated. In a STM a transition probability is derived for each mutually exclusive competing health state and these competing health states must be exhaustive, and it requires many health states to achieve a level of detail comparable to DES. In a DES it is also easier to manage multiple events at the same time and to include and exclude events [22]. In the STM the patient is in one of a variety of mutually exclusive health states at any one time, which need to be clearly defined in the model structure, hence including / excluding events is a complex task. Furthermore, DES models can capture a greater level of detail than STM allowing the model to capture more detail regarding uncertainty in the system and including time to event information [20, 21]; this is important for obesity as multifactorial conditions and complex interventions (e.g. in the context of prevention) need to be simulated.

Besides all these advantages it needs to be considered that there are also several disadvantages, which prevent a broad application of DES in the fields of health economics [18]. DES models are generally more complex, require more data (that is often not available), and take more time to develop and run than STM; furthermore this could lead to a DES-induced over-specification [23] where models may become more complex than necessary, which again leads to increased data needs for DES models compared to STM [23].

These issues prevent a broad application of DES in health economics of obesity. The STM is rated as a pragmatic, widely applied, practical, familiar and widely accepted approach by the expert panel. Especially the communication and dissemination of (complex) DES models to stakeholders and policy makers is seen as a key hurdle for a broad application, as usually the model approach needs to be understandable to achieve research impact.

4.5.4 Clinical Event Simulation

The obtained event simulation approaches are quite diverse but it was possible to identify two preferred approaches by the expert panel namely the risk equation approach (most preferred approach - 60% rated this as number one, and the BMI related RR (30% rated this as number one). Many reasons were provided by the experts why the risk equation approach is preferred. The most prominent ones

were that the risk equation approach describes the whole nature of a chronic disease and considers inter-event dependencies whereas within the BMI based approach the question remains whether everything can be explained only by the BMI and how strong the BMI association of a specific disease really is. A further point that was highlighted in the expert discussions was that the modelers' decision on the event simulation approach is often driven by data availability. Whereas for the BMI based approach only data on the BMI development (over time) is required, the risk equation approach requires data on all risk factors included in the equation, and is therefore far more data demanding. In the case that data on the risk factors is not available the BMI approach could be the most pragmatic way to estimate the health economic impact of an intervention, although the named limitations need to be considered and extra sensitivity analyses and scenario analysis may be required. Furthermore the experts agreed on the procedure that (if possible) comparable event simulation approaches should be applied for the different events, mainly to have comparable strengths and limitations for the simulation of the different events included in the obesity model.

4.5.5 External Validation

The systematic review identified only ten models (of 72 that simulated events) that performed an external validation [7]. As this procedure is a key part of testing the validity of the modelling results with regard to the predictiveness of the event simulation approach, this was in general regarded as a limitation of published obesity models. All the experts rated the external validation as (at least) important for a health economic obesity model and that this should be performed as standard together with the internal validation that is usually performed as part of the internal model testing.

4.5.6 General Issues of Obesity Models

Besides the key structural aspects that were investigated and discussed there are several other aspects that make it a challenge to model health economic assessments in obesity. As already mentioned one key difficulty is that the chronic events associated with obesity require a lifetime horizon and therefore several assumptions related to the sustainability of the effect size and the natural course of weight / BMI. It is recommended that these two factors require clear and transparent handling and need to be investigated in a sensitivity analysis.

One other aspect that makes obesity models so diverse is that an intervention might focus either on the therapy or on the prevention of obesity. Whereas prevention measures usually start in younger age groups (e.g. in the school setting), the therapy of obesity could either target young or older age populations. Modelling prevention measures are usually more complex than modelling therapy, as the prevention effect might have a positive influence on the whole community setting,

and would hence require simulating interactions amongst individuals or between individuals and the environment, whereas therapy is usually targeted to the patients receiving a specific intervention.

Besides the diversity in the setting and intervention there are quite some challenges related to the understanding of the etiology of obesity and of obesity associated diseases including so called obesity-paradoxes [24]. Whereas obesity implies increased risk for chronic diseases, it is in fact associated with decreased mortality risk compared with normal weight [24]. Another paradox concerns the observation that when fitness is taken into account, the mortality risk associated with obesity is offset [24]. Furthermore there is a paradox describing the presence of a sizeable subset of individuals with obesity who are otherwise healthy [24]. Even when some obese persons are healthy and for late phase of disease, obesity may be protective, it still is considered an important risk factor in the development of chronic disease. This has been recently stressed in a review on cardiovascular diseases [25]. Modelling may thus have to distinguish several subgroups, depending on time and diseases analyzed.

4.5.7 Limitations and Implications

As discussed above, challenges around the economic modelling of obesity are not purely structural, and hence one limitation of this study is the focus only on key structural aspects. However, especially as there are many challenges, it is important to offer guidance on the handling of some key structural aspects when simulating obesity. The rationale for this is that the basic structure of the model is integral, and each decision that is made in the key structural development is carried forward to each calculation step of the model. Therefore, the provided consensus on those fundamental structural issues could minimize the challenges modelers, stakeholders and decision makes face, while developing, interpreting and rating model-based health economic assessments in obesity.

For the expert panel, we focused on experts that were attending the EuHEA meeting in Maastricht (2018), as a result of this selection criterion we had only European experts participating. Hence one limitation of this approach was that researches from non-European countries were not able to contribute to this research. Considering that, according to the previously published systematic review, 47% of decision models focused on a European setting, 27% on US setting and 20% on an Australian setting, it would have been interesting to consider additionally the expert opinion of non-European experts.

Further, in the expert interviews and in the expert panel we only used basic quantitative methods in order to obtain an expert rating and an expert consensus, as the style of questions were not designed to involve more advanced

quantitative methods (e.g. discrete choice experiments) or qualitative techniques (such as the Delphi method). Furthermore, the set focus on health economists is a limitation related to the composition of the panel. The rationale for selecting health economists was that modelling is primarily driven by this discipline, but as a consequence it was not possible to get a clear expert rating on purely clinical aspects, such as the obesity associated event selection. In case of specialized epidemiologists and / or clinicians the discussion might have moved more into the direction of which events are nowadays considered as clearly obesity associated, a fact that we have tried to resolve by discussing the latest related literature.

Although we have observed consensus on many structural issues, there is no structural approach that covers all needs, and hence related to the decision problem, research question, and according to the data and resource availability there are different structural approaches that were rated as suitable for building a health economic obesity model.

One key question that remains in this context is, how the application of different approaches to the same decision problem, research question and population might influence the results of the clinical event prediction and subsequently of the whole health economic evaluation – which is seen as a valuable field of future research.

4.6 Conclusions

While the working group acknowledges the challenges in achieving consensus, several recommendations for the key structural approaches for a health economic obesity model were developed. The obtained insights, discussion content and consensus can provide valuable guidance for all decision makers, health economists and modelers for developing decision-analytic models to generate high-quality and transparent economic evidence for obesity interventions.

4.7 References

1. World Health Organization. Fact Sheet on Obesity. 2003. [https://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf] (Accessed on 09.02.2019).
2. World Health Organization. Updated Fact Sheet on Obesity. 2018. [https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight] (Accessed on 09.02.2019).
3. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, Mullany EC, Biryukov S, Abbafati C, Abera SF *et al*: Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)* 2014, 384(9945):766-781.
4. Tremmel M, Gerdtham UG, Nilsson PM, Saha S: Economic Burden of Obesity: A Systematic Literature Review. *International journal of environmental research and public health* 2017, 14(4).
5. Ananthapavan J, Sacks G, Moodie M, Carter R: Economics of obesity--learning from the past to contribute to a better future. *International journal of environmental research and public health* 2014, 11(4):4007-4025.
6. Schwander B, Hiligsmann M, Nuijten M, Evers S: Systematic review and overview of health economic evaluation models in obesity prevention and therapy. *Expert review of pharmacoeconomics & outcomes research* 2016, 16(5):561-570.
7. Schwander B, Nuijten M, Hiligsmann M, Evers S: Event simulation and external validation applied in published health economic models for obesity: a systematic review. *Expert review of pharmacoeconomics & outcomes research* 2018, 18(5):529-541.
8. Afzali HH, Karnon J, Merlin T: Improving the accuracy and comparability of model-based economic evaluations of health technologies for reimbursement decisions: a methodological framework for the development of reference models. *Medical decision making : an international journal of the Society for Medical Decision Making* 2013, 33(3):325-332.
9. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S: Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006, 24(4):355-371.
10. Moher D, Liberati A, Tetzlaff J, Altman DG: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of clinical epidemiology* 2009, 62(10):1006-1012.
11. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW: Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford: Oxford University Press; 2015.
12. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR: Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2003, 6(1):9-17.
13. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB: Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2012, 15(6):843-850.
14. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH: The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC public health* 2009, 9:88.

15. Golay A, Ybarra J: Link between obesity and type 2 diabetes. *Best practice & research Clinical endocrinology & metabolism* 2005, 19(4):649-663.

16. Aslibekyan S, Garvey WT: Obesity: Obesity and cardiometabolic disease - more than meets the eye. *Nature reviews Endocrinology* 2017, 13(10):566-568.

17. Stone TW, McPherson M, Gail Darlington L: Obesity and Cancer: Existing and New Hypotheses for a Causal Connection. *EBioMedicine* 2018, 30:14-28.

18. Standfield L, Comans T, Scuffham P: Markov modeling and discrete event simulation in health care: a systematic comparison. *International journal of technology assessment in health care* 2014, 30(2):165-172.

19. Brennan A, Chick SE, Davies R: A taxonomy of model structures for economic evaluation of health technologies. *Health economics* 2006, 15(12):1295-1310.

20. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J: Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2012, 15(6):821-827.

21. Caro JJ, Briggs AH, Siebert U, Kuntz KM: Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2012, 15(6):796-803.

22. Caro JJ, Moller J, Getsios D: Discrete event simulation: the preferred technique for health economic evaluations? *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2010, 13(8):1056-1060.

23. Karnon J, Brown J: Selecting a decision model for economic evaluation: a case study and review. *Health care management science* 1998, 1(2):133-140.

24. McAuley PA, Blair SN: Obesity paradoxes. *Journal of sports sciences* 2011, 29(8):773-782.

25. Elagizi A, Kachur S, Lavie CJ, Carbone S, Pandey A, Ortega FB, Milani RV: An Overview and Update on Obesity and the Obesity Paradox in Cardiovascular Diseases. *Progress in cardiovascular diseases* 2018, 61(2):142-150.

CHAPTER 5

Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles & Reproduction Success

Chapter 5 was informed by Schwander B., Nuijten M., Evers S., Hiligsmann M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. PharmacoEconomics. 2021;39(4):433-46. https://doi.org/10.1007/s40273-021-01008-7

5.1 Abstract

Objectives: This research aims to: (i) replicate published health economic models; (ii) compare reproduced results with original results; (iii) identify facilitators and hurdles to model replicability and determine reproduction success; (IV) suggest model replication reporting standards to enhance model reproducibility, in the context of health economic obesity models.

Methods: Four health economic obesity models simulating an adult UK population were identified, selected for replication and evaluated using the CHEERS checklist. Reproduction results were compared to original results, focusing on cost-effectiveness outcomes, and the resulting reproduction success was assessed by published criteria. Replication facilitators and hurdles were identified and transferred into related reporting standards.

Results: All four case studies were state-transition models simulating costs and quality adjusted life years (QALYs). Comparing original vs. reproduction outcomes, the following range of deviations was observed; costs: -3.9% to 16.1% (mean over all model simulations: 3.78%), QALYs: -3.7% to 2.1% (mean: -0.11%), and average cost-utility ratios: -3.0 to 17.9% (mean 4.28%). Applying different published criteria, an overall reproduction success was observed for three of four models. Key replication facilitators were input data tables and model diagrams, while missing standard deviations and missing formulas for equations were considered as key hurdles.

Conclusions: This study confirms the feasibility of rebuilding health economic obesity models, but minor to major assumptions were needed to fill reporting gaps. Model replications can help to assess the quality of health economic model documentation and can be used to validate current model reporting practices; simple changes to actual CHEERS reporting criteria may solve identified replication hurdles.

5.2 Introduction

Method replicability and reproduction of results, which in other disciplines are common criteria of adequate research reporting to assure scientific rigor, are gaining importance in the field of health economic modelling, and have been the subject of recent studies [1, 2]. In the field of health economic modelling, the topics of research reporting, model transparency and model quality have been commonly discussed and investigated in great detail; this is reflected in the availability and application of multiple quality and reporting standards for health economic assessments [3-5]. A recently published review investigated the definitions of replicability in other disciplines, and produced a set of definitions for the success of result reproduction in health economic modelling [6]. This approach goes beyond the usual topics of reporting standards, transparency and quality. The issues of model replication and the reproduction of results have not yet been explored within health economic obesity decision models, which is especially relevant because obesity is a complex disease with several comorbidities. Consequently, complex modelling frameworks simulated over long-term horizons are required, and these carry the potential risk of errors by the modeler and/or misinterpretations by the reader. In order to investigate the reproducibility of results in this context we have selected health economic obesity models for replication, on the basis of a previously published systematic review [7, 8] and on the basis of previously published structural quality criteria for health economic obesity models [9]. The field of obesity modelling is in general very diverse; this is driven by multiple preventive and therapeutic approaches and multiple complications and comorbidities, which have triggered the development of various unique modeling approaches. Of 87 systematically identified obesity model publications, 69 (79% of the total) were based on unique modelling approaches [7], whereas in type 2 diabetes, of 78 systematically identified published models, only 20 (26% of the total) were based on unique modelling approaches [10]. This observed difference might also be based on the fact that there are currently no attempts to align, compare and validate obesity modelling approaches, such as the still ongoing Mount Hood challenge for type 2 diabetes [11-13]. Furthermore, it was found that most of these unique obesity models lack an external event validation [8], making the replication of obesity models specifically an interesting research exercise.

According to previously published research in the field of model replication, comprehensive replicability is generally perceived to be desirable in health economic models [1], but additional work is needed to understand how to improve model transparency and in turn increase the chances of successful result reproduction [2]. These existing publications state that further work is needed to better understand facilitators and hurdles, and to define standards that could ultimately increase the chances of replication. Accordingly, our research goes beyond currently

published approaches and investigates model replication and result reproduction in complex obesity models, with a special focus on a systematic assessment of results reproduction success and on identifying solutions for improving current reporting standards to enhance model replicability.

Therefore the objectives of our research were: (i) to replicate published health economics models in obesity; (ii) to compare the reproduction results to the original results; (iii) to determine facilitators, hurdles, and challenges of the replication process and to assess the reproduction success measured by different definitions suggested by McManus et al. 2019 [6]; (iv) and finally to suggest model replication reporting standards to enhance model reproducibility.

5.3 Data & Methods

5.3.1 Model Selection & Model Overview

Based on a previous systematic review identifying 87 health economic obesity models [7, 8], the models for replication were selected using an expert panel consensus [9]. The panel assessed the key structural modelling approaches applied in published obesity models, and provided an expert consensus to improve the methodology and consistency of the application of decision-analytic modelling in obesity research. In order to select high quality obesity models, the related minimal structural requirements for health economic obesity models were applied, consisting of the following criteria: (i) simulation time horizon: long-term (lifetime or comparable) [n=55 of 87]; (ii) model type: state transition model (STM) or discrete event simulation (DES) [STM n=74 or DES n=2 of 87]; and (iii) events simulated: at least coronary heart disease, type 2 diabetes and stroke [n=39 of 87]. To assure that the models were simulating a comparable setting and patient population, the United Kingdom (UK) country setting [n=15 of 87], and the adult population were used [n=70 of 87] as final model selection criteria, which resulted in four health economic obesity models [14-17]. Additional details of this step-wise model selection process are presented in the appendix [Table 5-6].

The details of these health economic models are presented in Table 5-1.

Table 5-1. Health Economic Obesity Models selected for the Replication Process

Coding	Ref.	Model Type	Time Horizon/ Cycle Length	Events	Intervention / Comparator	Country / Perspective	Health Outcomes(s)	Software
Case Study 1	Ara (2012) [14]	Cohort STM	Lifetime/ 1-year	MI, T2D, Stroke, Death	Diet & exercise advice plus: Orlistat, Sibutramine, Rimonabant; Placebo	UK / NHS	QALYs	Simul8
Case Study 2	Au (2013) [15]	Cohort STM	40 Years*/ 6 months	MI, T2D, Stroke, CC, OA Death	Standard behavioral therapy (SBT); SBT combined with meal plans & shopping lists (SBT+list); Do Nothing	UK / NHS	QALYs	Excel
Case Study 3	Caro (2007) [16]	Cohort STM	Lifetime/ 1 month	MI, Stroke, AP, TIA, T2D, Death	Diet & exercise advice plus: Rimonabant; Placebo	UK / NHS	QALYs	Excel
Case Study 4	Meads (2014) [17]	Cohort STM	Lifetime/ 1-year	MI, T2D, Stroke, Death	Commercial weight loss program; Usual Care	UK / NHS	QALYs	Excel

*regarded as comparable to lifetime; AP = Angina Pectoris; CC = Colorectal Cancer; MI = Myocardial Infarction; NHS = National Health Service; OA = Osteoarthritis; STM = State Transition Model; T2D = Type 2 Diabetes; TIA = Transient Ischemic Attack; UK = United Kingdom; QALYs = Quality adjusted life-years

The first obesity model (Case Study 1) is based on extensive research, informed by a systematic review, a mixed treatment comparison and a lifetime health economic Markov modelling approach, consisting of 13 health states. This research was funded by the UK National Institute for Health Research Health Technology Assessment (HTA) program and is presented in a full-length HTA report, including an appendix with the health economic model, published in Health Technology Assessment [14].

The second obesity model (Case Study 2) is based on a systematic review focusing on interventions based on food purchasing patterns, and used a long-term health economic Markov modelling approach consisting of 9 main health states. Although the model description in the original paper, published in Nutrition & Diabetes, is very brief, the publication is accompanied by an extensive appendix in which all relevant information on the modelling approach and on the underlying input data can be found [15].

The third obesity model (Case Study 3), funded by an industry research grant, is based on intervention-related clinical trials simulated over a lifetime horizon, using a health economic Markov modelling approach consisting of 5 main health states. All relevant information on the modelling approach and the input values was provided in the original paper, published in the Journal of Medical Economics [16]; of the four case studies, this was the only one that presented information on internal and external validation of the model.

The fourth obesity model (Case Study 4), funded by an industry research grant, uses an intervention-related clinical trial, a company dataset and a lifetime health economic Markov modelling approach based on 9 health states. All relevant information on the modelling approach and the input values was provided in the original paper, published in Clinical Obesity [17].

5.3.2 Replication of Health Economic Obesity Models

To prepare the replication of a specific model, a predefined data / information availability check was performed and the results were recorded in table format for each selected model. This initial check was supplemented by the documentation of all identified issues, hurdles and facilitators observed during the model replication (this process is described in appendix and the results of this two-step procedure are presented in the appendix tables Table 5-7 to Table 5-10). The replication was performed in TreeAge Pro Healthcare (Version 2020 - TreeAge Software, Inc.) by one modeler; this specialized modelling software was used in order to minimize potential programming errors as all relevant calculations are automated, once the model structure and inputs have been defined by the modeler. A summary of

identified replication facilitators and hurdles is provided in the result section below (Table 5-2); details for each model are provided in the appendix section 5.8.2.

5.3.3 Comparison of Reproduction Results to the Original Results

For each replicated model, model simulations were performed according to those presented in the original paper. The results were then compared to the original results, focusing on the health economic model outcomes, namely costs, clinical effects (especially QALYs) and cost-utility (as all models used QALYs as the effectiveness parameter). For each case study all published long-term comparisons were analyzed and the related costs, QALY and cost-effectiveness (CE) results were presented, as average CE ratios (for each alternative) and as incremental CE ratios (ICER). These health economic outcomes are presented together with the deviation of results between the replication and the original (absolute and as a percentage) in table format. In order to achieve a better rating of the deviation between original and reproduction results, the incremental costs and the incremental QALY results are visualized for all comparisons of the underlying case studies in the incremental cost-effectiveness coordinate plane.

5.3.4 Assessment of the Reproduction Success

A recently published systematic review, presented in 2019 by McManus et al., investigated published definitions for replicability in health economics and other disciplines and produced a set of potential definitions for result reproduction success in health economic models, based on definitions from other scientific disciplines [6]. These definitions are: (i) the same conclusions for the intervention's cost-effectiveness were reached; (ii) costs and outcomes replicated for some treatment pathways/model scenarios and not others; (iii) results for the costs and outcomes vary by only a specific percentage and are consistent with the original conclusions in comparison with the original; (iv) the calculated incremental cost-effectiveness ratio varies by only a specific percentage in comparison with the original; (v) cost-effectiveness figures could be reproduced to a reasonable degree of success (for example, the ICER plane or the cost-effectiveness acceptability curve); (vi) identical results are produced. The findings according to these success criteria are presented in table format for each case study. On the basis of these findings the different replication success criteria are interpreted and combined in order to allow a final overall assessment of the success of the model reproduction of results. For each case study all published long-term comparisons were analyzed and the related results of the reproduction success assessment are indicated by "yes" (assessment criteria is fulfilled) and by "no" (assessment criteria is not fulfilled). For all criteria that are investigating a relative variation, expressed as a percentage, we investigated thresholds of 5%, 10% and 20%, for the intervention, the comparator and the incremental results, in order to see how this might influence the rating of the reproduction success.

5.3.5 Assessment of Model Replication Reporting Standards

The selected case studies were appraised for quality of reporting using the CHEERS checklist [18]. One reviewer assessed the reporting quality of the included studies. The twenty-four items of the CHEERS checklist were scored using 'Yes' (reported in full), 'Part' (Partially reported), 'No' (not reported), and 'Not Applicable'. According to a previously published approach [19] a score of 1 was assigned if the requirement of reporting for a specific item was fulfilled completely, 0.5 for partial fulfillment and otherwise 0; resulting in a maximum score of 24 for an article that reported all information completely.

On the basis of the assessed quality of reporting, how successful the reproduction of results is, and the identified facilitators and hurdles, specific model replication reporting standards are suggested. The detailed health economic model reporting recommendations, provided in the CHEERS statement, are then used as the basis for evaluating whether and which changes of these existing reporting criteria would enhance the reproducibility of model results.

5.4 Results

5.4.1 Replication Process - Facilitators and Hurdles

It was possible to replicate all selected models but in all cases there were hurdles, which needed to be overcome by specific assumptions, which potentially influenced the reproduction of results. A summary of the key facilitators and the key hurdles, identified during the publication review and during the model replication process, is presented in Table 5-2.

5.4.2 Comparison of Reproduction Results to the Original Results

The reproduced results, the original results and the comparison of both results as absolute and as relative (presented as percentage) variation are presented in Table 5-3 for all four obesity case studies. In addition, the incremental CUA results are visualized as a CE coordinate plane (Figure 5-1), presenting the ICER as cost per QALY gained, for both the original model and the replication.

Table 5-2. Summary of Key Facilitators and Key Hurdles for Model Replication

Category	Detailed Description of Key Replication Facilitator / Hurdle	Applies to specific case study (yes / no)			
		#1	#2	#3	#4
Key Facilitators	Model structure and possible state transitions were presented in a state transition diagram.	yes	yes	yes	yes
	Overview of input parameters was provided in table format.	yes	yes	yes	yes
Key Hurdles	Probabilistic sensitivity analyses were performed	yes	no	yes	yes
	Relevant PSA values for PSA result reproduction were provided (type of distribution and either mean and standard deviation or distribution parameters were provided).	no	no	no	no
	Clinical event simulation results were provided (which are very helpful to guide potential assumptions to be made for rebuilding the model and which provide an additional means of testing the fit of the replication).	no	no	yes	yes
	Relevant details on the underlying life tables were provided (including year of data).	yes	yes	no	no
	Several self-created regression equations were introduced but without details on how to apply/solve the provided regressions correctly.	yes	no	no	no

Table 5-3. Cost, Utility and CU Results: Original versus Reproduced Results by Case Study

Model	Scenario	Definitions	Costs			QALYs			CU-Ratio		
			INT	CP	INC	INT	CP	INC	INT	CP	INC
Case Study 1	LT (per patient) **Orlistat vs. Placebo**	LT (original)	3,097	2,806	291	15.303	15.128	0.175	202	185	1,663
		LT (reproduction)	3,258	2,937	321	15.133	15.107	0.026	215	194	12,346
		Difference	161	131	30	-0.170	-0.021	-0.149	13	9	10,683
		Difference in %	**5.2%**	**4.7%**	**10.3%**	**-1.1%**	**-0.1%**	**-85.1%**	**6.4%**	**4.9%**	**642.4%**
	LT (per patient) **Rimonabant vs. Placebo**	LT (original)	3,478	2,806	672	15.317	15.128	0.189	227	185	3,556
		LT (reproduction)	3,689	2,937	752	15.157	15.107	0.050	243	194	15,040
		Difference	211	131	80	-0.160	-0.021	-0.139	16	9	11,484
		Difference in %	**6.1%**	**4.7%**	**11.9%**	**-1.0%**	**-0.1%**	**-73.5%**	**7.0%**	**4.9%**	**322.9%**
	LT (per patient) **Sibutramine 10 mg vs. Placebo**	LT (original)	3,011	2,806	205	15.376	15.128	0.248	196	185	827
		LT (reproduction)	3,448	2,937	511	15.161	15.107	0.054	227	194	9,463
		Difference	437	131	306	-0.215	-0.021	-0.194	31	9	8,636
		Difference in %	**14.5%**	**4.7%**	**149.3%**	**-1.4%**	**-0.1%**	**-78.2%**	**15.8%**	**4.9%**	**1044.3%**
	LT (per patient) **Sibutramine 15 mg vs. Placebo**	LT (original)	2,967	2,806	161	15.418	15.128	0.290	192	185	555
		LT (reproduction)	3,445	2,937	508	15.190	15.107	0.083	227	194	6,120
		Difference	478	131	347	-0.228	-0.021	-0.207	35	9	5,565
		Difference in %	**16.1%**	**4.7%**	**215.5%**	**-1.5%**	**-0.1%**	**-71.4%**	**18.2%**	**4.9%**	**1002.7%**
Case Study 2	40Y (per patient); **SBT + list vs. SBT**	40Y (original)	9,359	9,369	-10	27.180	27.070	0.110	344	346	-91
		40Y (reproduction)	9,255	9,276	-21	27.700	27.610	0.090	334	336	-233
		Difference	-104	-93	-11	0.520	0.540	-0.020	-10	-10	-142
		Difference in %	**-1.1%**	**-1.0%**	**110.0%**	**1.9%**	**2.0%**	**-18.2%**	**-2.9%**	**-2.9%**	**156.0%**
	40Y (per patient); **SBT + list vs. Do Nothing**	40Y (original)	9,359	9,302	57	27.180	26.840	0.340	344	347	168
		40Y (reproduction)	9,255	9,261	-6	27.700	27.390	0.310	334	338	-19
		Difference	-104	-41	-63	0.520	0.550	-0.030	-10	-9	-187
		Difference in %	**-1.1%**	**-0.4%**	**-110.5%**	**1.9%**	**2.0%**	**-8.8%**	**-2.9%**	**-2.6%**	**-111.3%**

Table 5-3. Continued

Model	Scenario	Definitions	Costs			QALYs			CU-Ratio		
			INT	CP	INC	INT	CP	INC	INT	CP	INC
Case Study 3	LT (per patient) Rimonabant vs. Diet & Exercise	LT (original)	3,054	2,496	558	14.173	14.108	0.065	215	177	8,581
		LT (reproduction)	2,933	2,464	470	13.658	13.583	0.075	215	181	6,263
		Difference	-120	-32	-88	-0.515	-0.525	0.010	0	4	-2,318
		Difference in %	**-3.9%**	**-1.3%**	**-15.8%**	**-3.6%**	**-3.7%**	**15.4%**	**0.0%**	**2.3%**	**-27.0%**
Case Study 4	LT (per patient) Slimming World vs. Usual Care	LT (original)	9,065	9,988	-923	12.580	12.360	0.220	721	808	-4,195
		LT (reproduction)	9,481	10,392	-911	12.779	12.550	0.229	742	828	-3,978
		Difference	416	404	12	0.199	0.190	0.009	21	20	217
		Difference in %	**4.6%**	**4.0%**	**-1.3%**	**1.6%**	**1.5%**	**4.1%**	**2.9%**	**2.5%**	**-5.2%**

40Y = 40-year time horizon (simulated by the original model); BSC = Best Supportive Care/Usual Care; CU = Cost-Utility; CP = Comparator; INC = Incremental; INT = Intervention; LT = Lifetime time horizon; SBT = standard behavioral therapy; SBT+list = standard behavioral therapy combined with provision of detailed meal plans and corresponding shopping lists

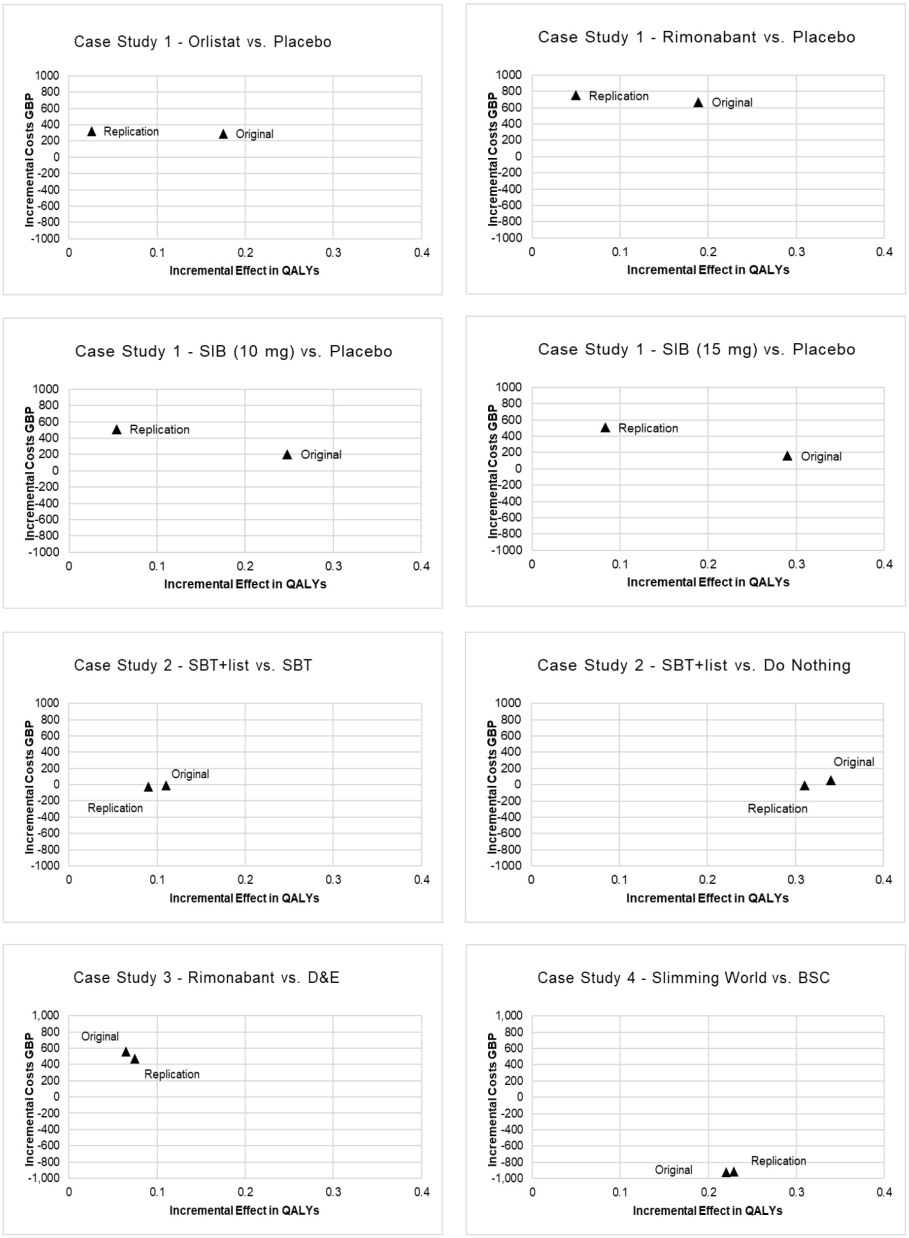


Figure 5-1. Incremental Cost-Effectiveness Results - Original versus Reproduction by Case Study and Comparison

BSC = Best Supportive Care / Usual Care; D&E = Diet & Exercise; SBT = standard behavioral therapy; SBT+list = standard behavioral therapy combined with provision of detailed meal plans and corresponding shopping lists; SIB = Sibutramine; QALY = quality-adjusted life years

In summary the intervention and comparator cost of the replication showed quite good results when compared to the original values; in Case Studies 2, 3 and 4, the variation in costs between the reproduction costs and the original costs was always <5%. This was also observed for the comparator in Case Study 1, but here the various intervention costs showed higher deviations (between 5.2 and 16.1%). Looking at the quality adjusted life year (QALY) result reproduction of the intervention and comparators, the variation observed was always <5%. However, when looking at the incremental cost and QALY results the relative deviation (in percent) increases substantially in all case studies. This comes about because the absolute incremental numbers are quite low and hence only a small deviation in absolute numbers translates into a much higher relative deviation. The same issue is observed when looking at the key outcome of the case studies, namely the ICER.

In Figure 5-1 it could be seen that the incremental costs were fairly comparable between the replication and the original (presented by the very similar height of the ICER point estimates for the replication and the original shown in the coordinate plane). This picture changes if looking at the incremental QALYs, where, especially in Case Study 1, a strong deviation is observed (presented by the horizontal distance between the ICER point estimates for the reproduction results and the original results). This distance is considerably smaller for Case Studies 2, 3 and 4, showing the best fit of reproduction results for Case Studies 3 and 4, in which the ICER point estimates almost overlap.

5.4.3 Assessment of the Success of Result Reproduction

The success ratings of reproduced results, according to the different criteria proposed in a recently published literature review [6], are presented in Table 5-4.

In summary the same conclusion for cost-effectiveness (in all studies defined as an ICER per QALY < 20,000 GBP) was reached in each investigated case study comparison; this reflects the broadest definition of reproduction success (success criteria #1).

Table 5-4. Assessment of the Success of Reproduced Results according to the Criteria proposed by McManus et al. 2019 [6]

Reproduction Success Definitions	#	Case study 1				Case Study 2		Case Study 3	Case Study 4
		Orlistat vs. Placebo	Rimonabant vs. Placebo	SIB 10 mg vs. Placebo	SIB 15 mg vs. Placebo	SBT+list vs. SBT	SBT+list vs. Do Nothing	Rimonabant vs. D&E	Slimming World vs. BSC
Same conclusions for intervention cost-effectiveness were reached	(i)	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Costs & outcomes reproduced for some treatment pathways / model scenarios and not others	(ii)	Best fit for simulating "Orlistat vs. Placebo"; worse result fit for the other intervention (Rimonabant & Sibutramine 10 / 15 mg)				Same direction of results for all pathways / model scenarios		Only one scenario	Only one scenario
Results for costs & utility vary by only 5% / 10% / 20% compared to the original and are consistent with the original conclusions	(iii)	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC
<5% variation (Costs)		no / yes / no	no / yes / no	no / yes / no	no / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes
<10% variation (Costs)		yes / yes / no	yes / yes / no	no / yes / no	no / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes
<20% variation (Costs)		yes / yes / yes	yes / yes / yes	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes	yes / yes / yes
<5% variation (Utility)		yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	no / no / no	yes / yes / no	yes / yes / no	yes / yes / yes
<10% variation (Utility)		yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes	yes / yes / no	yes / yes / yes
<20% variation (Utility)		yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes	yes / yes / yes	yes / yes / no	yes / yes / yes
Calculated CE-Ratio varies by only 5% /10% / 20% vs. the original	(iv)	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC	INT / CP / INC
<5% variation (CE-Ratio)		no / yes / no	no / yes / no	no / yes / no	no / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no
<10% variation (CE-Ratio)		yes / yes / no	yes / yes / no	no / yes / no	no / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes
<20% variation (CE-Ratio)		yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / no	yes / yes / yes
CE figures could be reproduced to a reasonable degree of success*	(v)	no	no	no	no	yes*	yes*	yes*	yes*
Identical results are produced	(vi)	no	no	no	no	No	no	no	no

* cost-effectiveness coordinate plane was reproduced; ** ICER <GBP 20,000; *BSC = Best Supportive Care / Usual Care; CP = Comparator; ICER = incremental cost-effectiveness ratio; INC = Incremental; INT = Intervention; SBT = standard behavioral therapy; SBT+list = standard behavioral therapy combined with provision of detailed meal plans and corresponding shopping lists; SIB = Sibutramine*

With regard to assessing a different degree of success in reproducing results, considering the different scenarios analyzed within one case study, for Case Study 1 the best reproduced results are observed for the comparison of "Orlistat vs. Placebo"; a worse result fit was observed for all the other comparative scenarios (10/15 mg Rimonabant & Sibutramine vs. placebo), whereas no such issues were identified for Case Studies 2, 3 and 4 (success criteria #2).

A smaller variation, of 5%, 10% or 20% in intervention and comparator costs, utilities and (intervention-specific) average CE ratios was observed in many cases. However, looking at incremental costs, utilities and the incremental CE ratio as well, this situation was rarely observed. This is due to the smaller absolute numbers when looking at incremental results; even small absolute variations might lead to a strong relative variation. A good example of this issue is observed in Case Study 2 for the comparison of "SBT + list vs. SBT". Here the original incremental costs are GBP -10, and in the reproduction the incremental costs are GBP -21, a result that is to be rated as quite comparable considering the 40-year simulation time horizon. However, when expressed as a percentage the relative variation comparing the original vs. the replication for this example is 110% (success criteria #3 and #4).

Therefore, for the assessment of incremental costs, QALYs and ICERs the calculation of relative variations may be misleading. This issue could be overcome by another success criteria, such as visualizing the original and reproduction of the incremental costs and QALYs in the cost-effectiveness coordinate plane. Here, the distance between the mean ICER estimates can be used to rate whether the result could be reproduced within a reasonable degree. On the basis of this approach, the rating of a successfully reproduced result was finally made for Case Studies 2, 3 and 4, but not for Case Study 1, where the variation of incremental QALYs was regarded as too strong. However, a partially successful reproduction could be seen in quite comparable incremental costs (success criteria #5).

The strictest criterion, namely the production of identical reproduction results, was observed in no case (success criteria #6). In order to rate the success of the final results of the reproduction, a combination of different broader and more specific criteria seems to be the most adequate approach. As a successful replication of a health economic model needs to result in the same cost-effectiveness conclusions, success criteria #1 needs to be considered. Furthermore, the assessment of the relative deviation of costs and utilities, as well of the average CE ratios (success criteria #3 & #4) should be considered (here the acceptable deviation could be set to 5%); whereas incremental results should not be assessed in a relative manner, due to the issue of small numbers described above. The application of this success factor assures that the reproduced results for the single interventions are within an acceptable error range. Finally, the ICER results should be visualized in the

cost-effectiveness plane in order to determine if the deviation presented is to be regarded as acceptable or not (success criteria #5), assuring that the ICER results are fairly comparable.

The proposed combination of success criteria were all clearly fulfilled for Case Studies 2, 3 and 4. In contrast, Case Study 1 shows strong variations (<10%) in relative cost, utility and CE-ratios, and also fails to present fairly acceptable ICER results (as visualized in Figure 5-1); accordingly, Case Study 1 needs to be rated as a failure in reproducing results.

5.4.4 Assessment of Model Replication Reporting Standards

The results of assessing the reporting quality according to the CHEERS checklist are presented in Table 5-5 for each case study of obesity model replication. With regard to the CHEERS total scoring outcomes there was no relevant difference in reporting quality observed between the case studies (the CHEERS score ranges between 18.0 and 20.0; the maximum possible CHEERS score is 24). The description of study (input) parameters (CHEERS item #18) is one of the most sensitive topics for a model replication; here Case Studies 1, 3 and 4 were rated as reporting the relevant data in part, whereas Case Study 2 was rated as reporting the relevant data in full.

Very specific information is required in order to enhance a successful model replication. Considering the identified key hurdles and applying the CHEERS guidance [18] on the quality of reporting related to these issues, it is determined whether the current consensus on reporting is adequate for successful model replications.

Table 5-5. CHEERS checklist results for all included obesity models / case studies

	Item #	Case Study 1 [14]		Case Study 2 [15]		Case Study 3 [16]		Case Study 4 [17]	
Title & Abstract		*Rating*	*Score*	*Rating*	*Score*	*Rating*	*Score*	*Rating*	*Score*
Title	1	yes	1	yes	1	yes	1	yes	1
Abstract	2	yes	1	yes	1	yes	1	yes	1
Introduction									
Background & objectives	3	yes	1	yes	1	yes	1	yes	1
Methods									
Target population & subgroups	4	yes	1	yes	1	yes	1	yes	1
Setting and location	5	yes	1	yes	1	yes	1	yes	1
Study perspective	6	yes	1	yes	1	yes	1	yes	1
Comparators	7	part	0.5	yes	1	part	0.5	part	0.5
Time horizon	8	part	0.5	part	0.5	part	0.5	part	0.5
Discount rate	9	part	0.5	yes	1	part	0.5	part	0.5
Choice of health outcomes	10	part	0.5	part	0.5	part	0.5	part	0.5
Measurement of effectiveness	11a	na		na		na		no	0
	11b	yes	1	yes	1	part	0.5	na	
Measurement & valuation of preference-based outcomes	12	yes	1	yes	1	yes	1	yes	1
Estimating resources and costs	13a	na		na		na			
	13b	part	0.5	yes	1	part	0.5	part	0.5
Currency, price date & conversion	14	yes	1	yes	1	yes	1	no	0
Choice of model	15	part	0.5	part	0.5	part	0.5	part	0.5
Assumptions	16	yes	1	yes	1	yes	1	yes	1
Analytical methods	17	yes	1	yes	1	yes	1	yes	1

Table 5-5. Continued

	Item #	Case Study 1 [14]		Case Study 2 [15]		Case Study 3 [16]		Case Study 4 [17]	
Results									
Study parameters	18	part	0.5	yes	1	part	0.5	part	0.5
Incremental costs & outcomes	19	yes	1	yes	1	yes	1	yes	1
Characterizing uncertainty	20a	na		na		na		na	
	20b	part	0.5	part	0.5	yes	1	yes	1
Characterizing heterogeneity	21	na	0	na	0	na	0	yes	1
Discussion									
Study findings, limitations, generalizability & current knowledge	22	yes	1	yes	1	yes	1	yes	1
Other									
Source of funding	23	yes	1	no	0	yes	1	yes	1
Conflicts of interest	24	yes	1	yes	1	no	0	yes	1
Total CHEERS Score			**19.0**		**20.0**		**18.0**		**18.5**

na = not applicable

With regard to the identified lack of reporting of standard deviations or distribution parameters, in order to enable the reproduction of PSA results (in three case studies), the CHEERS statement asks to "report the values, ranges, references, and if used, probability distributions for all parameters" [18]. However, it is not made clear in the related CHEERS example table that, in addition to the distribution type, the standard deviation is required to inform the PSA. This lack of clarity in the related CHEERS example table might have led to the observed situation, namely that all case studies that have applied a PSA (Case Studies 1, 3 and 4) have not provided all the required information. This resulted in the rating of "partial" compliance with the CHEERS criteria with regard to the quality of reporting study (input) parameters (CHEERS item #18).

Two further identified key hurdles for model replication are also related to the reporting of input parameters, namely the lack of reporting of details on life tables (Case Studies 3 and 4) as well as the introduction of several self-created regression analyses without providing details on how to apply/solve the provided regressions correctly (Case Study 1). All those aspects are also related to the CHEERS criteria related to the quality of reporting study (input) parameters (CHEERS item #18). These were already rated as being in "partial" compliance due to the PSA issue stated above.

With regard to the identified lack of reporting of clinical event results (in two case studies), the CHEERS statement offers no guidance or related requirements. Hence the missing information on clinical event results (Case Studies 1 and 2) had no impact on the CHEERS rating on the quality of reporting results (CHEERS item #19). As generally health economic models are driven by clinical events and related mortality, we believe that this issue should be addressed in future adaptations of the reporting standard. In this context it would be most helpful to present event and mortality results (for all simulated alternatives) over the whole time horizon of the model, for each model cycle; this would be most helpful for checking how adequately a model adaptation predicts the underlying clinical events. This information helps in identifying whether a potential result deviation (between the original and reproduction) is driven by clinical events or by the related cost and utility valuation approach of these health states.

5.5 Discussion

This study confirms the feasibility of rebuilding four identified health economic obesity models. However, success in reproducing results was observed in only three out of the four studies, and some challenges were observed. The replication

of health economic models is an important topic, especially as there is no broad application of open source models, although these were proposed by several authors in order to enhance model transparency and result credibility [20-22]. Such open source models would have the advantage of joint development, joint validation and ongoing improvement by the scientific community, but to date only a view open source models are available, mainly due to lack of funding and other challenges (e.g. organization, software & intellectual property restrictions) of such initiatives [23, 22]. The replication of a health economic decision analytic model is a complex exercise, and one should keep in mind that the more information and results of a model that are provided, the more information is available to investigate whether a result reproduction was successful or not. From the perspective of a modeler performing a replication of quite complex long-term obesity models, it is extremely helpful if the authors publish the simulated clinical event output frequencies, as these make it possible to check whether the event simulation and hence the clinical heart of the replicated model is working correctly (or not). If the clinical event frequencies are comparable, the replication of the structure of the model and transition probabilities can be considered correct. If the ICER is then different, the reason can be due to the inappropriate replication of the costs or utilities, or inappropriate reporting of costs or utilities by authors. This knowledge helped to determine the source of potentially observed mismatches between original and reproduction results, as it enabled the researchers to better locate the potential issue. It is no coincidence that this information was not provided for Case Study 1, for which we failed to perform a successful result reproduction.

However, it needs to be taken into account that the replication of a model itself is an error-prone exercise. Hence a failed reproduction could be based on errors made during programming, and might not necessarily come from a lack of documentation or inadequate reporting by the original authors. In order to minimize this potential source of errors we have used specialized modelling software (TreeAge Pro Healthcare) to rebuild the selected health economic obesity models. Consequently, potential errors might be due mainly to input data typos, as building the model structure (and related calculations) is widely automated. However, using TreeAge instead of the software used in the original study could also be an issue preventing a 1:1 reproduction of modelling results, due to the automatic application of some TreeAge features (e.g. automatic half cycle correction) as stated in detail in the appendix tables. Furthermore, it needs to be considered that the success of a model replication is also influenced by the skills of the programmer; hence one limitation is that the replication was performed by only one modeler. However, this modeler has over 20 years of experience as a professional health economist and all critical issues were reviewed and discussed within the team, which included experienced health economic modelers. On the other hand, programming errors in the original publication could not be ruled out completely, as especially complex

Excel models require complex testing and validation to assure the correctness of all calculations, and this might also impact the presented reproduction results.

For assessing the success of the reproduction results we have applied different criteria as defined and proposed in a recently published review on this topic [6]; to our knowledge these criteria were applied for the first time in this study to systematically assess the reproduction success. The six criteria applied range from very broad to very specific; accordingly, it is easier or harder to fulfill them. The strictest criterion, namely that identical results be produced, was not achieved by any of the case studies. This is not unexpected considering that all obesity models were simulating a long-term time horizon, and hence a small deviation (even a rounding issue) will get more and more pronounced over time. Another reason may be the high complexity of obesity models, triggered by including all the relevant complications of obesity. The larger the complexity, the larger the chance of misinterpretating the data, assumptions and model structure description in the original paper, combined with a higher probability of errors by either the replicator or the original programmer. This strictest definition does not seem to be very helpful as identical results have not yet been achieved with regard to the publications of other model replications [1, 2]. Moreover, the other proposed criteria were not rated as sufficient to adequately define reproduction success, as all were rated as too soft to act as stand-alone reproduction success criteria. Therefore, we have used a combination of various criteria in order to investigate and to determine the success of reproduction. Although this proposed combination does not assure identical results, it assures that the cost-effectiveness conclusion is identical, that the deviation in single components is acceptable (<5%) and that the incremental cost-effectiveness results are fairly comparable. As this study was to our knowledge the first application of these replication success criteria, and hence also of this criteria combination, further research and scientific dialogue is required to investigate and define how to best rate the success of a health economic model replication; we believe that the applied criteria developed by McManus et al. [6] and our research will help to inform this scientific dialogue.

The identified key model replication facilitators were input data tables and model diagrams showing the model structure and possible state transitions. Key replication hurdles were missing standard deviations for performing probabilistic analysis, missing clinical event results, missing details on applied life tables and missing formulas for equations based on own calculations. Whereas the key facilitators were quite in line with those identified by other research teams [2, 1], our identified key barriers seem to be more specific than those identified in previous research. This might be related primarily to the fact that we have selected long-term obesity models, whereas other research teams [2, 1] included a broader range of health economic models, including short and long-term time horizons

and different disease areas. This focus on only one disease area and on long-term models is also a limitation of our research; the transferability of our findings to other kinds of health economic models needs to be investigated by future research.

Looking specifically at the reproduction of the cost and utility results of single strategies, a previous study [2] found that there was a tendency for greater variation in the reproduced costs than outcomes, which is also seen in our research; costs ranged from -3.9% to 16.1% (mean over all model simulations: 3.78%) whereas QALYs varied by -3.7% to 2.1% (mean: -0.11%). However, looking at the comparison of reproduced results and original results in terms of incremental cost and QALYs (please refer to Figure 5-1), which was done for the first time in our study, the observed variations in incremental QALYs were more pronounced than the variations in incremental costs; this highlights the importance of reporting and visualizing incremental replication results.

As one key facilitator McManus et al. [2] suggested that cost and outcome results should be presented over time in an additional table to enhance model replication. We agree that this information would be very helpful for replication, especially to see from which point in time deviations between reproduced results and original results are observed. However, on the basis of this information it wouldn't be clear where the replication error might be located, which is why we are suggesting that the clinical events be presented over time. If it is possible to reproduce the results of the clinical events, the structure and related transition probabilities are replicated correctly. If a deviation in costs or outcomes is then observed, this is related to costs or to the parameter values for costs and utilities, and the methodology of including these parameters. Hence in the best case all model outcomes, including the underlying event rates, would be presented to facilitate model replication.

We applied the CHEERS checklist [18] as it looks particularly at the quality of reporting, a core criterion for successful model replication, and as it was found to be the most commonly used checklist since 2017 in a recently published systematic review [3]. Other frequently applied checklists (such as the Phillips checklist [24] or the CHEC project [25]) assess the quality of conducting the health economic study, which was not our key focus. We investigated whether the CHEERS score might be predictive for the success of model replication, but this was not the case, with scores ranging from 18 to 20. The non-successful replication Case Study 1 rated a score of 19 (the maximum possible CHEERS score is 24). A comparable finding was observed by another research team that investigated the Phillips checklist [24] in the context of model replication; they found that the Phillips checklist was not reliable for ensuring that studies are replicable [2]. However, we believe that simple changes in the CHEERS reporting criteria might be adequate to solve the

key hurdles for model replicability that we observed in our presented research. These proposed changes are:

(i) the probability distribution and all the necessary parameters to define its shape are to be presented for all input parameters, assuring the reproduction of probabilistic sensitivity analyses results;

(ii) When a model simulates (clinical) events, the event simulation results should be presented, to guide potential necessary assumptions and to better locate potential replication errors;

(iii) for all included regressions/risk equations (whether published or unpublished) applied in the model, the calculation formula should be presented, preferably with an application example, to assure the correct replication of formula-based transition probabilities, costs and outcomes.

Although the current CHEERS statement covers parts of these aspects - namely it asks for "probability distributions for all parameters" to be included and for "outcomes of interest" to be reported - we believe that these aspects need to be made clearer to adequately guide reporting on the model.

To our knowledge there are currently no other publications that suggest specific changes to CHEERS or other health economic reporting guidelines to enhance health economic model replication. However, we have identified a recently published paper that suggests a nine-item osteoporosis-specific addition to the CHEERS checklist, in order to address disease-specific issues adequately [26]. The further development of health economic reporting standards is an ongoing process and there is a specific ISPOR task force currently working on an update of the CHEERS criteria.

5.6 Conclusions

The small changes to existing reporting criteria, as presented above, may increase both the transparency of health economic model reporting and the success of reproducing its consequent results. Proofing the replicability of our health economic simulation "experiments" might increase the scientific rigor and acceptance of our field. In conclusion, model replications can help to assess the quality of health economic model documentation, can be used to validate and refine current model reporting practices, and might subsequently increase the transparency and acceptance of health economic modelling studies.

5.7 References

1. Bermejo I, Tappenden P, Youn JH. Replicating Health Economic Models: Firm Foundations or a House of Cards? PharmacoEconomics. 2017;35(11):1113-21. doi:10.1007/s40273-017-0553-x.
2. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and Facilitators to Model Replication Within Health Economics. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2019;22(9):1018-25. doi:10.1016/j.jval.2019.04.1928.
3. Watts RD, Li IW. Use of Checklists in Reviews of Health Economic Evaluations, 2010 to 2018. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2019;22(3):377-82. doi:10.1016/j.jval.2018.10.006.
4. Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB et al. AHRQ Methods for Effective Health Care. Best Practices for Conducting Economic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.
5. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50. doi:10.1016/j.jval.2012.04.012.
6. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. PharmacoEconomics. 2019;37(11):1371-81. doi:10.1007/s40273-019-00836-y.
7. Schwander B, Hiligsmann M, Nuijten M, Evers S. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert review of pharmacoeconomics & outcomes research. 2016;16(5):561-70. doi:10.1080/14737167.2016.1230497.
8. Schwander B, Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert review of pharmacoeconomics & outcomes research. 2018;18(5):529-41. doi:10.1080/14737167.2018.1501680.
9. Schwander B, Nuijten M, Hiligsmann M, Queally M, Leidl R, Joore M et al. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy and Technology. 2020;9(3):314-22. doi:https://doi.org/10.1016/j.hlpt.2020.03.005.
10. Yi Y, Philips Z, Bergman G, Burslem K. Economic models in type 2 diabetes. Curr Med Res Opin. 2010;26(9):2105-18. doi:10.1185/03007995.2010.494451.
11. Brown JB, Palmer AJ, Bisgaard P, Chan W, Pedula K, Russell A. The Mt. Hood challenge: cross-testing two diabetes simulation models. Diabetes Res Clin Pract. 2000;50 Suppl 3:S57-64. doi:10.1016/s0168-8227(00)00217-5.
12. Palmer AJ, Clarke P, Gray A, Leal J, Lloyd A, Grant D et al. Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(4):670-85. doi:10.1016/j.jval.2013.01.002.
13. Si L, Willis MS, Asseburg C, Nilsson A, Tew M, Clarke PM et al. Evaluating the Ability of Economic Models of Diabetes to Simulate New Cardiovascular Outcomes Trials: A Report on the Ninth Mount Hood Diabetes Challenge. Value in Health. 2020;23(9):1163-70. doi:10.1016/j.jval.2020.04.1832.

14. Ara R, Blake L, Gray L, Hernandez M, Crowther M, Dunkley A et al. What is the clinical effectiveness and cost-effectiveness of using drugs in treating obese patients in primary care? A systematic review. Health Technol Assess. 2012;16(5):iii-xiv, 1-195. doi:10.3310/hta16050.

15. Au N, Marsden G, Mortimer D, Lorgelly PK. The cost-effectiveness of shopping to a predetermined grocery list to reduce overweight and obesity. Nutr Diabetes. 2013;3:e77. doi:10.1038/nutd.2013.18.

16. Caro J, Stillman O, Danel A, Getsios D, McEwan P. Cost effectiveness of rimonabant use in patients at increased cardiometabolic risk: estimates from a Markov model. Journal of Medical Economics. 2007;10(3):239-54. doi:10.3111/13696990701438629.

17. Meads DM, Hulme CT, Hall P, Hill AJ. The cost-effectiveness of primary care referral to a UK commercial weight loss programme. Clin Obes. 2014;4(6):324-32. doi:10.1111/cob.12077.

18. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(2):231-50. doi:10.1016/j.jval.2013.02.002.

19. Hiligsmann M, Evers SM, Ben Sedrine W, Kanis JA, Ramaekers B, Reginster JY et al. A systematic review of cost-effectiveness analyses of drugs for postmenopausal osteoporosis. PharmacoEconomics. 2015;33(3):205-24. doi:10.1007/s40273-014-0231-1.

20. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. PharmacoEconomics. 2017;35(1):125-8. doi:10.1007/s40273-016-0479-8.

21. Schramm W, Sailer F, Pobiruchin M, Weiss C. PROSIT Open Source Disease Models for Diabetes Mellitus. Stud Health Technol Inform. 2016;226:115-8.

22. Jansen JP, Incerti D, Linthicum MT. Developing Open-Source Models for the US Health System: Practical Experiences and Challenges to Date with the Open-Source Value Project. PharmacoEconomics. 2019;37(11):1313-20. doi:10.1007/s40273-019-00827-z.

23. Emerson J, Bacon R, Kent A, Neumann PJ, Cohen JT. Publication of Decision Model Source Code: Attitudes of Health Economics Authors. PharmacoEconomics. 2019;37(11):1409-10. doi:10.1007/s40273-019-00796-3.

24. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. PharmacoEconomics. 2006;24(4):355-71. doi:10.2165/00019053-200624040-00006.

25. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess Health Care. 2005;21(2):240-5.

26. Hiligsmann M, Reginster JY, Tosteson ANA, Bukata SV, Saag KG, Gold DT et al. Recommendations for the conduct of economic evaluations in osteoporosis: outcomes of an experts' consensus meeting organized by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO) and the US branch of the International Osteoporosis Foundation. Osteoporos Int. 2019;30(1):45-57. doi:10.1007/s00198-018-4744-x.

5.8 Appendix

5.8.1 Details on Model Selection

The step-wise model selection process is presented in Table 5-6, which shows how many models complied with a specific selection criterion, how many were excluded in a specific selection step, and how many models remained after each selection step.

Table 5-6. Details on the step-wise model selection process

Model Selection Steps	Step 0	Selection Step 1	Selection Step 2	Selection Step 3	Selection Step 4	Selection Step 5
Selection Criteria	Total models	Lifetime or Comparable	Selection STM/DES	T2D / CHD / Stroke	UK	Adults
N complied with the specific Criterion	87	55	76	35	15	70
Excluded by Selection Step	0	-32	-7	-20	-23	-1
Remaining Models	87	55	48	28	5	**4**

STM = State Transition Model, DES = Discrete Event Simulation, UK = United Kingdom

5.8.2 Details on identified key replication facilitators and key hurdles

Below the identified key facilitators and key hurdles are presented as summary for each case study:

Case Study 1

An overview of key facilitators and hurdles for the replication process is provided below:

Key Facilitators (Case Study 1):

- Model structure and possible state transitions were presented in a state transition diagram
- Overview of input parameters was provided in table format

Key Hurdles (Case Study 1):

- Clarification on how to apply/solve the several self-created regression analyses (for primary event risks, secondary event risks and utilities) was missing (in the only case where the function was provided, it was misleading as instead of the required Weibull survival function, the formula for the Weibull hazard function was provided to estimate transition probabilities)

- For input parameters, which were sampled for probabilistic sensitivity analyses as beta/gamma distributions, neither the SD nor the distribution parameters were provided, hence it was not possible to replicate the PSA results
- Although detailed results of the underlying mixed treatment comparison focusing on the BMI reduction of specific obesity drugs were provided, some values for simulated interventions were missing and it was unclear how they were estimated on the basis of the given data
- Results of event simulation were not provided (which would have been very helpful to guide the various assumptions required for rebuilding the model)

Case Study 2

An overview of key facilitators and hurdles for the replication process is provided below:

Key Facilitators (Case Study 2):

- Model structure and possible state transitions were presented in a state transition diagram
- Details of BMI development were presented for both alternatives as figures
- The base risks, relative, risks, the related mortality, the base utilities, the disutilities, the intervention and the event costs were all presented in table format

Key Hurdles (Case Study 2):

- Gender distribution (in the base case) is not clearly stated in the manuscript
- BMI development for both alternatives was presented only as figures – hence the values needed to be measured within and pulled out from the figure
- There are no event simulation results provided (which would provide an additional means for testing the fit of the replication)

Case Study 3

An overview of key facilitators and hurdles for the replication process is provided below:

Key Facilitators (Case Study 3):

- Model structure and possible state transitions were presented in a state transition diagram
- Event-specific probabilities of dying, event costs, BMI-related utility-decrements and event-specific disutilities were presented in table format

Key Hurdles (Case Study 3):

- Only mean values (for risk equation parameters, costs and utilities) were provided; these informed only a deterministic approach (although a cost-

effectiveness acceptability curve presenting PSA results was published); hence a replication of PSA results was not possible
- In order to adjust the applied Framingham algorithms for the influence of BMI, the authors added an additional risk factor (BMI) to the equations. How this risk was added to the equations was not described in detail and therefore it was not replicable
- The applied risk equations were not provided in the paper (only referenced); a clear presentation of the calculation approach (e.g. as equation formula below the risk factor table) would have strongly simplified the replication process
- The underlying UK life tables were not provided, nor was it stated which life tables (year of data) were used in the model
- For the age-related utility norms only a reference was provided. This reference was not accessible, so an alternative study providing UK-specific norms was used

Case Study 4

An overview of key facilitators and hurdles for the replication process is provided below:

Key Facilitators (Case Study 4):
- Model structure and possible state transitions were presented in a state transition diagram
- For population key parameters (age, BMI, gender) the mean and standard deviations were provided
- Transition probabilities, costs and health utilities were presented as mean values for each health state in table format, including information of the kind of distribution used in PSA
- Event-specific probability of dying was presented in table format

Key Hurdles (Case Study 4):
- On the basis of the provided mean/SD population parameters (age, BMI, gender) it was not possible to exactly reproduce the real-world data correlations (meaning the exact parameter distribution in single patients)
- For transition probabilities, costs and health utilities, neither the distribution parameters nor the standard deviation of the mean were given; hence a replication of PSA results was not possible
- The underlying UK life tables were not provided, nor was it stated which life tables (year of data) were used in the model
- It is unclear whether the results presented in table format are based on deterministic or probabilistic analyses; it is unclear whether a half-cycle correction was performed

5.8.3 Model Replication

The following topics were investigated in detail: model type & model structure, baseline population, weight/body mass index (BMI) development over time, event and mortality simulation, utility estimation/ calculation (including discounting), cost estimation/calculation (including discounting), cost-effectiveness results (deterministic/probabilistic). For each of the presented topics the available information was reviewed in detail and facilitators, hurdles and barriers for the replication process were determined and documented. The findings of this initial assessment were extended by specific issues/findings determined during the model replication process. Interestingly, some of the model replication hurdles were only identified and added to the tables presented below during the programming process. For example, it was not obvious during the initial review that the provided information on the input value distributions (of case studies 1, 3 and 4) does not allow the PSA to be performed. Also the issues around how to apply/solve the several self-created regression analyses (Case Study 1) were not identified in the initial review, which explains the two-step strategy (a pre-defined data/information availability check, supplemented by findings during the replication procedure). The details related to each case study are presented below in table format (Table 5-6 to Table 5-10).

Table 5-7. Case Study 1 - detailed Facilitators and Hurdles of Model Replication

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Model Structure/ Health States	STM: T2D, MI, Stroke, Fatal CVD Events; Death from other causes	Model structure and possible state transitions were presented in a state transition diagram.	None	None
Baseline Population	1,000,000 patients (deterministic); 400,000 (stochastic); age 45.5 years, 25.7% male, 33.2% diabetic, mean BMI 34.92 kg/m^2	Population input values are stated in detail in the underlying HTA report.	Values in the text and in the diagram do not correspond to each other (from the replicators' point of view, values in the text are correct). Baseline covariate summary statistics are not provided for each covariate included in the Weibull event functions. E.g. Smoking status is not provided.	None
BMI Development / Impact of Treatment	Impact of the weight loss strategies are described in the HTA and presented in detail as result of the mixed treatment comparison (MTC).	The correlated changes in BMI at 3, 6 and 12 months from the MTC were presented in detail in the underlying HTA report.	Changes in BMI are not provided for each interventions for each time points, it is unclear which values were used in this case.	Assumptions needed to be performed related to missing BMI changes.
BMI influence on transition probabilities	BMI was part of the risk functions that were used for each event and mortality simulation; hence the BMI had a direct influence on the risk.	The approach of developing the risk functions is described in detail in the HTA report (Chapter 4); for each event the related time-to-event analysis is presented as variance–covariance matrix; the applied distributions are presented for each input parameter in table format in the underlying HTA report.	On the basis of the provided information and tables it is not possible to calculate time-dependent risks (transition probabilities for the simulated events and mortality) as only the related Weibull hazard function was provided. No related application formulas were provided for the regression analyses for the secondary events.	When applying the Weibull survival function (not provided in the paper) it is possible to estimate the primary event risks on the basis of the variance covariance matrixes. However, for the secondary events only different regression analyses results are provided to estimate the probability of secondary CV events (not separating MI or stroke) on the basis of the patient's age. On the basis of the given information it was unclear how to apply the different regressions. As the data source for the basis of the regression analyses was provided it was possible to extract the published values and to run own regression analyses that were used to estimate the probability of subsequent CV events. **These central assumptions necessary for the replication may be responsible for the differences observed in the reproduced costs and utilities**.
Transition Probabilities & Distributions (PSA)	Transition probabilities are based on the BMI risk functions; distributions are described as part of the variance–covariance matrices as well as for each input parameter.			
Mortality	Mortality was based on BMI risk functions.			

Table 5-7. Continued

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Utility	Age and health-condition related EQ-5D scores were used to simulate state-specific utility values.	EQ-5D scores and mean errors in predicted values were presented in table format.	The description lacks details on how the presented values were used/applied in the model (e.g. age-specific base utilities were decreased by event-specific disutilities). The formula to apply the presented regression analysis is not provided.	It is stated that "after controlling for age, gender and health status (history and time since heart attack, stroke, angina, and diabetes), the results of the regression showed an independent relationship between BMI and EQ-5D score". This BMI influence was not presented in the HTA report, which prevents a 1:1 reproduction of the utility values. However, we have used the underlying publication cited in the HTA in order to estimate utility values for the different health states, which might be the key reason for differences observed in the reproduced utilities.
Costs	Drug Costs, Monitoring Costs and Health State Costs were considered in the model.	All costs are provided in table format. Potential error in the table – stroke costs combined with T2D are lower in the first year than in the subsequent years.	Details on the gamma distribution applied for the health state costs during the PSA are not provided (at least SD should be reported to estimate the adequate distribution parameters).	As not even standard deviations of the means are provided for values that were sampled as beta or gamma distribution it was not possible to reproduce PSA results.
Cost-Effectiveness/ Event Results	Life Years, QALYs, Total Costs and ICER per QALY gained are presented for the simulated alternatives.	Cost-utility results are presented in table format, as well as the ICER results for one-way sensitivity analyses; furthermore cost-effectiveness planes and cost-acceptability curves are provided.	There are no event frequencies/results provided. Furthermore details on the beta and gamma distributions applied are not provided.	As there are no event frequencies/ results provided it was not possible to compare the number of events of the replicated model to the original results. As not even standard deviations of the means are provided for values that were sampled as beta or gamma distribution it was not possible to reproduce PSA results. **Missing SD or distribution parameters prevented replication of the PSA.** **Publication of event frequencies is always helpful for replication in order to guide the direction of assumptions that might need to be applied during the replication process.**

Table 5-7. Continued

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Simulation Settings	Deterministic, probabilistic and one-way sensitivity analyses were performed (lifetime horizon).	Definition of sensitivity analyses applied are provided in the HTA.	None	None
Software	The model was built in Simul8 version 17.0.	None	None	We have used TreeAge Pro to rebuild the model.

Table 5-8. Case Study 2 - detailed Facilitators and Hurdles of Model Replication

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Model Structure /Health States	STM: T2D, CHD, Stroke, Osteoarthritis, Colorectal Cancer, Related Death, Unrelated Death	Model structure and possible state transitions were presented in a state transition diagram.	None	Structure does not differentiate between primary and subsequent events. Furthermore there are no combined disease states (e.g. T2D & CHD etc.); hence the model forgets previous disease states once another disease is occurring.
Baseline Population	1,000 patients, aged 40 years with a BMI of 33 kg/m^2 and free of any obesity-related illnesses	Input values are stated in the underlying manuscript.	Gender distribution (in the base case) is not clearly stated in the underlying manuscript.	Unclear whether the base case simulated only women (as in the underlying trials) or a mixed population of men and women. According to the result comparability a 1:1 gender distribution was simulated.
BMI Development/ Impact of Treatment	The impact of the weight loss strategies is described in the underlying manuscript. and is presented in detail.	Details of BMI development were presented for both alternatives as figures.	Detailed BMI development presented only as figures, not as absolute BMI change from baseline	As in the model the BMI change (from baseline) is the key factor for changes of event rates; the details had to be extracted from the provided figure (by measuring the changes presented in the figure). Here it would have been helpful if the numbers had been presented.

Table 5-8. Continued

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
BMI influence on transition probabilities	The change in BMI influences the relative risk of obesity-related events.	The relative risk for a 1-point change in BMI was presented in detail.	None	The change in BMI is the central impact factor that influences the event risk/ transition probabilities; therefore the base event risk needs to clearly reflect the risk for the population to be simulated (which was considered by the authors for the base risk selection).
Transition Probabilities & Distributions (PSA)	Transition probabilities are based on a combination of base risks and BMI-related relative risks.	The base risks and the BMI related risks are presented in detail in table format.	None	
Mortality	"Unrelated" mortality is estimated on the basis of age and gender-specific life tables; event-specific mortality is estimated by gender-specific probabilities.	The event-specific mortality rates were presented in table format.	"Unrelated mortality" was based on UK life tables; the link provided was no longer working.	The best-fitting UK life tables (year 2011) were determined and downloaded; a deviation of results might be based on the use of different UK life tables
Utility	Gender-related base utilities which were reduced by BMI-group and by event-specific disutilities	Base-utilities and disutilities were provided in table format.	None	None
Costs	Intervention costs and event-specific costs were considered in the model.	Both intervention and event-specific costs were presented in table format.	None	None
Cost-Effectiveness/ Event Results	QALYs, total Costs and ICER per QALY gained are presented for the base case, the best case and the worst case.	Cost-utility results are presented in table format, as well as the ICER results for one-way sensitivity analyses.	There are no event frequencies/ results provided.	As there are no event frequencies/ results provided it is not possible to compare the number of events of the replicated model to the original results.
Simulation Settings	Deterministic analyses of the base case, best case and worst case as well as one-way sensitivity analysis (probabilistic sensitivity analyses were not performed); a 40-year time horizon was simulated.	Definition of best case/ worst case scenarios was briefly provided in the manuscript.	An additional LT horizon simulation would have made sense due to the impact of obesity-associated events on survival.	Reproducing the best case and worst case results or the one-way sensitivity analysis was not attempted; however, with the information provided this seemed to be possible.
Software	The model was built in Microsoft Excel.	None	It's difficult to test and validate complex Excel models.	We used TreeAge Pro to rebuild the model; in this context we also applied the standard half-cycle correction which was not applied in Excel; this may have prevented a 1:1 reproduction of modelling results.

Table 5-9. Case Study 3 - detailed Facilitators and Hurdles of Model Replication

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Model Structure/ Health States	STM: T2D, Primary CVD and CBV events (MI, AP, TIA, Stroke), Secondary CVD (MI or Stroke), Death	Model structure and possible state transitions were presented in a state transition diagram.	None	The structure presented does not differentiate between the different CVD and CBV events – during replication a comparable structure was applied, and hence the specific events were not presented in the tree. Thus, the impact of specific events on cost and utility were considered in the related cost and utility formulas.
Baseline Population	Baseline characteristics/ risk factors of 1,000 patients (T2D status, gender, age, smoker, BMI, SBP, cholesterol, HDL, triglycerides, HbA1C, fasting glucose, family history of diabetes)	For key parameters (age, BMI, gender) the mean values were provided for a "disease free" population and for a population with "T2D".	Only mean values were provided, which informed a deterministic approach – hence no probabilistic analyses were performed.	The population characteristics form the basis for feeding the Framingham, UKPDS and Stern algorithms that were used as the basis for estimating the disease-specific risks/transition probabilities.
BMI Development / Impact of Treatment	The impact of treatment on the various risk factors was provided for the intervention and the control group based on evidence from a randomized controlled trial.	The effect of therapy was obtained for the treatment duration (12 months) and was then assumed to decline to "zero" in a linear manner over the same period of time.	None	A natural increase of risk factors (other than "age") was neither assumed nor simulated in the model.
BMI influence on transition probabilities	BMI was a risk factor in the T2D algorithm but not in the (applied) Framingham / UKPDS algorithms.	In order to adjust for the BMI an additional risk was added to the CVD/CBV algorithms.	A detailed description of how this risk was added to the equations was not provided.	As we have identified no valid format for integrating such an additional BMI- related risk, we applied only the original equations, which might explain the small deviations in event, utility and cost outcomes.
Transition Probabilities & Distributions (PSA)	Transition probabilities are based on the Framingham, UKPDS and Stern algorithms, or no distributions were provided; for subsequent events fixed transition probabilities were provided.	References to the applied algorithms were provided, as well as the approach of estimating monthly transition probabilities.	The applied risk equations were not provided in the paper. This would have simplified the replication process, especially if the approach to calculation had been clearly presented – e.g. as a formula below the risk factor table.	First of all it was necessary to retrieve and to understand the papers/ equations on the different equations, which is much more difficult if no calculation example is provided; **there might be an issue with the T2D equation applied in the model, as a correction for this was published.**

Table 5-9. Continued

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Mortality	General mortality is based on UK life tables; events-specific mortality is simulated as specified probability of death in the acute phase and afterwards.	Event-specific probabilities of death were presented in table format.	The underlying UK life tables were not provided – it was also not stated which life tables (year of data) were used in the model.	As the model was published in 2007 we have used UK life tables reflecting the year 2005 – however this issue might impact the survival simulation and hence prevent a 1:1 reproduction of modelling results.
Utility	Age-related utilities norms; BMI-related decrements and events-specific disutilities were considered in the model.	Details on BMI-related decrements and events-specific disutilities were provided in table format.	For the age-related norms only the reference was provided. This reference was not accessible so an alternative study providing UK-specific norms was used..	As another study had to be used for the age-specific utility norms, this might have impacted the utility simulation and hence prevented a 1:1 reproduction of modelling results.
Costs	Intervention costs & event costs for the acute and the chronic phase were considered in the model.	Details on the event costs were provided in table format.	Details on intervention costs were provided only in the text. The costs of doctor and nurse visits were not provided, so these could not be included.	As the costs of doctor and nurse visits were not provided, these could not be included in the intervention costs; this might have prevented a 1:1 reproduction of modelling results.
Cost-Effectiveness Results	ICER per QALY gained for LT is provided; Furthermore a WTP acceptability curve for LT is presented.	Information is provided in table and figure format and adequate for comparison to reproduced results.	None	None
Simulation Settings	In general the model enables deterministic analyses; PSA results are provided (WTP acceptability curve).	None	No distributions nor SD are provided for the mean data; only the minimum and maximum parameters of single variables were investigated and provided in a tornado diagram.	As the key information on the applied distributions and the related standard deviation for the PSA was not provided, it was not possible to reproduce the WTP acceptability curve.
Software	All Analyses were conducted in MS Excel.	None	It is difficult to test and validate complex Excel models.	We have therefore used TreeAge Pro to rebuild the model – we simulated a microsimulation in order to reflect the 1,000 patients – in this context we also applied the standard half-cycle correction which was not applied in Excel; this might have prevented a 1:1 reproduction of modelling results.

Table 5-10. Case Study 4 - detailed Facilitators and Hurdles of Model Replication

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Model Structure / Health States	STM: T2D, Primary MI, Secondary MI, Primary Stroke, Secondary Stroke, Death	Model structure and possible state transitions were presented in a state transition diagram	None	Structure is conservatively simplified. This means it is not possible to develop T2D after MI or Stroke or to have both MI and Stroke events.
Baseline Population	Real-world dataset on baseline characteristics of 1,000 patients (gender, age, BMI)	For key parameters (age, BMI, gender) the mean and standard deviations were provided.	On the basis of the provided mean/SD it was not possible to simulate the real-world data correlations (these are usually tied to one person).	The population was estimated on the basis of "unconnected distributions" of patient characteristics. This issue is seen as a major factor in preventing a 1:1 reproduction of the underlying cohort.
BMI Development	Impact of the 12-month weight loss program: at 3 months −4.00 kg; at 6 months −4.25 kg ; at 12 months −4.5; afterwards natural weight gain of +0.429 kg (starting at year 2)	Detailed information is provided on weight loss related to specific time points.	It is unclear how this weight loss was simulated in annual cycles.	It was assumed that the mean weight loss of -4.25 kg is to be applied during the first annual cycles.
BMI influence on transition probabilities	BMI group-specific transition probabilities were provided for each event.	Details on underlying BMI groups were provided in the state transition diagram and for each health state.	None	None
Transition Probabilities & Distributions (PSA)	Transition probabilities (mean value) and kind of distribution were provided.	Transition probabilities were presented as mean value for each health state in table format, including information of the kind of distribution used in PSA.	Neither the distribution parameters nor the standard deviation of the mean were given. Furthermore, in the paper the increased likelihood of stroke & MI in T2D patients is stated – but the related risk increase was not stated.	Both the distribution parameters (on the basis of an assumed SD – assumption SD = 10% of the mean) and the risk increase of stroke & MI due to T2D (doubled risk for MI or Stroke according to ACC/AHA risk calculator) needed to be assumed. This issue is seen as one major factor in preventing a 1:1 reproduction of the PSA results, and of the event simulation results as well.
Mortality	General mortality is based on UK life tables; events-specific mortality is simulated either as a factor that increases general mortality (T2D) or as a specified probability of death (MI and Stroke).	Event-specific probability of death was presented in table format.	The underlying UK life tables were not provided, nor was it stated which life tables (year of data) were used in the model.	As the model was published in 2014 we have used UK life tables reflecting the years 2011-2013 – however this issue might impact the survival simulation and hence prevent a 1:1 reproduction of modelling results.

Table 5-10. Continued

Topic / Step	Basic Info	Facilitators	Hurdles / Barriers	Comments
Utility	Only the mean and the kind of distribution (beta) were provided for age-related utilities norms, BMI-related decrements and event-specific disutilities and for utilities.	Details on BMI-related decrements and events-specific disutilities were provided in table format, together with the information on the kind of distribution applied in the PSA (beta).	Only the reference was provided for the age-related norms. However, it was possible to determine the underlying data by retrieving the base publication. Neither the distribution parameters nor the standard deviation of the mean was given.	The distribution parameters needed to be assumed (on the basis of an assumed SD – assumption SD = 10% of the mean) for both utilities and costs. The missing standard deviation or distribution values for the utilities are seen as a major factor in preventing a 1:1 reproduction of the PSA results.
Costs	Program costs & event costs (T2D, MI 1st year, MI subsequent years; Stroke 1st year and Stroke subsequent years).	Details on mean costs were provided in table format together with the information on the kind of distribution applied in the PSA (gamma).	Neither the distribution parameters nor the standard deviation of the mean was given.	
Cost-Effectiveness Results	ICER per QALY gained for 1 year and for LT is provided. Furthermore, an ICE scatterplot and a WTP acceptability curve for LT were presented.	Information is provided in table and figure format and is adequate for comparison to reproduced results.	None	None
Simulation Settings	The model enables deterministic and probabilistic analyses; PSA are based on 10,000 iterations based on Monte-Carlo simulation.	The number of runs performed for the PSA was provided.	It is unclear whether the results presented in table format are based on deterministic or probabilistic analyses; it is unclear whether a half-cycle correction was performed.	As deterministic results are leading to quite different event results, it is expected that both the event simulations and the CE data presented in table format are based on PSA; we applied a half-cycle correction.
Software	All analyses were conducted in MS Excel.	None	It is difficult to test and validate complex Excel models.	We have therefore used TreeAge Pro to rebuild the model – we simulated a combined microsimulation (1,000 patients) and Monte Carlo simulation (10,000 iterations).

CHAPTER 6

Does the Structure Matter? An External Validation and Health Economic Results Comparison of Structural Obesity Event Modelling Simulation Approaches in Severe Obesity

Chapter 6 was informed by Schwander B., Kaier K., Hiligsmann M., Evers S., Nuijten M. Does the Structure Matter? An External Validation and Health Economic Results Comparison of Event Simulation Approaches in Severe Obesity. PharmacoEconomics 2022;40(9):901-15. https://doi.org/10.1007/s40273-022-01162-6

6.1 Abstract

Objectives: As obesity associated events impact long-term survival, health economic (HE) modelling is commonly applied, but modelling approaches are diverse. This research aims to compare the events simulation and the HE outcomes produced by different obesity modelling approaches.

Methods: An external validation, using the Swedish obesity subjects (SOS) study, of three main structural event modelling approaches was performed: 1) continuous body mass index (BMI) approach; 2) risk equation approach; and 3) categorical BMI-related approach. Outcomes evaluated were mortality, cardiovascular events, and type 2 diabetes for both the surgery and control arms. Concordance between modelling results and the SOS study were investigated by different state of the art measurements, and categorized by the grade of deviation observed (from grade 1-4 expressing mild, moderate, severe and very severe deviations). Furthermore, the cost per quality-adjusted life year (QALY) gained of surgery vs. controls were compared.

Results: Overall and by study arm, the risk equation approach presented the lowest average grade of deviation (overall grade 2.50; control arm 2.25; surgery arm 2.75), followed by the continuous BMI approach (overall 3.25; control 3.50; surgery 3.00) and by the categorial BMI approach (overall 3.63; control 3.50; surgery 3.75). Considering different confidence interval limits, the cost per QALY gained were fairly comparable between all structural approaches (ranging from £2,055 to £6,206 simulating a lifetime horizon).

Conclusion: None of the structural approaches provided perfect external event validation, although the risk equation approach showed the lowest overall deviations. The economic outcomes resulting from the three approaches were fairly comparable.

6.2 Introduction

Obesity is a multifactorial, chronic disorder that is usually defined as a body mass index (BMI) >30 kg/m² [1]. Recent clinical guidelines point out, that obesity can only be adequately diagnosed by BMI in combination with waist circumference (WC) [2, 3]. According to the World Health Organization, obesity is a major contributor to the global burden of chronic disease and disability [4]. In a systematic literature review of health economic obesity models, a large variation in health economic modelling approaches was identified [5].

Different modelling approaches are available to simulate obesity associated diseases and mortality on the basis of surrogate markers. Most commonly the BMI (as continuous or categorial variable) is used as central surrogate marker influenced by anti-obesity measures, but the application of widely used risk equations (e.g. UKPDS & Framingham), which include a broader set of surrogate parameters (e.g. blood pressure, HDL and total, cholesterol, triglycerides, fasting glucose, HbA1c, etc. but not necessarily BMI) to simulate a disease risk, is also quite common. These different event simulation approaches are addressed as structural (event simulation) approaches throughout the manuscript, as the approach of simulating events is usually categorized as a structural health economic modelling component, e.g. according to the Phillips checklist [6].

According to the ISPOR/SMDM modelling good research practices, trust and confidence are critical to the acceptance of health economic models [7]. According to this paper, there are two main methods for achieving this: transparency (people can see how the model is built) and validation (how well the model reproduces reality) [7]. In order to investigate and proof the validity of a health economic model, an external validation (comparing model results with real-world results) and a structural sensitivity analysis need to be performed [7]. External validation tests the model's ability to calculate actual real-world outcomes, and hence investigates the model's ability of predicting the expected development of outcomes in the real-world. By definition an external validation compares a model's results with actual event data; and involves simulating events that have occurred, such as those in a clinical trial, and examining how well the results correspond [7]. Although the obesity modelling landscape is very diverse, the published (obesity modeling) literature lacks structural sensitivity analyses and provides only limited information on external validation [8].

Up to now it has not been investigated which impact these frequently applied structural obesity-associated event simulation approaches have on the validity of event prediction and on health economic results. Consequently, the objective of this study was to assess the external validity (in terms of clinical event prediction)

of different structural obesity event simulation approaches, and to investigate their impact on the health economic results. This research could help offering a better guidance for outcome researchers, health economists and decision makers on choosing and rating the structural approaches applied in health economic obesity models.

6.3 Methods

As basis for this research, three previously replicated obesity models were used [9-13]. These models reflect three main structural obesity event simulation approaches, commonly used in health economic obesity modelling [8]. Using the clinical input data from the Swedish obesity subjects (SOS) intervention study [14, 15] (selected validation study) and health economic inputs (costs & utilities) from a recent NICE appraisal [16], model simulations were performed. On the basis of these analyses, an external validation of clinical event modelling results was performed, by comparing the simulation outcomes to the actual event data observed in the SOS-intervention study. Further, we compare key health economic outcomes between the different structural approaches. The details and methodology of these different research steps are described below.

6.3.1 External Validation Study

As external validation study, the SOS-study was selected, as this is currently the only available prospective long-term intervention study in obese subjects that has presented statistically significant improvements in mortality, incidence of type 2 diabetes, and fatal / non-fatal cardiovascular events (myocardial infarction and stroke) for obesity surgery compared to matched controls over an 18-year period [14, 15]. The SOS study reflects a population of severely obese patients that were treated with bariatric surgery intervention in the surgery arm. We have extracted the annual event rates from the published Kaplan-Meier curves, for both the surgery arm and the control arm using the GetData Graph Digitizer 2.26. This obesity-associated event data of the SOS intervention study was then compared to events simulated by three different structural event simulation approaches.

6.3.2 Description of Obesity Models

The different structural event simulation approaches are reflected in three published health economic models [9-11]. These models were selected on the basis of a previously published systematic review by our research group [8], and on the basis of minimal quality requirements based on an expert consensus [12]. All models were previously successfully replicated in TreeAge Pro (Version 2021 R1.1) on the basis of the published data [13]. For assessing the success of the model replications, we applied different criteria as defined and proposed in a recently published review on this topic [17].

Each of these health economic obesity models is reflecting another structural approach for the obesity-associated event simulation and was hence named according to the underlying structural event simulation approach as continuous BMI approach [9], risk equation approach [10] and categorical BMI approach [11]. All models are to be categorized as individual-level Markovian models without interaction and hence reflect category 2C of the revised version of Brennan's taxonomy [18].

In the model reflecting the "continuous BMI approach", the baseline risks for obesity associated events were estimated for a UK population [19-22], depending on the diabetes status and altered by relative risks for each change in BMI [23, 24]; hence each change in the BMI altered the obesity-associated event risks.

In the model reflecting the "risk equation approach" stroke and myocardial infarction were simulated via the Framingham risk equations [25-28] in non-diabetics and by the UKPDS risk equations [29-31] in diabetics. The type 2 diabetes evidence was simulated by the San Antonio Heart Study algorithm [32]; hence each change in a risk factor of these equations altered the obesity-associated event risks.

In the model reflecting the "categorical BMI approach", the risks for obesity-associated events were based on BMI-group specific risks [33-37]; namely the following BMI categories were simulated: BMI<25; BMI 25-<30; BMI 30-<35; BMI 35-<40 and BMI>40 kg/m². Accordingly the event risks were only influenced in patients changing between the BMI categories.

Mortality was simulated by disease state-specific mortality risks and by an UK life-table based background mortality in each model [38].

Simulating a severely obese population, the base risks of the "continuous BMI approach" were reviewed and adjusted (increased) for type 2 diabetes on the basis of the original publication informing this model; no adjustments were made to the "risk equation approach" and to the "categorical BMI approach", as both models have been developed flexible enough to self-adjust the risk for changing population characteristics. The details on the influence factors considered for the different event simulation approaches, as well as the applied event rates are presented in Table 6-2 in the appendix. A further calibration of the models was not performed.

6.3.3 Input Data & Model Simulations

All of those models were developed for the UK setting, and were informed for validation purposes with the population and clinical input data of the SOS intervention study. Depending on the underlying structural approach, these models were either informed by the SOS-study risk factor data (risk equation approach) or

the BMI data (continuous & categorial BMI approaches), in order to simulate the events over time. The related SOS-study data applied in the models is presented in detail in Table 6-3 (baseline values) and Table 6-4 (risk factor development over time) in the appendix.

The cost and health utility data for each model was informed by the data used in the latest UK NICE appraisal on obesity [16], which is presented in Table 6-5 of the appendix. This allows a comparison of the health economic modelling results in terms of total costs, total quality-adjusted life years (QALYs) and of the related cost-effectiveness expressed as cost per QALY gained.

Model simulations were performed for the SOS-study time horizon (18 years) and for a life-time horizon using a Monte-Carlo microsimulation approach with 10,000 iterations, which was the minimum number to achieve stable average results. Hence when simulating the same input profile consistent results were obtained.

6.3.4 External Event Validation Methodology

In the ISPOR/SMDM recommendations on results presentation and validation [7], the methods of quantitative measures to assess and present the results of an external validation are not clearly defined. However, there are recently published external validations [39, 40], that have proposed and applied different measurements (described below) for assessing the level of concordance between modelling results and validation study results, and we have used a comparable approach.

In order to allow a visual inspection of concordance, the annual cumulative events incidences corresponding to the predicted outcomes (Y axis) against those of the empirical study end-points (X axis) were plotted for each key event by model and study arm (surgery or control). In case of perfect concordance, the results would be placed on the visualized 45-degree line. If the points are located over this 45-degree line, this means overprediction of event rates by the model, and a placement below means underprediction.

Furthermore, the slope and intercept of the best-fitting linear regression line were estimated in order to quantify the visualization. In the optimal case (perfect concordance) the slope is 1 and the intercept is 0, consistent with the 45-degree line. The higher the slope is over 1 the stronger the overprediction of event rates by the model, the lower the slope is under 1 the stronger the underprediction. The figures are optimized for the comparison between the three modelling approaches within one study arm; hence the figure scaling is different for each study arm and each obesity associated key event. For an easier interpretation of findings related to the linear regression, we have categorized the level of over- and underprediction

on the basis of the variation from the optimal slope value of "1" into: mild (±25% variation from the optimal slope value "1"; grade 1), moderate (>25% and ≤±50% variation, grade 2), severe (>50% and ≤±100% variation, grade 3) and very severe (>100% variation, grade 4) over- or underprediction. In order to calculate an overall score representing the combined level of over- and underprediction, an average grade was calculated on the basis of the grade values for each endpoint.

Additionally, the R^2 coefficient was estimated; an R^2 close to 1 indicates that the relationship between the predicted and the observed data points is explained well by the linear regression line.

As the R^2 coefficient alone is not sufficient investigating whether the fitted line coincides with the identity line; an F test was performed. This test investigates whether the null hypothesis of the regression line having intercept 0 and slope 1 (perfect concordance) can be rejected. Hence the F test investigates whether there is sufficient evidence that the estimated regression line does not coincide with the identity line. Finally, the root mean squared error (RMSE) was calculated, which is zero in case of perfect concordance. Hence the smaller the RMSE value the better the model fit.

6.3.5 Comparison of Health Economic Outcomes

The health economic results are then presented in table and figure format. For each case study and study arm, the mean total costs, mean total QALYs and the related mean incremental results are presented in a summary table. Additionally, the incremental costs, utility and cost-utility results are visualized as box plots. These standard box plots reflect the 25% and 75% quartile as lower and upper end of the box; the median as line within the box; the mean as "x" within the box; and the upper and lower fence reflecting the 1.5-fold deviation of the difference between the 25% and 75% quartiles. Furthermore, to add an additional dimension of result variability, we have visualized the cost-effectiveness acceptability curves for the three approaches, in order to present the probability of being a cost-effectiveness intervention considering varying cost-effectiveness thresholds.

6.4 Results

6.4.1 Event Validation Results

Looking at the detailed external event validation results presented in Figure 6-1 to Figure 6-4 and summarized in Table 6-6 in the appendix, it could be seen that the optimal fit represented by an intercept of "0" and a slope of "1" was never observed; this is also reflected by the p-values, which are always <0.001, showing that the observed events were never exactly comparable to the identity line. The R^2 coefficient was however always quite close to 1, reflecting a good linear relationship

of the event results predicted by the models. The RMSE was always quite low but never zero, which would reflect a perfect concordance.

According to the visualization of the external event validation by event (Figure 6-1 to Figure 6-4) and according to the slope values, the following levels of over- and underprediction were observed: For the event mortality (Figure 6-1), very severe overpredictions (grade 4) were observed for the continuous and categorial BMI approaches irrespective of the study arm, whereas the risk equation approach presented a mild overprediction (grade 1) for the control arm and a moderate overprediction (grade 2) for the surgery arm.

The total cardiovascular events (Figure 6-2) presented a more diverse picture with a very severe overprediction (grade 4) observed in both study arms by the categorial BMI approach. The continuous BMI approach showed a severe overprediction (grade 3) in the control arm, but in the surgery arm a mild underprediction (grade 1) was observed. The risk equation approach showed a mild overprediction (grade 1) of total cardiovascular events in the control arm and a mild underprediction (grade 1) in the surgery arm.

The fatal cardiovascular events (Figure 6-3) were very severely overpredicted (grade 4) by all approaches irrespective of the study arm, whereas also here the risk equation approach presented the smallest overprediction, which was slightly more pronounced in the control arm than in the surgery arm.

The event diabetes (Figure 6-4) was severely underpredicted (grade 3) by the continuous BMI approach, irrespective of the study arm. For the risk equation approach a severe overprediction (grade 3) was observed in the control arm, whereas the overprediction in the surgery arm was very severe (grade 4). For the categorial BMI approach a moderate underprediction (grade 2) of diabetes was observed in the control arm and a severe underprediction (grade 3) was observed in the surgery arm.

Overall and by study arm, the risk equation approach presented the lowest average grade of over- and underprediction (overall grade 2.50; control arm 2.25; surgery arm 2.75); followed by the continuous BMI approach (overall grade 3.25; control arm 3.50; surgery arm 3,00) and by the categorial BMI approach (overall grade 3.63; control arm 3.50; surgery arm 3.75). An overview of the grades by approach, event and by study arm, as well as the average grades is provided in Table 6-7 in the appendix.

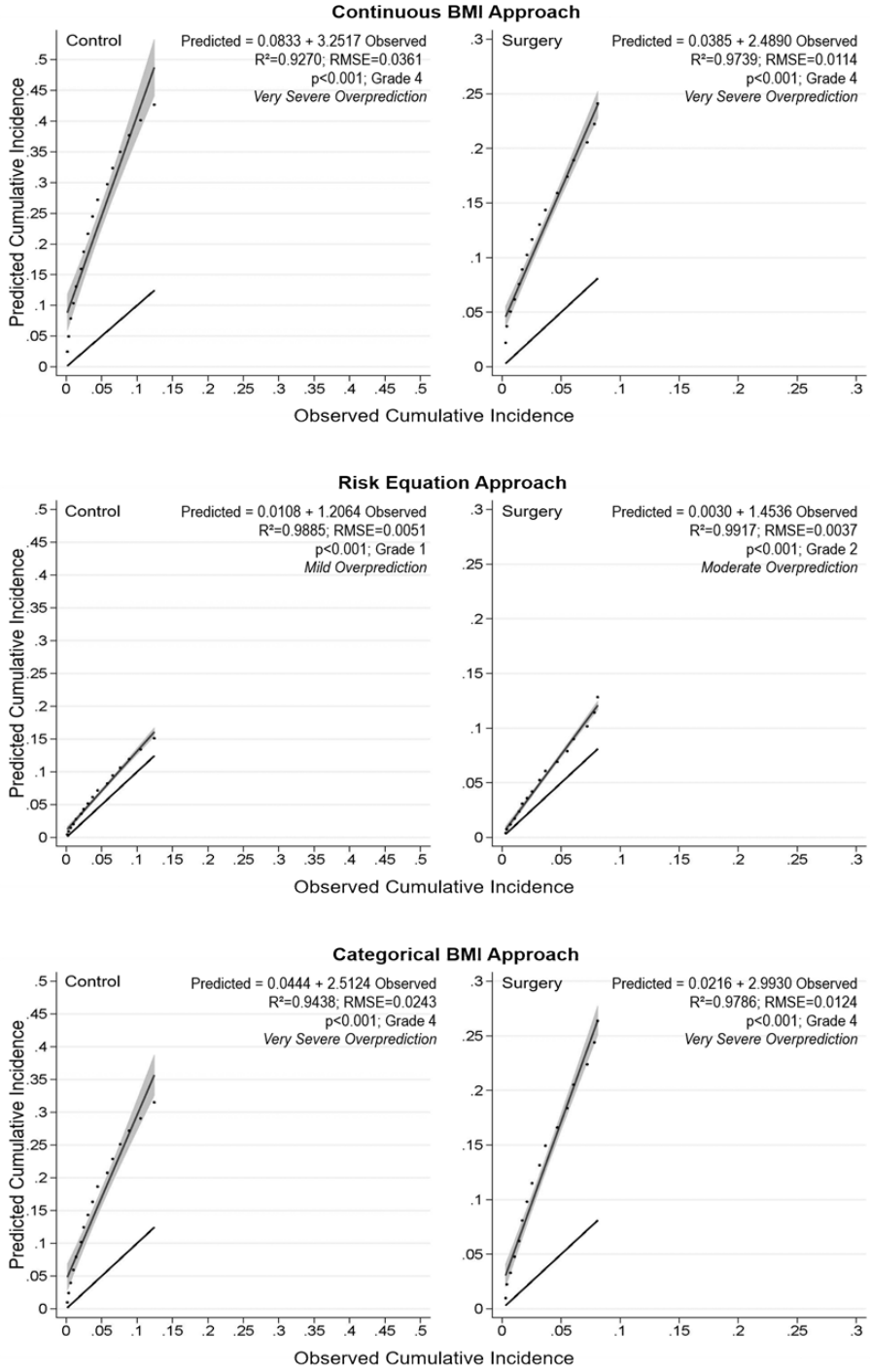


Figure 6-1. Results of the External Validation for Overall Mortality

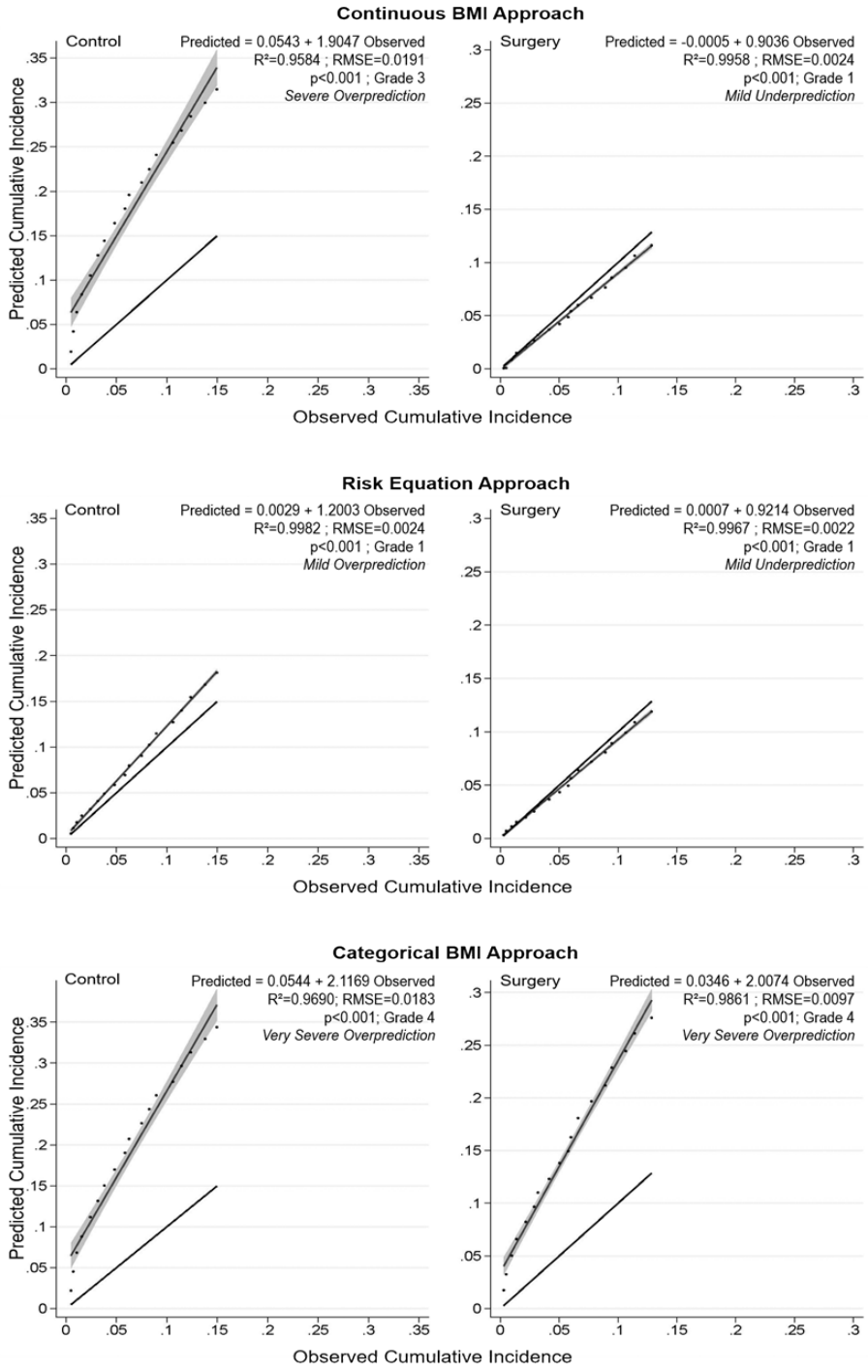


Figure 6-2. Results of the External Validation for Total Cardiovascular Events

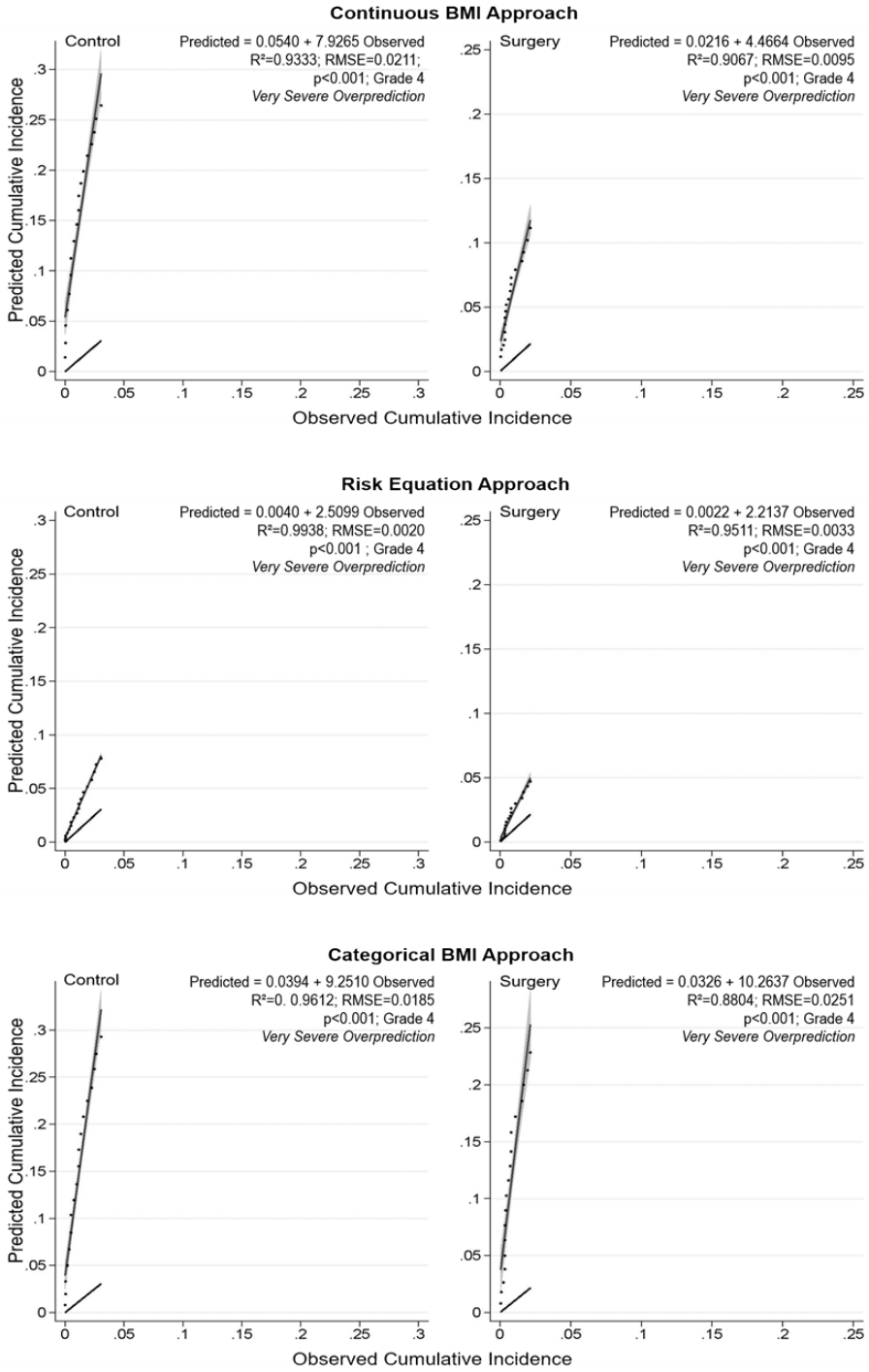


Figure 6-3. Results of the External Validation for Fatal Cardiovascular Events

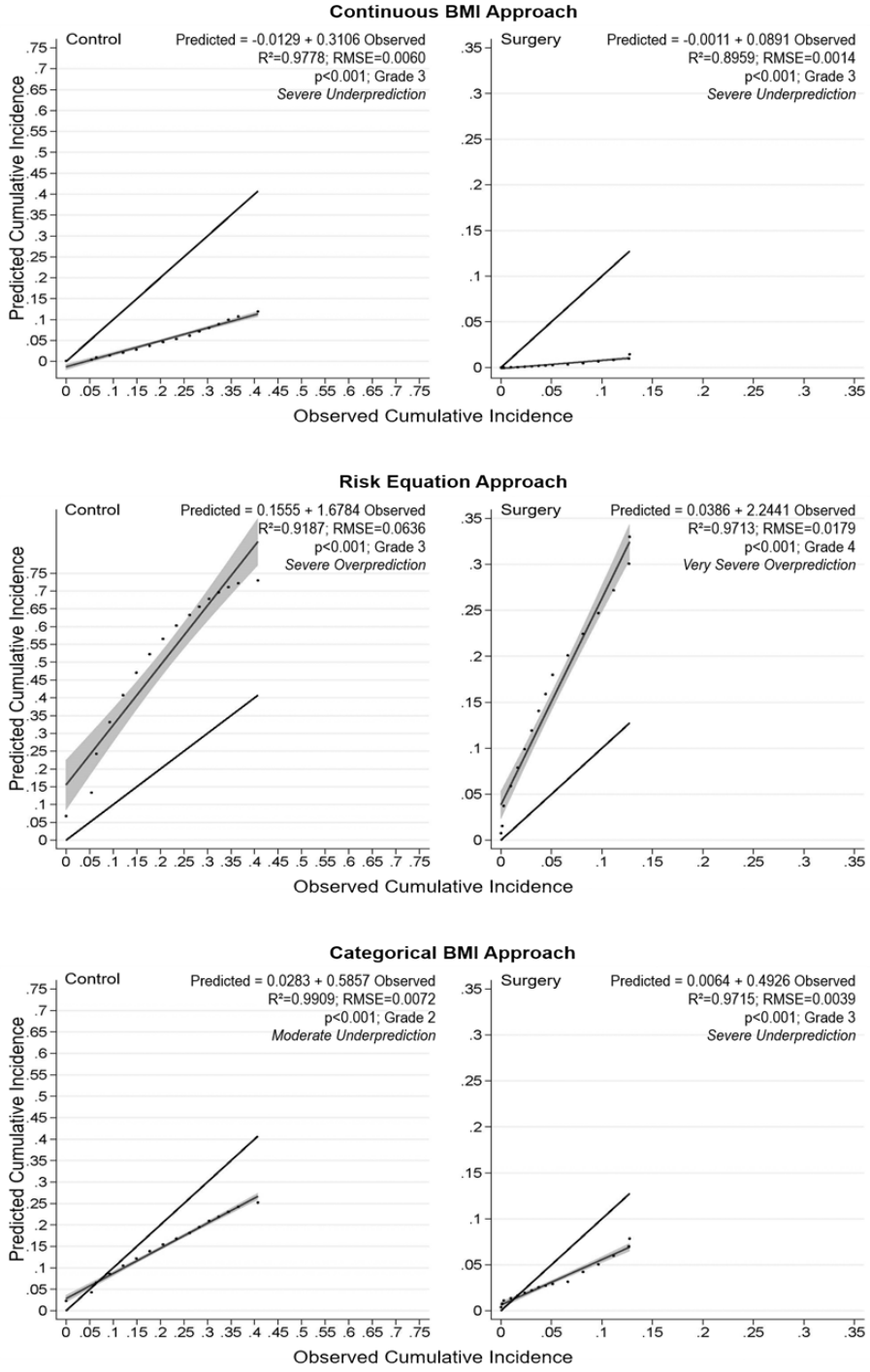


Figure 6-4. Results of the External Validation for Type 2 Diabetes

6.4.2 Health Economic Results

The health economic results, comparing the control arm vs. the surgery arm, related to the three structural approaches are presented in Table 6-1 and in Figure 6-5. Considering the mean results, presented in Table 6-1, the incremental cost-effectiveness ratio (ICER) was lowest for the continuous BMI approach, followed by the risk equation approach, and was highest for the categorial BMI approach, irrespective of the model time horizon. However, looking at the distribution of the ICER values, presented in Figure 6-5, the different confidence interval levels presented in the box plots are largely overlapping, making the ICER outcomes comparable from a statistical point of view, as even the boxes representing the 25% and 75% quantiles, and hence the 25% confidence intervals, are overlapping.

Table 6-1. Overview of Mean Health Economic Outcomes

Time Horizon	Approach	Costs (UKP)			Utility			ICER
		Surgery	Control	Incr.	Surgery	Control	Incr.	
18 Years	Continuous BMI	13,695	6,598	7,097	11.39	9.13	2.26	3,143
	Risk Equation	14,410	7,834	6,576	14.57	12.60	1.97	3,338
	Categorical BMI	14,873	4,350	10,522	10.75	9.49	1.26	8,328
Lifetime	Continuous BMI	18,126	10,162	7,965	15.37	11.49	3.88	2,055
	Risk Equation	26,354	19,637	6,717	23.00	20.00	3.00	2,241
	Categorical BMI	16,867	6,599	10,268	13.92	12.27	1.65	6,206

ICER = incremental cost-effectiveness ratio

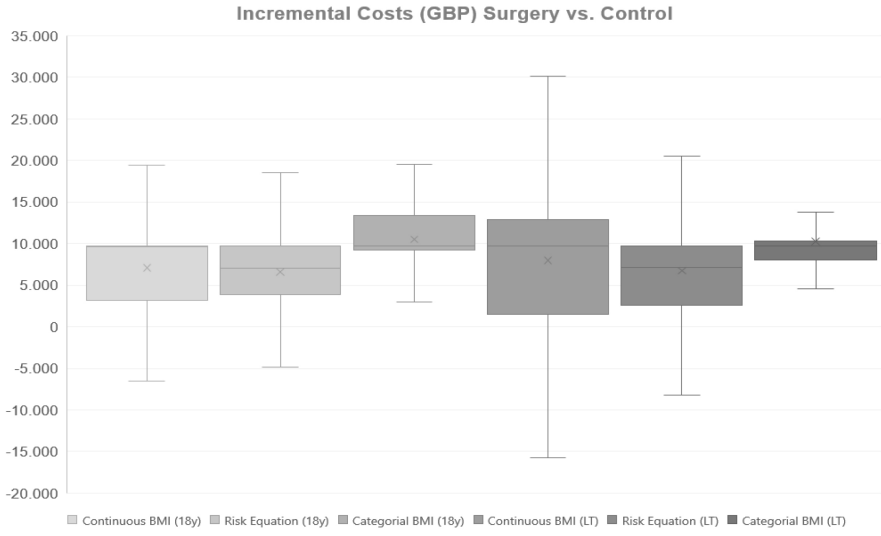
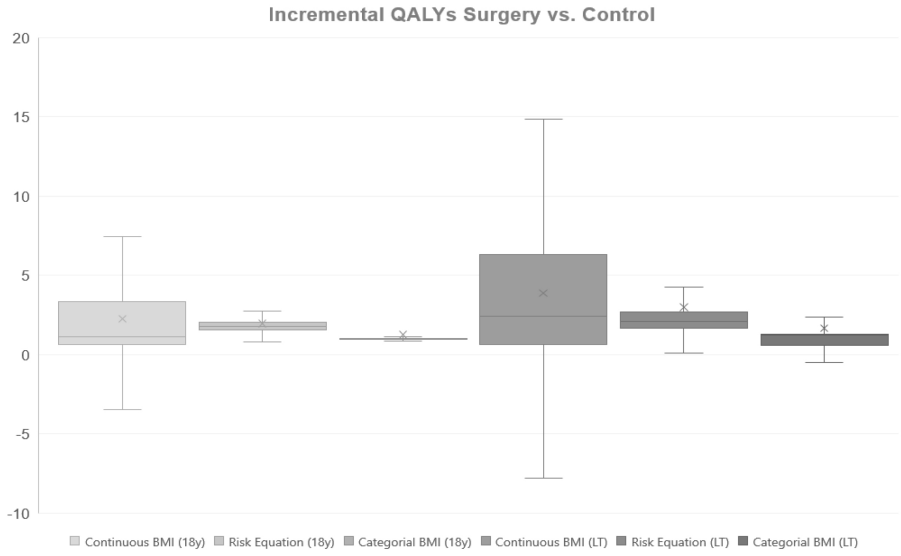


Figure 6-5. Overview of Incremental Health Economic Outcomes



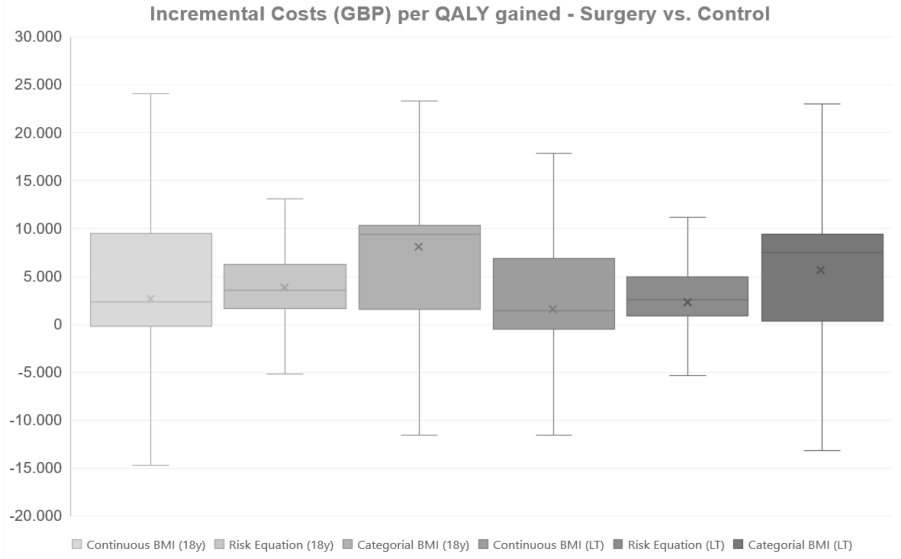


Figure 6-5. Continued

The cost-effectiveness acceptability curves are visualized in Figure 6-6 for both the study time horizon and the life-time horizon. Irrespective of the time horizon, the risk equation approach showed the highest probability of being cost-effective, followed by the continuous and the categorial BMI approaches.

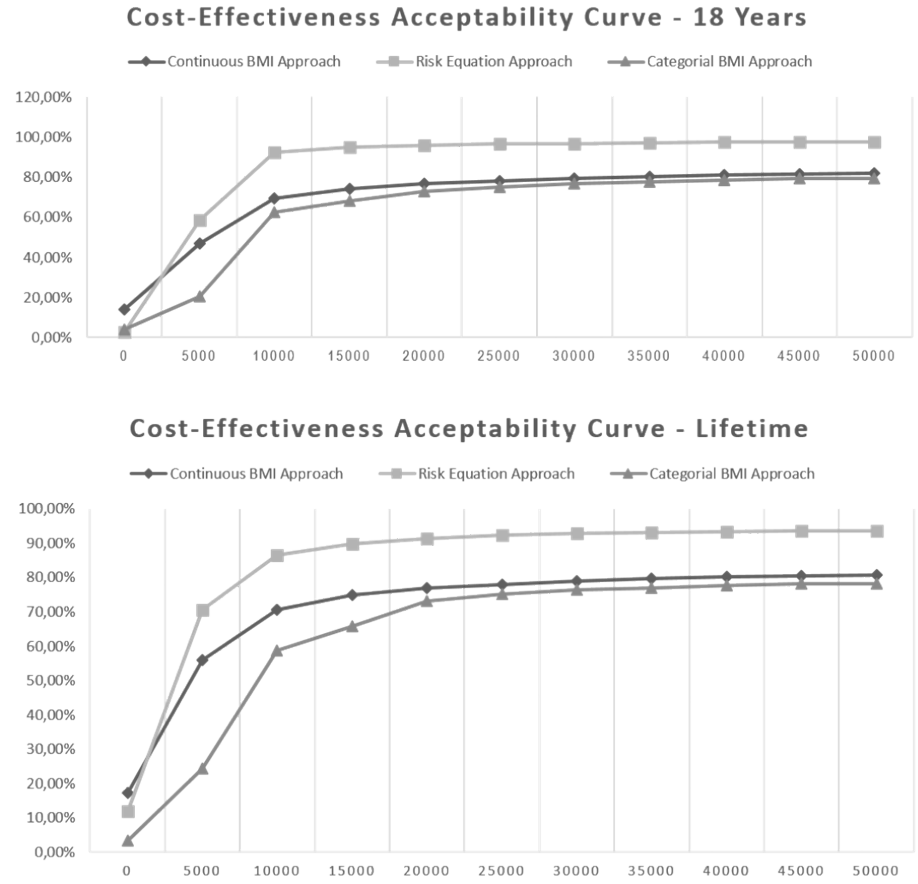


Figure 6-6. Overview of Cost-Effectiveness Acceptability Curves

6.5 Discussion

This study performed an external validation of structural event simulation approaches commonly applied in health economic obesity models (discussed first), as well as a comparison of health economic outcomes between those approaches (discussed second).

Looking at the results of the external validation, none of the investigated approaches provided an optimal event prediction, when simulating the severely obese SOS-study cohort over time. Each approach had specific findings of over- and underprediction of specific events. However, overall and by study arm, the risk equation approach showed the smallest grade of over- and underprediction, followed by the continuous BMI approach and by the categorial BMI approach.

Only with regard to the prediction of type 2 diabetes, the BMI-based approaches presented a better grade of prediction than the risk equation approach. A potential reason for this might be that the presented risk equation approach used the algorithms of the San Antonio diabetes study [32]. This southern US-based algorithm does not seem to be adequate for the prediction of type 2 diabetes in a Swedish cohort of severely obese patients, as according to our findings the type 2 diabetes incidence was severely to very severely overpredicted by the risk equation approach. This issue might be solved by selecting a Northern Europe-based T2D risk algorithm; e.g. the UK-based QDiabetes algorithm [41]; however also here the predictive quality would still needed to be investigated by an external validation.

In contrast to the risk equation approach, the external validation results of the continuous and categorial BMI approaches showed stronger deviations from the validation study. These findings are supported by ongoing discussions that not each obesity-related disease is fully and best predicted by the BMI alone [42, 43]. Obesity is a health risk defined by abnormal or excessive fast accumulation, for which WC in combination with the BMI is the best indicator. This is already reflected in recent clinical obesity definitions [2, 3], but have not yet been transferred (broadly) into health-economic modelling. The reason why many health economic models still rely only on the BMI as central risk predictor, is often based on the fact that BMI measurements are widely assessed in underlying clinical studies in obesity, whereas additional information on the development of other risk factors over time is often not available, in the desired detail, to inform more-sophisticated risk equations. Due to the shift of clinical guidelines from BMI alone to BMI plus WC it is expected, that future health economic models will also shift to BMI plus WC as central predictive variable, which might improve the predictive quality of event simulation approaches.

Previous published external validations [39, 40], that have used a comparable statistical analysis methodology, have not looked at single events or single treatment arms but on a mix of different events and treatment arms, which may have increased the likelihood of a better concordance of predicted and observed event results. On one hand the mix of different events enables overpredicted events to be balanced by underpredicted events. On the other hand, simulating and comparing the development of single events over time, as we did by including the annual cumulative event rates over time, is pronouncing observed deviations of modelling and validation study results. In contrast to our approach, other published studies have only used one point in time by study and mixed those point estimates with the results of other studies within one graph and hence within one linear regression. This approach would have also been desirable for our research, but there is a lack of long-term intervention studies in obesity that prevented the inclusion of a broader study base. For the external validation presented in this paper, we selected the SOS-study, as it is still the only prospective long-term intervention study in obesity, that have shown a significant reduction in obesity associated events and mortality, in the bariatric surgery arm [15]. These findings supported the positive reimbursement decisions on obesity surgery in many health care systems all over the world. Another prospective long-term intervention study ("Look AHEAD") has failed to prove a positive prospectively-assessed impact of diet and exercise on obesity associated events [44], which is why the external validation focused on the SOS-study.

The external validation results presented in this manuscript are based on simulations performed with three different models, that were aligned with regard to the aspects of population input parameters, BMI, risk factor development, costs, utilities and discounting. However, there are still some structural differences between the models, namely the cycle length and additional events simulated. The variation of cycle length (6 months for the categorical BMI approach, 1 month for the risk equation approach and 1 year for the categorial BMI approach) are not expected having any major impact on the event simulation results, as for all models comparable time horizons were simulated. With regard to additional events, the model reflecting the continuous BMI approach simulated additionally osteoarthritis and colorectal cancer, the later influencing survival. From both states simulated patients could move to other disease states, as long as they are not dying. Hence only patients dying from colorectal cancer have a major influence on the rates of other events; as patients dying will on one hand increase the mortality count and would reduce the rates of other events (as patients can no more move into these states).

The incidence of colorectal cancer, was about 1% in each arm simulated, with 0,5% of patients dying due to colorectal cancer, over the study time horizon which is

relevant for the external validation. Therefore, the impact of this event is rated to be minor and could neither explain the strong overprediction of mortality (indeed also the SOS-study included cancer death) nor the strong underprediction of type two diabetes observed for the continuous BMI approach. Overall the impact of still existing structural differences between the models is therefore rated as negligible.

As a limitation it has to be considered that none of the underlying structural approaches was explicitly designed for predicting obesity associated events correctly, but to investigate the health economic impact of different therapeutic measures. However, as comparable structural approaches are frequently used for various health economic evaluations in obesity, we found it justified to perform the presented external validation.

As a further limitation it needs to be considered that the obesity surgery approach, reflected in the SOS-study, is the most invasive and most efficient intervention approach in obesity, targeting especially severely obese patients (reflected by a mean BMI ≥40 mg/m² in the SOS-study population). This means that the observed variations in BMI and other risk factors, which are translating into disease risk changes and so in the number of events simulated, are strongest for surgery compared to any other less invasive obesity interventions; which also could lead to higher deviations observed in the external validation. Hence the findings of our study are referring to a very specific severely obese patient population and to a very invasive bariatric surgery approach, and may not be transferable to other less severely obese populations treated with less invasive therapy approaches.

An additional limitation to be considered is that the three underlying models were designed for a UK healthcare setting and hence for a UK population, whereas the validation study is reflecting a Swedish cohort. Although the population characteristics of the SOS-study were used to inform all simulations, this could also have had an impact on the over- and underpredictions observed in the external validation.

In addition, the external validation of health economic obesity models was found to be an exercise not frequently performed [8]; which might partly be explained by the lack of long-term intervention studies in obesity providing adequate information on the development of obesity-associated events and mortality over time. Consequently, many published external model validations used validation studies that were not reflecting an obese population. In a published systematic review on this topic, it was found that only for 14% (10 of 72) of published model-based health economic assessments in obesity, an external event validation was performed; and only for one the predictiveness and validity of the event simulation was investigated in a cohort of obese subjects [8].

Furthermore, there are no adequate published guidelines available that allow to categorize and compare the observed level of over- and underprediction. Due to this lack of published guidance we defined a classification differentiating mild, moderate, severe and very severe over- and underprediction. Although this categorization was found to be useful for our study, its value beyond the presented application in obesity, needs to be evaluated by future research.

Although we found that structure matters if considering the prediction of obesity-associated events, is this also true from a health economic outcomes perspective? We have compared the health economic key outcomes between the three structural approaches. Our main focus was set on the comparison of the incremental cost per QALY gained, comparing the surgery vs. the control arm, as this is observed as central cost-effectiveness outcome by most cost-effectiveness driven payers and decision makers. Considering this health economic key result and considering the different confidence limits presented in the box plots, there was interestingly no large difference found between the structural obesity modelling approaches. This finding might be primarily triggered by the fact that for the purpose of health economic comparison, in the presented case of surgery versus control, the incremental results are of upmost importance for the health care payers and decision makers. Hence if using comparable methods in both arms, there might be a strong difference in the single arm results (as reflected in table 6-1), but if looking at the incremental results these differences are almost "absorbed" / "no more identifiable".

However, in case that the mean ICER is to be presented and seen as the "main health economic result", the categorial event simulation approach has to be rated as the most conservative approach, as here the highest mean ICER is produced, whereas no difference was observed between the risk factor and continuous BMI approaches. Looking the cost-effectiveness acceptability curves, again the categorial BMI approach is the most conservative one, presenting the lowest probabilities of being cost-effective. The continues BMI approach presented slightly higher probabilities of being cost-effective, and the risk factor approach presented the highest probabilities of being cost-effective.

These findings are logical, as in case of the categorical BMI approach the effect size needs to be stronger for achieving to reach another BMI category and hence a related change in event risks, if compared to the risk equation and continuous BMI approaches, where each small change in risk factors or BMI is translating into a change in event risks. Hence, the hurdles for positive intervention effects are higher for the categorial BMI approach, which translates into a higher mean ICER per QALY gained and into a lower probability of being cost-effective.

To our knowledge, this is the first published research that investigated the impact of different structural event simulation approaches in obesity modelling on the event prediction and on health economic results. The reasons for the lack of previous such investigations are diverse, but research budget constraints and the intention of not putting into question an already chosen modelling approach too strongly, may be seen as two key aspects. This study provides first insights on the influence of structural event modelling approaches in obesity modelling on the accuracy of event prediction and on the health economic key outcomes. Further research is required in order to obtain a deeper understanding of the influence of structural event simulation approaches in health economic obesity modelling. In addition, it would be interesting to compare the effects of different modelling approaches on the health economic outcomes in other obese populations and in other disease areas.

6.6 Conclusions

In conclusion, this study suggests that the structure of a health economic model matters if clinical events are to be predicted most accurately, in a severely obese population. Although it was found that none of the structural approaches showed perfect external event validation results, the risk equation approach showed the smallest deviations. Combined with a careful selection of risk equations, this risk equation approach would be the method of choice for a most accurate prediction of obesity associated events.

However, if the purpose of a health economic model is purely the incremental health economic comparison, this study suggests that the structure does not matter that much, which seems positive for the credibility and comparability of health economic key results based on different structural modelling approaches. The different structural approaches provided fairly comparable probabilistic health economic results, whereas looking at the mean results (in a purely deterministic manner) and on the cost-effectiveness acceptability curves, the categorical BMI approach produced the most conservative estimates. Further research in other obese populations and other disease areas would be interesting to confirm this finding.

6.7 References

1. World Health Organization. Fact Sheet on Obesity. 2003. https://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf. Accessed 09.02.2019.
2. Wharton S, Lau DCW, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D et al. Obesity in adults: a clinical practice guideline. Canadian Medical Association Journal. 2020;192(31):E875-E91. doi:10.1503/cmaj.191707.
3. Yumuk V, Tsigos C, Fried M, Schindler K, Busetto L, Micic D et al. European Guidelines for Obesity Management in Adults. Obes Facts. 2015;8(6):402-24. doi:10.1159/000442721.
4. World Health Organization. Updated Fact Sheet on Obesity. 2018. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed 09.02.2019.
5. Schwander B, Hiligsmann M, Nuijten M, Evers S. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert review of pharmacoeconomics & outcomes research. 2016;16(5):561-70. doi:10.1080/14737167.2016.1230497.
6. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. PharmacoEconomics. 2006;24(4):355-71. doi:10.2165/00019053-200624040-00006.
7. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50. doi:10.1016/j.jval.2012.04.012.
8. Schwander B, Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert review of pharmacoeconomics & outcomes research. 2018;18(5):529-41. doi:10.1080/14737167.2018.1501680.
9. Au N, Marsden G, Mortimer D, Lorgelly PK. The cost-effectiveness of shopping to a predetermined grocery list to reduce overweight and obesity. Nutr Diabetes. 2013;3(6):e77. doi:10.1038/nutd.2013.18.
10. Caro J, Stillman O, Danel A, Getsios D, McEwan P. Cost effectiveness of rimonabant use in patients at increased cardiometabolic risk: estimates from a Markov model. Journal of Medical Economics. 2007;10(3):239-54. doi:10.3111/13696990701438629.
11. Meads DM, Hulme CT, Hall P, Hill AJ. The cost-effectiveness of primary care referral to a UK commercial weight loss programme. Clin Obes. 2014;4(6):324-32. doi:10.1111/cob.12077.
12. Schwander B, Nuijten M, Hiligsmann M, Queally M, Leidl R, Joore M et al. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy and Technology. 2020;9(3):314-22. doi:https://doi.org/10.1016/j.hlpt.2020.03.005.
13. Schwander B, Nuijten M, Evers S, Hiligsmann M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. PharmacoEconomics. 2021;39(4):433-46. doi:10.1007/s40273-021-01008-7.
14. Sjöström L, Lindroos AK, Peltonen M, Torgerson J, Bouchard C, Carlsson B et al. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. The New England journal of medicine. 2004;351(26):2683-93. doi:10.1056/NEJMoa035622.

15. Sjöström L. Review of the key results from the Swedish Obese Subjects (SOS) trial - a prospective controlled intervention study of bariatric surgery. J Intern Med. 2013;273(3):219-34. doi:10.1111/joim.12012.

16. National Institute for Health and Care Excellence. Single Technology Appraisal - Liraglutide for managing overweight and obesity [ID740]. 2020. https://www.nice.org.uk/guidance/ta664/evidence/appraisal-consultation-committee-papers-pdf-8952596893. Accessed 19.01.2022.

17. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. PharmacoEconomics. 2019;37(11):1371-81. doi:10.1007/s40273-019-00836-y.

18. Briggs ADM, Wolstenholme J, Blakely T, Scarborough P. Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. Population Health Metrics. 2016;14(1):17. doi:10.1186/s12963-016-0085-1.

19. Sutcliffe SJ, Fox KF, Wood DA, Sutcliffe A, Stock K, Wright M et al. Incidence of coronary heart disease in a health authority in London: review of a community register. Bmj. 2003;326(7379):20. doi:10.1136/bmj.326.7379.20.

20. Gibbs RG, Newson R, Lawrenson R, Greenhalgh RM, Davies AH. Diagnosis and initial management of stroke and transient ischemic attack across UK health regions from 1992 to 1996: experience of a national primary care database. Stroke. 2001;32(5):1085-90. doi:10.1161/01.str.32.5.1085.

21. González EL, Johansson S, Wallander MA, Rodríguez LA. Trends in the prevalence and incidence of diabetes in the UK: 1996-2005. J Epidemiol Community Health. 2009;63(4):332-6. doi:10.1136/jech.2008.080382.

22. Mulnier HE, Seaman HE, Raleigh VS, Soedamah-Muthu SS, Colhoun HM, Lawrenson RA et al. Risk of stroke in people with type 2 diabetes in the UK: a study using the General Practice Research Database. Diabetologia. 2006;49(12):2859-65. doi:10.1007/s00125-006-0493-z.

23. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. Jama. 1999;282(16):1523-9. doi:10.1001/jama.282.16.1523.

24. Ni Mhurchu C, Parag V, Nakamura M, Patel A, Rodgers A, Lam TH. Body mass index and risk of diabetes mellitus in the Asia-Pacific region. Asia Pac J Clin Nutr. 2006;15(2):127-33.

25. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. Am Heart J. 1991;121(1 Pt 2):293-8. doi:10.1016/0002-8703(91)90861-b.

26. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. Circulation. 1991;83(1):356-62. doi:10.1161/01.cir.83.1.356.

27. D'Agostino RB, Russell MW, Huse DM, Ellison RC, Silbershatz H, Wilson PW et al. Primary and subsequent coronary risk appraisal: new results from the Framingham study. Am Heart J. 2000;139(2 Pt 1):272-81. doi:10.1067/mhj.2000.96469.

28. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. Stroke. 1991;22(3):312-8. doi:10.1161/01.str.22.3.312.

29. Clarke PM, Gray AM, Briggs A, Farmer AJ, Fenn P, Stevens RJ et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). Diabetologia. 2004;47(10):1747-59. doi:10.1007/s00125-004-1527-z.

30. Kothari V, Stevens RJ, Adler AI, Stratton IM, Manley SE, Neil HA et al. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. Stroke. 2002;33(7):1776-81. doi:10.1161/01.str.0000020091.07144.c7.

31. Stevens RJ, Kothari V, Adler AI, Stratton IM. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). Clin Sci (Lond). 2001;101(6):671-9.

32. Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? Ann Intern Med. 2002;136(8):575-81. doi:10.7326/0003-4819-136-8-200204160-00006.

33. Myint PK, Luben RN, Wareham NJ, Bingham SA, Khaw KT. Combined effect of health behaviours and risk of first ever stroke in 20,040 men and women over 11 years' follow-up in Norfolk cohort of European Prospective Investigation of Cancer (EPIC Norfolk): prospective population study. Bmj. 2009;338:b349. doi:10.1136/bmj.b349.

34. Mohan KM, Wolfe CD, Rudd AG, Heuschmann PU, Kolominsky-Rabas PL, Grieve AP. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. Stroke. 2011;42(5):1489-94. doi:10.1161/strokeaha.110.602615.

35. Labounty TM, Gomez MJ, Achenbach S, Al-Mallah M, Berman DS, Budoff MJ et al. Body mass index and the prevalence, severity, and risk of coronary artery disease: an international multicentre study of 13,874 patients. Eur Heart J Cardiovasc Imaging. 2013;14(5):456-63. doi:10.1093/ehjci/jes179.

36. Smolina K, Wright FL, Rayner M, Goldacre MJ. Long-term survival and recurrence after acute myocardial infarction in England, 2004 to 2010. Circ Cardiovasc Qual Outcomes. 2012;5(4):532-40. doi:10.1161/circoutcomes.111.964700.

37. Warren E, Brennan A, Akehurst R. Cost-effectiveness of sibutramine in the treatment of obesity. Medical decision making : an international journal of the Society for Medical Decision Making. 2004;24(1):9-19. doi:10.1177/0272989x03261565.

38. UK Office for National Statistics. National Life Tables, United Kingdom, 1980-1982 to 2018-2020. 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables. Accessed 17.05.2022.

39. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM). PharmacoEconomics. 2017;35(3):375-96. doi:10.1007/s40273-016-0471-3.

40. Lopes S, Johansen P, Lamotte M, McEwan P, Olivieri A-V, Foos V. External Validation of the Core Obesity Model to Assess the Cost-Effectiveness of Weight Management Interventions. PharmacoEconomics. 2020;38(10):1123-33. doi:10.1007/s40273-020-00941-3.

41. Hippisley-Cox J, Coupland C. Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. Bmj. 2017;359:j5019. doi:10.1136/bmj.j5019.

42. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. Nutr Today. 2015;50(3):117-28. doi:10.1097/nt.0000000000000092.

43. Gutin I. In BMI We Trust: Reframing the Body Mass Index as a Measure of Health. Soc Theory Health. 2018;16(3):256-71. doi:10.1057/s41285-017-0055-0.

44. Look Ahead Research Group. Eight-year weight losses with an intensive lifestyle intervention: the look AHEAD study. Obesity (Silver Spring). 2014;22(1):5-13. doi:10.1002/oby.20662.

6.8 Appendix

Table 6-2. Overview of Factors Influencing the Event Simulation and of Event Rates (annual incidence per 1,000 patients)

Events	Continuous BMI Approach	Risk Equation Approach	Categorial BMI Approach
General Mortality	UK LifeTables	UK LifeTables	UK LifeTables
T2D	Base Risk (annual incidence per 1,000 patients: Male 48.6 / Female 36.6 (adjusted to severe obesity*); Relative Risk of 1 unit increase of BMI by age: age <35-44 = 1.19; age 45-69; age 1.14; age≥70 = 1.10*	Risk Calculation based on the Algorithm of the San Antonio Diabetes Study; includes age, gender, ethnicity, fasting close, systolic blood pressure, body mass index, family history of diabetes	Risk = (annual incidence per 1,000 patients) per BMI group; <25=0.6; 25.0-<30,0 =2.2; 30-<35=4.1; 35-<40=15.8; 40+=28.3
T2D Mortality	Male 0.031 / Female 0.031	General Mortality	General Mortality × 1.36
CHD	Base Risk (without T2D): Male 4.14 / Female 1.47 Base Risk (with T2D) Male 36.30 / Female 31.60 Relative Risk of 1 unit increase of BMI by age: age <35-44 = 1.12; age 45-59; age 1.10; age≥60 = 1.06*	Without T2D: Framingham Algorithm: age, gender, systolic blood pressure (SBP), smoker status and total cholesterol and HDL cholesterol (TC:HDL) ratio Patients with Diabetes: UKPDS Algorithm: age, gender, SBP, TC:HDL ratio, smoking and glycemic control.	CHD Risk <25=0.007; 25.0-<30,0 =0.010; 30-<35=0.011; 35-<40=0.016; 40+= 0,016 Secondary CHD Risk Year 1 = 0.0406 Year 2+ = 0.0203
CHD Death	Annual mortality rates Male 0.483 / Female 0.464	Month 1: Case fatality rate = 0.459 Month 2+: Case fatality = 0.005	Mortality year 1 = 0,392 Mortality year 2+ = 0,196
Stroke	Base Risk (without T2D): Male 1.42 / Female 1.42 Base Risk (with T2D) Male 10,82 / Female 13,16 Relative Risk of 1 unit increase of BMI by age: age <35-44 = 1.14; age 45-59; age 1.10; age≥60 = 1.08*	Without T2D: Framingham Algorithm: age, gender, systolic blood pressure (SBP), smoker status and total cholesterol and HDL cholesterol (TC:HDL) ratio Patients with Diabetes: UKPDS Algorithm: age, gender, SBP, TC:HDL ratio, smoking and glycemic control.	Primary Stroke Risk <25=0.0024; 25.0-<30,0 =0.0027; 30-<35=0.0029; 35-<40=0.0029; 40+=28.3 Secondary Stroke Risk Year 1 = 0.111 Year 2+ =0.036
Stroke Death	Annual mortality rates Male 0.118 / Female 0.159	Month 1: Case fatality rate = 0.149 Month 2+: Case fatality = 0.0086	Mortality year 1 = 0,28 Mortality year 2+ = 0,14

*original rates = 1 unit BMI decrease calculated by male 48.6 / female 36.6; **reversed RR 1 unit BMI decrease calculated by reversed relative risk)

Table 6-3. Population Characteristics of the SOS-Study used to inform the model simulations

Population Characteristics	Control & Surgery Arms
Age (years)	47.0
Height (m)	1.7
Weight (kg)	118.0
Total Cholesterol (mmol/liter)	5.8
Fasting Glucose (mmol/liter)	5.3
HDL Cholesterol (mmol/liter)	1.2
Family history T2D (%)	2.7%
Systolic Blood Pressure (mmHg)	140.0
Triglycerides (mmol/liter)	2.1
Males (%)	30.0%
Smoker (%)	22.0%
HbA1c (%)	8.1%
Previous T2D (%)	15.3%
Previous MI (%)	1.5%
Previous Stroke (%)	0.9%
Disease-Free (%)	82.4%

Table 6-4. Risk Factor Development (as Percentage Change Relative to Baseline) in the SOS-Study over Time by Study Arm

Study Arm	Control Arm		Surgery Arm	
Risk Factor / Observation Period	Year 2	Year 10*	Year 2	Year 10*
Weight (kg)	0.10%	1.60%	-23.40%	-16.10%
Total Cholesterol (mmol/liter)	0.10%	6.00%	-2.90%	-5.40%
Fasting Glucose (mmol/liter)	5.10%	18.70%	-13.60%	-2.50%
HDL Cholesterol (mmol/liter)	3.50%	10.80%	22.00%	24.00%
Systolic Blood Pressure (mmHg)	0.50%	4.40%	-4.40%	0.50%
Triglycerides (mmol/liter)	6.30%	2.20%	-27.20%	-16.30%
HbA1C (%)	-10.00%	-5.00%	-30.12%	-15.06%

*in the model it was assumed that after the 10-year time horizon the values slowly started to develop into the direction of the baseline values, reaching the baseline values after a 20 years period and hen staying constant over lifetime

Table 6-5. UK Cost and Utility Input Data Informing Health Economic Model Analysis

Cost Item	**Costs**	**Comment**
Surgery Costs	9,753	Mean Costs of Bariatric Surgery
T2D per Year	372	Monitoring (prediabetes) plus T2D medication costs
MI acute fatal costs	2,265	Fatal acute MI costs
MI acute non-fatal costs	5,788	Fist year cost plus acute cost
ACS subsequent years	223	Used for MI/AP in subsequent years
AP acute fatal costs	1,466	Fatal acute AP costs
AP acute non-fatal costs	2,039	Fist year cost plus acute cost
Stoke acute fatal costs	4,351	Fatal acute stroke costs
Stroke non-fatal acute	10,471	Fist year cost plus acute cost
TIA acute event costs	1,945	
Stroke Subsequent Year s	2,815	Used for Stroke / TIA subsequent years
Base Utilities by Age*	Utility	
Age < 25	0.938	Base utility by age
Age ≥ 25 and < 35	0.897	Base utility by age
Age ≥ 35 and < 45	0.856	Base utility by age
Age ≥ 45 and < 55	0.856	Base utility by age
Age ≥ 55 and < 65	0.818	Base utility by age
Age ≥ 65 and < 75	0.779	Base utility by age
Age ≥ 75	0.710	Base utility by age
Female	0.014	Additional base utility for females
Health Utilities Decrements	Utility	**Comment**
T2D	-0.0374	Ongoing Disutility
Post-Acute Coronary Syndrome	-0.0368	Ongoing Disutility
Post Stroke	-0.0349	Ongoing Disutility
Bariatric Surgery	-0.1840	Acute Disutility
Acute MI/AP	-0.0630	Acute Disutility
Acute Stroke	-0.1170	Acute Disutility
Acute TIA	-0.0330	Acute Disutility

Table 6-6. Overview of External Validation Results

Outcome	SOS-Study vs. Event Approach	Study Arm	N	Intercept	Intercept (SE)	Slope	Slope (SE)	Grade	R^2	P-Value (F)	RMSE
Mortality	Continuous BMI	Control	16	0.0833	0.0141	3.2517	0.2438	4	0.927	<0.001	0.0361
		Surgery	16	0.0385	0.0048	2.4890	0.1090	4	0.974	<0.001	0.0114
	Risk Equation	Control	16	0.0108	0.0020	1.2064	0.0348	1	0.989	<0.001	0.0051
		Surgery	16	0.0030	0.0016	1.4536	0.0356	2	0.992	<0.001	0.0037
	Categorial BMI	Control	16	0.0444	0.0095	2.5124	0.1639	4	0.944	<0.001	0.0243
		Surgery	16	0.0216	0.0052	2.9930	0.1183	4	0.979	<0.001	0.0124
Total CVE	Continuous BMI	Control	18	0.0543	0.0079	1.9047	0.0992	3	0.958	<0.001	0.0191
		Surgery	18	-0.0005	0.0010	0.9036	0.0147	1	0.996	<0.001	0.0024
	Risk Equation	Control	18	0.0029	0.0010	1.2003	0.0127	1	0.998	<0.001	0.0024
		Surgery	18	0.0007	0.0009	0.9214	0.0133	1	0.997	<0.001	0.0022
	Categorial BMI	Control	18	0.0544	0.0075	2.1169	0.0946	4	0.969	<0.001	0.0183
		Surgery	18	0.0346	0.0040	2.0074	0.0597	4	0.986	<0.001	0.0097
Fatal CVE	Continuous BMI	Control	18	0.0540	0.0079	7.9265	0.5299	4	0.933	<0.001	0.0211
		Surgery	18	0.0216	0.0036	4.4664	0.3581	4	0.907	<0.001	0.0095
	Risk Equation	Control	18	0.0040	0.0007	2.5099	0.0496	4	0.994	<0.001	0.0020
		Surgery	18	0.0022	0.0012	2.2137	0.1255	4	0.951	<0.001	0.0033
	Categorial BMI	Control	18	0.0394	0.0069	9.2510	0.4645	4	0.961	<0.001	0.0185
		Surgery	18	0.0326	0.0094	10.2637	0.9455	4	0.880	<0.001	0.0251
Diabetes	Continuous BMI	Control	16	-0.0129	0.0030	0.3106	0.0125	3	0.978	<0.001	0.0060
		Surgery	16	-0.0011	0.0005	0.0891	0.0081	3	0.896	<0.001	0.0014
	Risk Equation	Control	16	0.1555	0.0324	1.6784	0.1334	3	0.919	<0.001	0.0636
		Surgery	16	0.0386	0.0070	2.2441	0.1031	4	0.971	<0.001	0.0179
	Categorial BMI	Control	16	0.0283	0.0036	0.5857	0.0150	2	0.991	<0.001	0.0072
		Surgery	16	0.0064	0.0015	0.4926	0.0226	3	0.972	<0.001	0.0039

Table 6-7. Overview of Grade of Deviation observed in the External Validation

Approach	Event	Control Arm	Surgery Arm	Both Arms
Continuous BMI	Mortality	4.00	4.00	4.00
	Total CVE	3.00	1.00	2.00
	Fatal CVE	4.00	4.00	4.00
	T2D	3.00	3.00	3.00
	Overall	**3.50**	**3.00**	**3.25**
Risk Equation	Mortality	1.00	2.00	1.50
	Total CVE	1.00	1.00	1.00
	Fatal CVE	4.00	4.00	4.00
	T2D	3.00	4.00	3.50
	Overall	**2.25**	**2.75**	**2.50**
Categorial BMI	Mortality	4.00	4.00	4.00
	Total CVE	4.00	4.00	4.00
	Fatal CVE	4.00	4.00	4.00
	T2D	2.00	3.00	2.50
	Overall	**3.50**	**3.75**	**3.63**

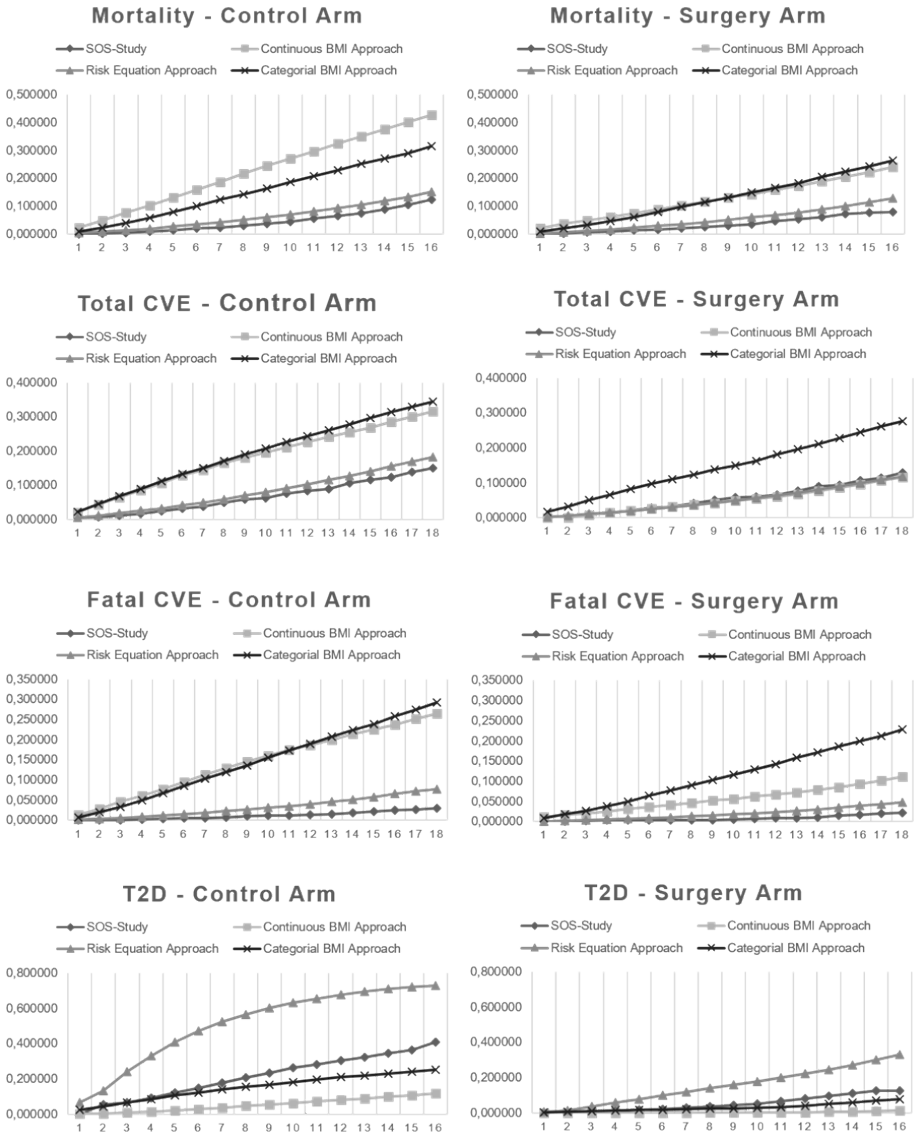


Figure 6-7. Overview of Cumulative Event Rates over Time

CHAPTER 7

General Discussion

7 General Discussion

Health economic research often has practical implications, namely informing decision makers on the most efficient way of allocating scarce resources within a given healthcare system. Decision makers in the healthcare setting can be payers, politicians, advisors, clinicians etc. or other central member of decision-making boards. Irrespective of a decision maker's background and specific perspective, they need to rely on the valid information provided by researchers. Health economists especially occupy a crucial position for informing such decisions, as they combine and synthesize information from different disciplines and sources, in order to simulate the clinical and economic consequences of such decisions, for individuals and for society. Therefore, confidence and trust in health economic research, and hence the research integrity, is crucial to ensure the best allocation of scarce resources in order to improve the health of individuals and society as a whole. One central basis for informing decision makers is producing health economic models which synthesize the clinical and economic consequences of a (usually innovative) healthcare intervention, and compare it to alternative routes of action (usually to current standards). Major health economic associations, namely the "International Society for Pharmacoeconomics and Outcomes Research" (ISPOR) and the "Society for Medical Decision Making" (SMDM), point out that trust and confidence are critical to the acceptance of such health economic models and their use for decision making [1, 2].

The research presented in this dissertation aims to increase this trust and confidence, by focusing on health economic models in obesity. Especially in chronic conditions, such as obesity, health economic modeling is required to translate short term surrogate parameters (e.g. a change in the BMI) into long-term patient-relevant endpoints (e.g. cardiovascular events, and related survival). As a result, the question is: which modeling approach produces the most reliable results, considering both the clinical and the economic consequences? With regard to this question the model structure is a key aspect, as this influences how surrogate parameters are translated into patient-relevant endpoints. These patient-relevant endpoints have a significant impact on survival, quality of life and costs, and hence on all central outcomes of a health economic assessment.

7.1 Main Objectives and Main Results

This dissertation intends to study the methodology and the validity of published health economic models in the field of obesity. In particular it evaluates, replicates and validates the current structural modelling landscape in obesity, with an

emphasis on commonly applied obesity-associated event simulation approaches. This research aims to increase trust and confidence in the selection and interpretation of results related to a specific methodological approach, used as a basis for health economic models in obesity.

Accordingly, this dissertation identified and evaluated the different methods used for such health economic obesity models, investigated how accurately these models predict the (clinical) reality, and studied the impact of the modelling methodology on the health economic model outcomes. Furthermore, it tested whether the information usually published for such health economic models allows replication of the model and reproduction of the results, a criterion for the quality of reporting of scientific experiments irrespective of the research field [3].

As presented in this dissertation, it was found that in the context of obesity (almost) every research team builds its own obesity model; this is reflected by the huge diversity of obesity modelling approaches (chapter 2) [4]. This makes it difficult to compare model outcomes potentially affecting the validity of the study, as the structural and methodological differences could have a major impact on the modelling results, as observed in other disease areas [5-7]. Furthermore, it was found, that one key limitation of these models is the lack of published external validation results which could provide valuable information on the predictiveness, and therefore on the quality, of their event simulation approaches (chapter 3) [8]. Hence it is unclear whether decision makers in the healthcare setting can rely on (trust) the results of those models. The different modeling approaches were presented and discussed with health economic experts in order to create best practice recommendations on key structural approaches for health economic obesity models (chapter 4) [9]. Using these expert recommendations, high quality health economic obesity models were selected and replicated (chapter 5) [10]. Here it was found that small changes to existing reporting criteria have the potential to increase the transparency of model reporting and may increase the reproduction success of health economic modelling results [10]. This may subsequently increase the transparency and acceptance of health economic modelling studies. Finally (in chapter 6), it was found that in a severely obese population, the structure of a health economic model matters if clinical events are to be predicted most accurately [11]. However, if the purpose of a health economic model is purely the incremental health economic comparison, this study suggests that the differences in structure are of less consequence, as incremental health economic results are fairly comparable [11].

7.2 Contribution to Health Economic Research and to Scientific Debate

Prior to our systematic review, there were several existing systematic reviews [12-20] on health economic assessments in obesity available. These focused either on specific therapeutic interventions [12, 14-16], on specific prevention measures [13, 18] or on specific populations [20, 17, 19]. On the basis of these systematic reviews it was already known that methodological differences can be expected to impact the modelling results [20]. Furthermore, it was pointed out that there is a need for future research to enhance reporting transparency, and to investigate external validity [20].

In addition to these existing systematic reviews on health economic assessments in obesity, the systematic review presented in chapter 2 did not focus on a specific intervention or population, and also included preventive approaches. This broadened the spectrum of included studies, in order to obtain a general picture on the health economic modelling landscape, in the field of obesity. It was found that the modelling landscape was very diverse also in comparison to other chronic conditions (e.g. in comparison to type 2 diabetes models [21] as presented in chapter 2 [4]), which justified and informed further research.

On the basis of this systematic review, we extracted and categorized the structural approaches of translating surrogate endpoints into patient-relevant endpoints (chapter 3) [8]. This was performed for the first time ever at this level of detail, which enabled us to define and differentiate the available key event simulation approaches. Besides the diversity of approaches, it was additionally found that there was a lack of external validations of these structural approaches, which raised the question of their validity [8].

This identified diversity of structural modelling approaches also highlighted the need to form an expert panel, in order to discuss and rate the various approaches, and to define structural quality criteria for the health economic obesity models (chapter 4) [9].

Using these quality criteria, high quality health economic obesity models were selected for replication (chapter 5) [10]. In this research step, we reproduced four different obesity models, reflecting the key approaches usually applied in the obesity-associated event simulation. This was done for the first time ever with regards to obesity. Furthermore, we tested different proposed criteria intended to define the replication success, and suggested a combination of different criteria on the basis of our findings. This was the first published application of these criteria, and hence forms the first basis for scientific discussion of how to define a model replication success in health economics. In addition, we have highlighted specific

needs for updating the original CHEERS reporting criteria [22] to enhance model reproduction. These highlighted aspects were partially considered in the newly published CHEERS II criteria [23].

The successful replicated models were then used to perform an external validation of obesity associated events, and a comparison of health economic key results (chapter 6) [11]; this research was performed in the context of obesity for the first time. Here it was found that the event simulation approach affects the prediction of clinical events, but that the influence on the incremental health economic results was limited. These findings are of great value for health economists and decision makers, as they highlight the strengths and weaknesses of current obesity models relating to event prediction, but provide confidence in the incremental cost-effectiveness outcomes, that were fairly comparable between the approaches.

Hence, each of the described research steps provided a specific contribution to the health economics literature. Although the focus of our research was obesity, the applied research approach and research methods are not limited to obesity. In all situations, where surrogate parameters are to be translated into patient relevant endpoints, the presented research would be indicated, which is mainly the case in chronic diseases and conditions, in which health consequences of interventions are observed over a long period of time.

7.3 Methodological Key Considerations / Reflections

7.3.1 Systematic Review

In order to obtain clear insights on the quantity and methodology of health economic assessments (HEAs) in obesity, a systematic review was performed as the basis for our research [4]. As a central aim of our research was to inform and support future HEAs of interventions in the prevention and treatment of obesity, we decided to focus the systematic review on available full HEAs (eligibility criteria). Due to this, we might have excluded obesity modelling approaches applied in pure epidemiologic or clinical obesity models. Such epidemiological disease models are expected to be more complex, as they may try to provide an exact simulation of the disease progress, in comparison to the HEA models, which are usually designed to provide insights to a specific decision problem. According to standard modelling guidelines, the HEA models are usually complex enough to adequately reflect the decision problem, but as simple as possible to allow for better transparency and easier validation [24]. Hence the clinical HEA modelling approaches for obesity are expected to be a simplification of reality, in comparison to more complex epidemiological disease models.

However, each HEA modeling approach identified tries to estimate the rates of specific obesity associated diseases, that then impact survival, quality of life and related costs. Therefore, this clinical modelling structure has a central impact on the key outcomes of the health economic assessment, and needs to work within acceptable accuracy parameters. This was the rationale for performing the further research steps presented in this dissertation, including the external validation of obesity-associated events and the structural sensitivity analysis of obesity-associated events.

One limitation of our systematic review is, that it does not reflect research published after the date of the systematic review execution (end of May 2015). In the meantime, the evidence body has constantly increased and further HEA obesity models have been published [25-34]. We have identified at least one performing an external validation of obesity associated events, which was according to the authors also driven by our research findings [34]. Hence it could be interesting to update our research in the future, in order to determine its impact on the HEA modelling in obesity.

7.3.2 Expert Panel

Due to the large methodological variations identified during the systematic review, an expert panel was formed to discuss the identified obesity modeling approaches. The aim was to define (health economic) expert recommendations on key structural approaches for a HEA obesity model.

Comparable to the approach used for the literature research, the expert panel was formed by health economists, which is a limitation related to the composition of the panel. The rationale for selecting health economists was that modelling is primarily driven by this discipline, but as a consequence it was not possible to obtain a clear expert rating on purely clinical aspects, such as the obesity associated event selection.

If we had involved specialized epidemiologists and / or clinicians, the discussion might have moved more in the direction of which events are clearly obesity associated, such as myocardial infarction, stroke and type 2 diabetes. Furthermore, the discussion might have focused on the best-fitting risk factors or surrogate parameters to estimate the risk of these obesity associated diseases. In order to close this gap, we have considered the latest related published literature on these topics, in the discussion part of chapter 4.

Due to the fact that there were no health economic recommendations on obesity modelling published previously, our research began with web-meeting based expert interviews, in order to obtain the individual opinions relating to various key

aspects identified by the systematic review. Afterwards, in a personal round-table meeting with these health economic experts, we then discussed the condensed individual answers and tried to obtain consensus decisions. In the expert interviews and in the expert panel we only used basic quantitative methods in order to obtain an expert rating and an expert consensus, as the style of questions were not designed to apply to more advanced methods (e.g. discrete choice experiments or the Delphi method).

7.3.3 Model Replication

Decision-analytic models have become an essential tool used to inform health technology assessments [35]. They tend to be complex and are rarely fully validated against external data; yet, use of their forecasts requires trust in their accuracy and lack of bias [35]. Thus, decision makers and other stakeholders want to be able to review their structure, inputs, and assumptions fully, which necessitates that these models be available and transparent enough to permit adequate review [35].

The best measure for transparency of reporting are the method replicability and reproduction of results, which are common criteria of adequate research reporting to assure scientific integrity. A recently published systematic review, presented in 2019 by McManus et al., investigated published definitions for replicability in health economics and other disciplines, and produced a set of potential definitions for result reproduction success in health economic models [3].

The methodological challenge in this context was that the different criteria proposed were applied to health economic models for the first time, as presented in chapter 5. In order to rate the result reproduction success, we have proposed a combination of different broader and more specific criteria as the most adequate approach. Although this proposed combination does not ensure identical results, it ensures that the cost-effectiveness conclusion is identical, that the deviation in single components is acceptable (<5%), and that the incremental cost-effectiveness results are fairly comparable. However, future research is required to better investigate and define the criteria for replication success in health economics.

7.3.4 External Validation

According to the ISPOR/SMDM modelling good research practices, trust and confidence are critical to the acceptance of health economic models [1]; these aspects form the preconditions for health economic research integrity.

According to this paper, there are two main methods to achieve this: transparency (people can see how the model is built), and validation (how well the model reproduces reality) [1]. In order to investigate and proof the validity of a health

economic model, an external validation (comparing model results with real-world results) and a structural sensitivity analysis need to be performed [1].

However, the ISPOR/SMDM recommendations do not clearly define the methods of quantitative measures to assess and present the results of an external validation [1]. Thus, we have used the approach suggested in recently published external validations [36, 34], in which the authors have proposed and applied different measurements for assessing the level of concordance between modelling results and validation study results.

In contrast to our research, these previous published external validations [36, 34], which used a comparable statistical analysis methodology, have not looked at single events or single treatment arms but at a mix of different events and treatment arms, which may have increased the likelihood of a better concordance of predicted and observed event results. On one hand the mix of different events enables overpredicted events to be balanced by underpredicted events. On the other, simulating and comparing the development of single events over time, as we did by including the annual cumulative event rates over time, is pronouncing observed deviations of modelling and validation study results, as each point in time informs and influences the linear regression. In contrast to our approach, other published studies have only used one point in time for each study and mixed those point estimates with the results of other studies within one graph and hence within one linear regression. This approach could have been useful to our research, but there is a lack of long-term intervention studies in obesity that prevented the inclusion of a broader study base. With the inclusion of more studies, it would be expected to have a broader variability of results, and therefore a broader confidence area around the linear regression. This could have supported better predictive results of the different event simulation approaches.

For the external validation presented in this dissertation, we selected the SOS-study, as it is still the only prospective long-term intervention study in obesity, which has shown a significant reduction in obesity associated events and mortality, in the bariatric surgery arm [37]. The lack of long-term intervention studies in obesity is one key limitation for performing external validations of health economic decision models in this area. Furthermore, external data availability could be an additional hurdle. Also, in our case we had no access to the full data of the SOS-study. Hence, we had to rely on the published data, which was detailed enough to perform an external validation. However, not every study might be published at the desired level of detail.

Consequently, many published external model validations used validation studies that did not reflect an obese population. In our systematic review on this topic, it

was found that an external event validation was performed for only 14% (10 of 72) of published model-based health economic assessments in obesity; and only one assessment investigated the predictiveness and validity of the event simulation in a cohort of obese subjects [8].

7.4 Implications and Recommendations for Future Research

7.4.1 Systematic Review

As our systematic review found that modelling approaches in obesity are very diverse and largely lack external validations, it would be interesting to see, which impact our findings might have on future health economic assessments in obesity. The lack of external validations in particular could be a field that may potentially improve by highlighting this issue; but it is also conceivable that larger research teams will begin to develop validated obesity models, that will then be offered to other researches and/or the industry. This situation is present in the field of type 2 diabetes, where modeling approaches are less diverse [21], and the key diabetes models are cross validated by the ongoing Mount Hood challenges [38-40]. Such key models, e.g. the CORE diabetes model, are recommended to be used for (industry) submissions to the UK National Institute for Health and Care Excellence. In the context of obesity, recently the CORE obesity model was developed [34], externally validated and published, which could initiate the development of a situation as observed in type 2 diabetes. This development might, alongside other factors, have been kick-started by our research, as our findings were cited as argumentation for the need of validated obesity models. Further, it would be of interest to see similar work in other disease areas and observe any similarities or differences.

7.4.2 Expert Panel

One limitation of our expert panel was the focus on health economists. However, even though the focus was set on one discipline, it was sometimes difficult to achieve consensus on specific topics; this might have been even more difficult if panelists from different disciplines had been invited.

One suitable approach for future panels could hence be to hold separate panels for different disciplines (health economists, clinicians, epidemiologists) focusing on different key questions related to the specific discipline. This approach might avoid an expected interdisciplinary disagreement, but conversely might also avoid possible interdisciplinary agreement on some aspects. To potentially combine the strengths and weaknesses of both approaches, a sequential conduct of separate panels for each discipline, followed by an interdisciplinary panel, could be an interesting approach, as it would allow us to investigate which discipline-related

consensus and which interdisciplinary consensus might be achieved, related to the clinical and structural methodology for health economic obesity modelling.

7.4.3 Model Replication

Model replication and reproducibility of results is strongly connected to quality of reporting, which is why the grade of compliance to the updated CHEERS II reporting criteria [23] is assumed to have a strong connection to enabling a model replication. In addition to the previous CHEERS version [22], it has become clear in the updated example CHEERS II reporting tables, that the details and parameters of probability distributions are to be reported. This enables a 1:1 replication of the probability distributions, which was not possible for our replications, reported in chapter 5, as the related details were missing. However, the latest CHEERS II update did not request a presentation of clinical events (best over time), as we suggested, to enhance the result reproduction.

Further research is required in order to further investigate the needs for model replicability and result reproduction and the correlation between the CHEERS II information/score and the model replication success.

Furthermore, as our replication study (presented in chapter 5) was to our knowledge the first application of replication success criteria proposed by McManus et al. [3] to health economic models, further research and scientific dialogue is required to investigate and define how to best rate the success of a health economic model replication.

7.4.4 External Validation

The main issue for performing external validations of health economic models, in the context of obesity, is the lack of long-term (intervention) studies. Other research teams have used long-term studies performed in non-obese populations, but the question remains whether an external validation based on non-obese cohorts is sufficient to investigate the prognostic validity of a model in the context of obesity. Hence, there is a need for additional long-term (intervention) studies in order to obtain a better understanding of the influence of the obesity status on the development and prevention of obesity-associated events. It might be interesting for further research to investigate if and how observational long-term studies and real-world evidence related to obese populations could be used for the external validation of health economic obesity models.

7.4.5 COVID-19 and Obesity

COVID-19 may be an interesting factor to consider in future obesity modelling. First studies indicate that the COVID-19 pandemic led to an increase in the prevalence of obesity, as due to lock-down and related contact restrictions, physical activity

and healthy eating habits were negatively impacted [41-44]. Furthermore, obesity was found to be an important risk factor for a severe COVID-19 course of disease, which increases the risk for COVID-19 related hospitalization, intensive care unit stay, invasive mechanical ventilation and death [45-47].

Hence treating or preventing obesity could also have a major impact on the severity of consequences of a COVID-19 infection. Whether this should be captured in future obesity modelling frameworks depends on the development of the characteristics of COVID-19 virus variants as well as on the future infection dynamics.

7.4.6 Scientific and Social Impact

A detailed reflection on the scientific and social impact of the research presented in this dissertation is presented in the "Impact" chapter in the annex of this document (please refer to page 172 ff).

7.5 Conclusions

As presented above, this thesis provided valuable insights on the systematic evaluation, replication and validation of structural health economic modelling approaches in the field of obesity. In particular it evaluated, replicated and validated the current structural modelling landscape in obesity, with an emphasis on commonly applied obesity-associated event simulation approaches.

This research was able to identify some important aspects related to the health economic modelling methodology in general, and key aspects specifically related to the field of obesity. Besides highlighting and investigating the aspects related to research integrity of published health economic models, our research formed a basis for evaluating the strengths and weaknesses of different structural event simulation approaches. Furthermore, we defined valuable future areas of research to further enhance trust and confidence in health economic modelling, especially in the field of obesity.

7.6 References

1. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50. doi:10.1016/j.jval.2012.04.012.
2. Ghabri S, Stevenson M, Möller J, Caro JJ. Trusting the Results of Model-Based Economic Analyses: Is there a Pragmatic Validation Solution? PharmacoEconomics. 2019;37(1):1-6. doi:10.1007/s40273-018-0711-9.
3. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. PharmacoEconomics. 2019;37(11):1371-81. doi:10.1007/s40273-019-00836-y.
4. Schwander B, Hiligsmann M, Nuijten M, Evers S. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert review of pharmacoeconomics & outcomes research. 2016;16(5):561-70. doi:10.1080/14737167.2016.1230497.
5. Brisson M, Edmunds WJ. Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. Medical decision making : an international journal of the Society for Medical Decision Making. 2006;26(5):434-46. doi:10.1177/0272989x06290485.
6. Moriña D, de Sanjosé S, Diaz M. Impact of model calibration on cost-effectiveness analysis of cervical cancer prevention. Scientific Reports. 2017;7(1):17208. doi:10.1038/s41598-017-17215-2.
7. Peñaloza-Ramos MC, Jowett S, Sutton AJ, McManus RJ, Barton P. The Importance of Model Structure in the Cost-Effectiveness Analysis of Primary Care Interventions for the Management of Hypertension. Value in Health. 2018;21(3):351-63. doi:https://doi.org/10.1016/j.jval.2017.03.003.
8. Schwander B, Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert review of pharmacoeconomics & outcomes research. 2018;18(5):529-41. doi:10.1080/14737167.2018.1501680.
9. Schwander B, Nuijten M, Hiligsmann M, Queally M, Leidl R, Joore M et al. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy and Technology. 2020;9(3):314-22. doi:https://doi.org/10.1016/j.hlpt.2020.03.005.
10. Schwander B, Nuijten M, Evers S, Hiligsmann M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. PharmacoEconomics. 2021;39(4):433-46. doi:10.1007/s40273-021-01008-7.
11. Schwander B, Kaier K, Hiligsmann M, Evers S, Nuijten M. Does the Structure Matter? An External Validation and Health Economic Results Comparison of Event Simulation Approaches in Severe Obesity. PharmacoEconomics. 2022;40(9):901-15. doi:10.1007/s40273-022-01162-6.
12. Boyers D, Avenell A, Stewart F, Robertson C, Archibald D, Douglas F et al. A systematic review of the cost-effectiveness of non-surgical obesity interventions in men. Obes Res Clin Pract. 2015;9(4):310-27. doi:10.1016/j.orcp.2015.03.001.
13. Flego A, Keating C, Moodie M. Cost-effectiveness of whole-of-community obesity prevention programs: an overview of the evidence. Expert review of pharmacoeconomics & outcomes research. 2014;14(5):719-27. doi:10.1586/14737167.2014.953933.

14. Wang BC, Furnback W. Modelling the long-term outcomes of bariatric surgery: A review of cost-effectiveness studies. Best Pract Res Clin Gastroenterol. 2013;27(6):987-95. doi:10.1016/j.bpg.2013.08.022.

15. Henteleff HJ, Birch DW, Hallowell PT. Cost-effectiveness of bariatric surgery for severely obese adults with diabetes. Can J Surg. 2013;56(5):353-5. doi:10.1503/cjs.020713.

16. Gandjour A. Cost-effectiveness of preventing weight gain and obesity: what we know and what we need to know. Expert review of pharmacoeconomics & outcomes research. 2012;12(3):297-305. doi:10.1586/erp.12.13.

17. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. Nutrition. 2012;28(9):829-39. doi:10.1016/j.nut.2011.11.016.

18. Lehnert T, Sonntag D, Konnopka A, Riedel-Heller S, König HH. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. Obes Rev. 2012;13(6):537-53. doi:10.1111/j.1467-789X.2011.00980.x.

19. Loveman E, Frampton GK, Shepherd J, Picot J, Cooper K, Bryant J et al. The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: a systematic review. Health Technol Assess. 2011;15(2):1-182. doi:10.3310/hta15020.

20. Griffiths UK, Anigbogu B, Nanchahal K. Economic evaluations of adult weight management interventions: a systematic literature review focusing on methods used for determining health impacts. Appl Health Econ Health Policy. 2012;10(3):145-62. doi:10.2165/11599250-000000000-00000.

21. Yi Y, Philips Z, Bergman G, Burslem K. Economic models in type 2 diabetes. Curr Med Res Opin. 2010;26(9):2105-18. doi:10.1185/03007995.2010.494451.

22. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(2):231-50. doi:10.1016/j.jval.2013.02.002.

23. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2022;25(1):10-31. doi:10.1016/j.jval.2021.10.008.

24. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):796-803. doi:10.1016/j.jval.2012.06.012.

25. Sharifi M, Franz C, Horan CM, Giles CM, Long MW, Ward ZJ et al. Cost-Effectiveness of a Clinical Childhood Obesity Intervention. Pediatrics. 2017;140(5). doi:10.1542/peds.2016-2998.

26. Alsumali A, Eguale T, Bairdain S, Samnaliev M. Cost-Effectiveness Analysis of Bariatric Surgery for Morbid Obesity. Obes Surg. 2018;28(8):2203-14. doi:10.1007/s11695-017-3100-0.

27. Lee M, Lauren BN, Zhan T, Choi J, Klebanoff M, Abu Dayyeh B et al. The cost-effectiveness of pharmacotherapy and lifestyle intervention in the treatment of obesity. Obes Sci Pract. 2020;6(2):162-70. doi:10.1002/osp4.390.

28. Boyers D, Retat L, Jacobsen E, Avenell A, Aveyard P, Corbould E et al. Cost-effectiveness of bariatric surgery and non-surgical weight management programmes for adults with severe obesity: a decision analysis model. Int J Obes (Lond). 2021;45(10):2179-90. doi:10.1038/s41366-021-00849-8.

29. Gandjour A. A parsimonious model to validate cost-effectiveness analyses on preventive health care. BMC Health Serv Res. 2021;21(1):1213. doi:10.1186/s12913-021-07217-2.

30. Klebanoff MJ, Chhatwal J, Nudel JD, Corey KE, Kaplan LM, Hur C. Cost-effectiveness of Bariatric Surgery in Adolescents With Obesity. JAMA Surg. 2017;152(2):136-41. doi:10.1001/jamasurg.2016.3640.

31. Ananthapavan J, Nguyen PK, Bowe SJ, Sacks G, Mantilla Herrera AM, Swinburn B et al. Cost-effectiveness of community-based childhood obesity prevention interventions in Australia. Int J Obes (Lond). 2019;43(5):1102-12. doi:10.1038/s41366-019-0341-0.

32. Kenney EL, Mozaffarian RS, Long MW, Barrett JL, Cradock AL, Giles CM et al. Limiting Television to Reduce Childhood Obesity: Cost-Effectiveness of Five Population Strategies. Child Obes. 2021;17(7):442-8. doi:10.1089/chi.2021.0016.

33. Nuijten M, Marczewska A, Araujo Torres K, Rasouli B, Perugini M. A health economic model to assess the cost-effectiveness of OPTIFAST for the treatment of obesity in the United States. J Med Econ. 2018;21(9):835-44. doi:10.1080/13696998.2018.1468334.

34. Lopes S, Johansen P, Lamotte M, McEwan P, Olivieri A-V, Foos V. External Validation of the Core Obesity Model to Assess the Cost-Effectiveness of Weight Management Interventions. PharmacoEconomics. 2020;38(10):1123-33. doi:10.1007/s40273-020-00941-3.

35. Tappenden P, Caro JJ. Improving Transparency in Decision Models: Current Issues and Potential Solutions. PharmacoEconomics. 2019;37(11):1303-4. doi:10.1007/s40273-019-00850-0.

36. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM). PharmacoEconomics. 2017;35(3):375-96. doi:10.1007/s40273-016-0471-3.

37. Sjöström L. Review of the key results from the Swedish Obese Subjects (SOS) trial - a prospective controlled intervention study of bariatric surgery. J Intern Med. 2013;273(3):219-34. doi:10.1111/joim.12012.

38. Brown JB, Palmer AJ, Bisgaard P, Chan W, Pedula K, Russell A. The Mt. Hood challenge: cross-testing two diabetes simulation models. Diabetes Res Clin Pract. 2000;50 Suppl 3:S57-64. doi:10.1016/s0168-8227(00)00217-5.

39. Palmer AJ, Clarke P, Gray A, Leal J, Lloyd A, Grant D et al. Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(4):670-85. doi:10.1016/j.jval.2013.01.002.

40. Si L, Willis MS, Asseburg C, Nilsson A, Tew M, Clarke PM et al. Evaluating the Ability of Economic Models of Diabetes to Simulate New Cardiovascular Outcomes Trials: A Report on the Ninth Mount Hood Diabetes Challenge. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2020;23(9):1163-70. doi:10.1016/j.jval.2020.04.1832.

41. Akter T, Zeba Z, Hosen I, Al-Mamun F, Mamun MA. Impact of the COVID-19 pandemic on BMI: Its changes in relation to socio-demographic and physical activity patterns based on a short period. PLoS One. 2022;17(3):e0266024. doi:10.1371/journal.pone.0266024.

42. Al-Domi H, Al-Dalaeen A, Al-Rosan S, Batarseh N, Nawaiseh H. Healthy nutritional behavior during COVID-19 lockdown: A cross-sectional study. Clin Nutr ESPEN. 2021;42:132-7. doi:10.1016/j.clnesp.2021.02.003.

43. Kriaucioniene V, Bagdonaviciene L, Rodríguez-Pérez C, Petkeviciene J. Associations between Changes in Health Behaviours and Body Weight during the COVID-19 Quarantine in Lithuania: The Lithuanian COVIDiet Study. Nutrients. 2020;12(10). doi:10.3390/nu12103119.

44. La Fauci G, Montalti M, Di Valerio Z, Gori D, Salomoni MG, Salussolia A et al. Obesity and COVID-19 in Children and Adolescents: Reciprocal Detrimental Influence-Systematic Literature Review and Meta-Analysis. International journal of environmental research and public health. 2022;19(13). doi:10.3390/ijerph19137603.

45. Sawadogo W, Tsegaye M, Gizaw A, Adera T. Overweight and obesity as risk factors for COVID-19-associated hospitalisations and death: systematic review and meta-analysis. BMJ Nutr Prev Health. 2022;5(1):10-8. doi:10.1136/bmjnph-2021-000375.

46. Raeisi T, Mozaffari H, Sepehri N, Darand M, Razi B, Garousi N et al. The negative impact of obesity on the occurrence and prognosis of the 2019 novel coronavirus (COVID-19) disease: a systematic review and meta-analysis. Eat Weight Disord. 2022;27(3):893-911. doi:10.1007/s40519-021-01269-3.

47. Földi M, Farkas N, Kiss S, Zádori N, Váncsa S, Szakó L et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. Obes Rev. 2020;21(10):e13095. doi:10.1111/obr.13095.

SUMMARY

Summary

This dissertation studies the systematic evaluation, replication and validation of structural health economic modelling approaches in the field of obesity.

In particular it evaluates, replicates and validates the current structural modelling landscape in obesity, with an emphasis on commonly applied obesity-associated event simulation approaches. This research aims to increase trust and confidence in the selection and interpretation of results related to a specific methodological approach used as basis for decision analytic models in obesity. The research presented in this document is mainly informed by the content of five connected scientific publications (chapters 2-6).

In chapter 2, a systematic review on health economic obesity models is reported, which identified a total of 87 scientific articles. These 87 articles reported 69 unique modelling approaches, hence a huge diversity of obesity modelling approaches have been identified. This makes it difficult to compare and comprehend the model outcomes, as the structural and methodological differences could also have a major impact on the modelling results.

Chapter 3 focuses especially on the (diverse) **clinical event simulation approaches and the (lack of) external validation** in the health economic obesity models identified in chapter 2. This research found that one key limitation of these models is the lack of published external validation results. This is in spite of the valuable information provided by such methods on the predictiveness, and hence on the quality, of their event simulation approaches. Only ten model-based health economic assessments in obesity (14%; 10 of 72) performed an external validation and the predictiveness of the event simulation was investigated in a cohort of obese subjects in only one study. Future work on quality assessment of key structural approaches (expert panel) and on the comparison of most common event simulation approaches (cross validation & external validation) is required in order to guide future modelling in the field of obesity. This is presented in chapters 4, 5 and 6. Furthermore, the wide range of modelling approaches (identified in chapter 2) suggested the need to develop best practice recommendations for model-based health economic assessments in obesity.

Accordingly, **chapter 4,** reports on the methodology and results of an **expert panel rating** on **key structural approaches** used in the identified health economic obesity models, which were transformed into **(best practice) expert recommendations**. Focusing on the key structural aspects outlined in the Philips checklist, this research presents main findings relevant to obesity models that have been identified (systematic literature search), rated (expert interviews) and

discussed (expert panel). While the expert panel acknowledged the challenges in achieving consensus, several recommendations for key structural approaches for a health economic obesity model were developed.

Chapter 5, based on the systematic review and the expert panel recommendations (chapters 2 to 4), focuses on the selection and **replication** of **high-quality health economic obesity models** and on the **assessment of reproduction success.** This study confirms the feasibility of replicating complex obesity models, although some challenges were identified. Small changes to existing reporting criteria have the potential to increase the transparency of model reporting, and may increase the reproduction success of health economic modelling results, which may subsequently increase the transparency and acceptance of health economic modelling studies.

In **chapter 6,** the **influence of** the (different) **structural modelling approaches on** the **clinical event simulation and** the **health economics outcomes** is further investigated, and hence targets the research needs identified in the previous chapters. This research identifies that in a severely obese population, the structure of a health economic model matters if clinical events are to be predicted most accurately. However, if the purpose of a health economic model is purely the incremental cost-effectiveness ratio, this study suggests that the structure does not matter as much, as health economic results are fairly comparable. Further similar studies in other obese populations and in other disease areas would be needed to confirm the findings.

Finally, in **chapter 7** the main objectives and main results of the thesis findings are summarized and discussed in relation to the broader research context. In this chapter the main contributions of the thesis to the health economic research and to scientific debate are reported. Furthermore, the methodological challenges and considerations are discussed, and implications and recommendations for future research are provided.

This chapter highlighted that our research was able to identify some important aspects related to the health economic modelling methodology in general, and key aspects specifically related to the context of obesity. Our research could form a basis for evaluating the strength and weaknesses of different structural event simulation approaches, but also identified valuable future areas of research to further enhance trust and confidence in health economic modelling, especially in the context of obesity.

IMPACT

Impact

Main Objective and Main Results

This dissertation aimed to study the methodology and validity of published health economic models in the context of obesity, usually defined by a BMI >30 kg/m² [1]. As obesity is a complex disease which impacts the human body in different ways, there are many diseases associated with obesity [2]. This means that the risk for a specific disease (e.g. coronary heart disease) is much higher in an obese person compared to a normal weight person [3]. In order to reduce this risk in obese persons, different approaches are available to reduce the person's weight, and other associated risk factors of obesity (e.g. high blood pressure), which can have positive impact on the life expectancy and quality of life.

To measure such long-term consequences with clinical studies would require a very long observation period, which would require massive funds to be invested in such studies. As time and funds are often not available, health economic models are instead used to predict the potential long-term consequences. This dissertation investigated the different methods used for such predictive obesity models, investigated how accurately these models predict the reality, and studied the impact of the modelling methodology on the health economic model outcomes. Furthermore, it tested whether the information usually published for such predictive models allows reprogramming of the model and reproduction of the results, a criterion for the quality of reporting of scientific experiments irrespective of the research field.

As presented in this dissertation, it was found that in the context of obesity (almost) every research team builds its own obesity model; this is reflected by the obtained diversity of obesity modelling approaches (chapter 2). This makes it difficult to compare model outcomes, as the structural and methodological differences could have a major impact on the modelling results. Furthermore, it was found, that one key limitation of these models is the lack of published external validation results which could provide valuable information on the predictiveness, and quality, of their event simulation approaches (chapter 3). Hence it is unclear whether decision-makers in the healthcare setting can rely on (trust) the results of those models. Therefore, the different modeling approaches were presented and discussed with experienced health economists, in order to create best practice recommendations for the key structural approaches for health economic obesity models (chapter 4). Using these expert recommendations, high quality health economic obesity models were selected and replicated (chapter 5). Here it was found that small changes to existing reporting criteria have the potential to increase the transparency of model reporting, and may increase the reproduction success of health economic modelling results. This may subsequently increase the transparency and acceptance

of health economic modelling studies. Finally (in chapter 6) it was found that in a severely obese population, the structure of a health economic model matters if clinical events are to be predicted most accurately. However, if the purpose of a health economic model is purely the incremental health economic comparison, this study suggests that the structure does not matter as much, as incremental health economic results are fairly comparable.

Scientific Impact

The research presented in this dissertation highlighted the increasing importance of health economic models in obesity, which is primarily triggered by the increasing burden of obesity and the related increased need for efficient allocation of resources. This has also been confirmed by the large number of health economic obesity modelling studies identified by the systematic review reported in chapters 2 & 3.

This systematic review has furthermore shown strong variability in predictive modelling in obesity. This variability was investigated, for the first time, with a special emphasis on the presentation and categorization of different approaches for predicting obesity associated events. This strong variability in the structural modelling approaches highlighted the need for recommendations and/or minimal requirements to inform obesity models. In order to offer guidance for scientists and modelers, best practice criteria were developed (chapter 4). It is expected that these best practice criteria can help to better harmonize the applied modelling methodologies in obesity.

Using these best practice criteria, high quality obesity models were selected for replication. This replication exercise (chapter 5) provided evidence that even complex obesity models can be rebuilt if the reporting and hence the transparence is sufficient for those exercises. This study provided important input for the reporting criteria of health economic models and, as we shared the outcomes of our research with the committee responsible for updating the CHEERS II reporting criteria [4], led to changes in the newest CHEERS II update. In addition to the previous CHEERS version [25], it was now made clear in the updated example CHEERS II reporting tables, that the details and parameters of probability distributions are to be reported. This enables a 1:1 replication of the probability distributions, which was not possible for our replications, reported in chapter 5, as the related details were missing. Furthermore, for assessing the success of the reproduction results, we have applied different criteria as defined and proposed in a recently published review on this topic [5], and we proposed a combination of different criteria to determine “replication success” specifically for health economic modeling. As this was, to our knowledge, the first application of these success

criteria, further research and scientific dialogue is required to investigate and define how best rate the success of a health economic model replication.

The external validation and the health economic result comparison (chapter 6) shows there is still a need for more long-term intervention studies in obesity, to provide better understanding of the condition, and a broader information source for the external validation. Using the currently available evidence base, focusing on the SOS study reflecting a severely obese population [6], it was shown that BMI alone is no good predictor for obesity associated events, but that a broader approach, considering a broader set of risk factors, provides better event prediction results. Interestingly, considering the incremental health economic results, no large difference was observed between the approaches, which should enhance trust in the health economic outcomes produced by obesity models, irrespective of the chosen approach.

The findings of this thesis will help researchers, health economists and modelers to make better informed decisions on the choice of a suitable modelling methodology for obesity models, and offers guidance for future fields of research. The research and findings of this thesis are relevant for all chronic diseases, in which health economic modelling is frequently applied to translate surrogate parameters (such as BMI, high blood pressure, fasting glucose levels) into patient relevant endpoints (such as stroke or myocardial infarction). In all such cases the transparency of research reporting and the validation of a modeling approach are crucial to gaining trust and confidence in the health economic outcomes. Future research in the field of obesity and other chronic conditions is required to complement the findings of this thesis.

Social Impact

Health economic research often has practical implications, namely informing decision makers on the most efficient and cost-efficient way of allocating scarce resources within a given healthcare system. Decision makers in the healthcare setting can be payers, politicians, administrators, clinicians or other central member of decision-making boards. Irrespective of the background and the specific perspective a decision maker has, they need to rely on the information provided by researchers. Health economists especially play a crucial role in informing such decisions, as they combine and synthesize information from different disciplines, in order to simulate the clinical and economic consequences of such decisions for individuals and for society. Therefore, trust and confidence in the health economic research are central factors in ensuring the best allocation of scarce resources. One central basis for informing decision makers are health economic models, that simulate the clinical and economic consequences of a (usually innovative) healthcare intervention and compare it to alternative routes of action (usually to

current standards). Health economists are aware of their responsibility as central health economic associations; the ISPOR and SMDM in particular point out that trust and confidence are critical to the success of such health care models [7].

The research presented in this thesis aims to increase this trust and confidence in health economic models used for decision making in the context of obesity. Hence a potential social impact of this thesis is that decision makers have better guidance on how a specific modelling approach might influence clinical and health economic model outcomes. This might lead to better informed decision making, and potentially to a better acceptance of health economic modeling studies in the context of obesity.

Dissemination of Research Results

Besides the publication of this thesis as a whole, single components of this thesis (chapters 2-6) were all published in peer-reviewed scientific journals [8-12], whereas two papers were published open-access (chapter 5 and 6) [11, 12]. In addition, the publication of each paper was announced via social media channels to increase the awareness of researches and decision makers.

In order to enhance the dissemination of these findings, each chapter was additionally presented to at least one scientific congress, in which researchers and decision makers commonly participate. These congresses were organized by the following associations (in brackets the number of thesis related congress contributions is shown): German health economic association (n=3 congress contributions); International Society for Pharmacoeconomics and Outcomes Research (n=3 congress contributions); European Health Economic Association (n=1 congress contribution); International Health Economic Association (n=1 congress contribution); the society for Health Technology Assessment International (n=1 congress contribution).

These congress contributions were always presented before the publication of the full manuscript in order to obtain first feedback for the related research, and to potentially include a broader perspective in the related discussion of a specific research paper.

In addition to these presentation and publication activities, the findings of our research were shared with the International Health Economic Association special interest group members, “Economics of Obesity”.

References

1. World Health Organization. Fact Sheet on Obesity. 2003. https://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf. Accessed 09.02.2019.
2. World Health Organization. Updated Fact Sheet on Obesity. 2018. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed 09.02.2019.
3. Field AE, Coakley EH, Must A, Spadano JL, Laird N, Dietz WH et al. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. Arch Intern Med. 2001;161(13):1581-6. doi:10.1001/archinte.161.13.1581.
4. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2022;25(1):10-31. doi:10.1016/j.jval.2021.10.008.
5. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. PharmacoEconomics. 2019;37(11):1371-81. doi:10.1007/s40273-019-00836-y.
6. Sjöström L. Review of the key results from the Swedish Obese Subjects (SOS) trial - a prospective controlled intervention study of bariatric surgery. J Intern Med. 2013;273(3):219-34. doi:10.1111/joim.12012.
7. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012;15(6):843-50. doi:10.1016/j.jval.2012.04.012.
8. Schwander B, Hiligsmann M, Nuijten M, Evers S. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert review of pharmacoeconomics & outcomes research. 2016;16(5):561-70. doi:10.1080/14737167.2016.1230497.
9. Schwander B, Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert review of pharmacoeconomics & outcomes research. 2018;18(5):529-41. doi:10.1080/14737167.2018.1501680.
10. Schwander B, Nuijten M, Hiligsmann M, Queally M, Leidl R, Joore M et al. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy and Technology. 2020;9(3):314-22. doi:https://doi.org/10.1016/j.hlpt.2020.03.005.
11. Schwander B, Nuijten M, Evers S, Hiligsmann M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. PharmacoEconomics. 2021;39(4):433-46. doi:10.1007/s40273-021-01008-7.
12. Schwander B, Kaier K, Hiligsmann M, Evers S, Nuijten M. Does the Structure Matter? An External Validation and Health Economic Results Comparison of Event Simulation Approaches in Severe Obesity. PharmacoEconomics. 2022. doi:10.1007/s40273-022-01162-6.

DISSEMINATION ACTIVITIES

Dissemination Activities

Scientific Articles in Peer-Reviewed Journals

Schwander B, Hiligsmann M, Nuijten M, Evers S. Systematic review and overview of health economic evaluation models in obesity prevention and therapy. Expert review of pharmacoeconomics & outcomes research. 2016;16(5):561-70. https://doi.org/10.1080/14737167.2016.1230497

Schwander B, Nuijten M, Hiligsmann M, Evers S. Event simulation and external validation applied in published health economic models for obesity: a systematic review. Expert review of pharmacoeconomics & outcomes research. 2018;18(5):529-41. https://doi.org/10.1080/14737167.2018.1501680

Schwander B, Nuijten M, Hiligsmann M, Queally M, Leidl R, Joore M et al. Identification and expert panel rating of key structural approaches applied in health economic obesity models. Health Policy and Technology. 2020;9(3):314-22. https://doi.org/10.1016/j.hlpt.2020.03.005.

Schwander B., Nuijten M., Evers S., Hiligsmann M. Replication of Published Health Economic Obesity Models: Assessment of Facilitators, Hurdles and Reproduction Success. PharmacoEconomics. 2021;39(4):433-46. https://doi.org/10.1007/s40273-021-01008-7

Schwander B., Kaier K., Hiligsmann M., Evers S., Nuijten M. Does the Structure Matter? An External Validation and Health Economic Results Comparison of Event Simulation Approaches in Severe Obesity. PharmacoEconomics 2022;40(9):901-15. https://doi.org/10.1007/s40273-022-01162-6

Congress Contributions

Schwander B, Hiligsmann M, Nuijten M, Evers S: Systematische Übersichtsarbeit zu gesundheitsökonomischen Evaluationen im Kontext der Behandlung und Prävention der Adipositas. Short presentation and poster at the 8th annual conference of the German Society of Health Economics (DGGÖ), Berlin, Germany, 15th March 2016.

Schwander B, Hiligsmann M, Nuijten M, Evers S: Systematic review and overview on health economic evaluation models in the context of obesity therapy and prevention. Poster presentation at the European Health Economic Association (EuHEA), Hamburg, Germany, 15th July 2016.

Schwander B, Nuijten M, Hiligsmann M, Evers S: Evaluation of methodological variations in the event simulation approaches of published health economic models in obesity prevention and therapy. 19th ISPOR Annual European Congress, Vienna, Austria, 31st October 2016 (PRM102).

Schwander B, Nuijten M, Hiligsmann M, Evers S. Evaluation methodischer Vorgehensweisen zur Simulationen klinischer Ereignisse im Kontext der Prävention und Behandlung der Adipositas [Evaluation of methodological variations in the event simulation approaches of published health economic models in the context of obesity therapy and prevention]. Presentation at the 9th annual conference of the German Society of Health Economics (DGGÖ), Basel, Switzerland, 10th March 2017.

Schwander B, Nuijten M, Hiligsmann M, Evers S: Quantity and quality of external event validation procedures performed in published health economic models in obesity: outcomes of a systematic review. 20th ISPOR Annual European Congress, Glasgow, Scotland, 8th November 2017 (PRM131).

Schwander B, Nuijten M, Hiligsmann M, Queally Q, Leidl R, Joore M; Oosterhoff M, Frew E, van Wilder P, Postma M, Evers S: Identification and Expert Panel Rating of Key Structural Approaches applied in Health Economic Obesity Models. Presentation at the 11th annual conference of the German Society of Health Economics (DGGÖ), Augsburg, Germany, 19th March 2019.

Schwander B, Nuijten M, Hiligsmann M, Evers S: Reproduction of Published Health Economic Models in Obesity: Assessment of Facilitators & Hurdles. Oral Presentation at the ISPOR Annual European Congress, Virtual Meeting, 30th October 2020.

Schwander B, Kaier K, Hiligsmann M, Evers S, Nuijten M. Does The Health Economic Modelling Structure Matter? An External Validation Of Three Approaches Commonly Used In Obesity Modelling. Oral Presentation at HTAi 2022 Annual Meeting, Utrecht, Netherlands, 27th June 2020.

ABOUT THE AUTHOR

About the author

Björn Schwander was born on 26th January 1977 in Freiburg im Breisgau, Germany. After receiving his nursing diploma from the Freiburg Regional Administrative Council in 2001, he started as a consultant focusing on clinical and health economic outcomes research. As an employed consultant, in different companies, he has led and worked on several projects with respect to market access, outcomes research, health economic modelling and reimbursement strategy and its implementation for various pharmaceutical, biotech and medical device companies.

In parallel to his consulting activities he obtained a Bachelor in Applied Health Sciences from the University of Applied Sciences Magdeburg (Germany) in 2005 and an International Master of Health Economics and Pharmacoeconomics from the Barcelona Business School based at the Pompeu Fabra University (Spain) in 2014. Afterwards he started as an external PhD student (Health Economics) at the Care And Public Health Research Institute (CAPHRI) of the Faculty of Health, Medicine and Life Sciences of the Maastricht University.

Bjoern has more than 20 years' experience in market access, outcomes research and health economic consultancy. Since 2012 he has been the founder and general manager of the AHEAD GmbH, which offers global health economic contract research and strategies and solutions for the German healthcare setting.

In the beginning of his consulting career he mainly supported global activities, such as global value dossier and core health economic model developments. In recent years he has focused on supporting different clients in the development, preparation and implementation of integrated market access and reimbursement strategies, with a special focus on the German healthcare setting. Alongside these activities he was involved in disseminating the results of the contract research activities by publishing several scientific congress contributions and peer-reviewed articles.

Bjoern acts as peer reviewer at health economic journals, is a guest lecturer (health economic modeling) at the Hamburg University of Applied Sciences, and video lecturer (health economics) at the International University of Applied Sciences in Erfurt.

ACKNOWLEDGEMENTS

Acknowledgements

The journey of this dissertation began when I had a great Saturday evening dinner meeting with my co-supervisor Mark, during the ISPOR EU congress in Amsterdam in November 2014. Here I learned about the opportunity of obtaining a PhD in Health Economics, as an external student, at the Maastricht University. Now, end of 2022, I am in the final stage of writing this doctoral thesis.

Hence my first great thank you is to my co-supervisor Dr. Mark Nuijten. Mark, without you I would not have had the opportunity to apply for an external PhD, at the Maastricht University. Thank you also for your ongoing support during this year-long journey. You always kept in close contact with me and had an open ear for any issues. This was especially necessary in the beginning, as I had some difficulties adapting from the more pragmatic and timing-oriented research approaches of contract research, to the more sophisticated and time-demanding academic research.

This topic brings me to my secondary supervisor Dr. Mickaël Hiligsmann. Mickaël, thank you for never getting tired of challenging me to broaden my perspective, to be more academic and less pragmatic, and to better explain and discuss the research methodologies, results and implications. I am sure that your ongoing and detailed input on my research enabled the high ranked publications of the manuscripts, that informed the main chapters of the PhD dissertation. I have the feeling you achieved the goal of developing a pragmatic contract-researcher into an academic researcher.

This brings me to my primary supervisor Prof. Dr. Silvia Evers. Silvia, I hope I don't get too close in calling you my real "doctor mother". Your open-minded and positive attitude in all talks and discussions we had carried me through some of the more challenging phases of this PhD journey. You always managed to bring key topics and discussions to a broader research perspective, and to bring your long academic experience to the table, which helped me to better focus on the essentials.

I am really looking forward to staying connected with all of you, and I already look forward to all meetings and exchanges, whether more academic or more private, that we will have in the future.

Besides my supervisor team, I would like to thank Janet van Caulil. Janet, thank you for your ongoing support related to the several administrative challenges of the PhD dissertation journey. Thank you also to the PhD office team for guiding me through the final stages, namely the preparation of the PhD defense ceremony.

This brings me to the members of the assessment committee. My gratitude goes to the esteemed and highly esteemed members of the committee, namely Dr. Karen Freijer, Dr. Ken Redekop, Prof. Dr. Stef Kremers, Prof. Dr. Carmen Dirksen and Prof. Dr. Aggie Paulus. Thank you very much, for reading and evaluating this PhD thesis and for your contribution to the PhD defense committee. I would like to give a special thanks to Prof. Dr. Aggie Paulus who chaired the assessment committee.

Besides the committee I thank my paranymphs, Dr. Bram Ramaekers and Carmen Heilmann-Lieberum, for their support and assistance during the PhD defense ceremony.

In addition, I would like to thank all participants of the expert panel presented in chapter 4, namely: Dr. Michelle Queally, Prof. Dr. Reiner Leidl, Prof. Dr. Manuela Joore, Dr. Marije Oosterhoff, Prof. Dr. Emma Frew, Prof. Dr. Philippe Van Wilder and Prof. Dr. Maarten Postma. Further I like to thank Dr. Klaus Kaier for his statistical input and support, related to the external validation presented in chapter 6.

I would also like to express my great gratitude to the Maastricht university, for giving me and many other researchers the opportunity to obtain a PhD as an external student, and for enabling this without asking for any financial contribution.

Last but not least, this whole journey would have not been possible without the ongoing support and love of my family, starting with my parents. Mum (Irmgard) and dad (Günter), I want to thank you greatly for always being there if there was any need for support. Great thanks also for assisting me so strongly with raising your grandson, Ewan. Without that support this journey would have not been possible.

This brings me to my son. Ewan, I thank you for many great and funny years of living and developing together. Thank you for always being considerate of me having to work for business, the PhD, or to bring forward any other idea that jumped into my mind.

Finally, I want to thank my wife, Carmen. Carmen, thank you for adopting me and Ewan as your new family. You are my daily sunshine and my ongoing supporter during good and bad phases of the journey called life.

I greatly love you all!