



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

Björn Schwander^{1,2} · Klaus Kaier³ · Mickaël Hiligsmann¹ · Silvia Evers^{1,4} · Mark Nuijten⁵

Accepted: 29 May 2022 / Published online: 30 June 2022
© The Author(s) 2022

Abstract

Objectives As obesity-associated events impact long-term survival, health economic (HE) modelling is commonly applied, but modelling approaches are diverse. This research aimed 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 (T2D) for both the surgery and the 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 (grades 1–4 expressing mild, moderate, severe, and very severe deviations). Furthermore, the costs per quality-adjusted life-year (QALY) gained of surgery versus 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 categorical BMI approach (overall 3.63; control 3.50; surgery 3.75). Considering different confidence interval limits, the costs 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.

1 Introduction

Obesity is a multifactorial, chronic disorder that is usually defined as a body mass index (BMI) $> 30 \text{ kg/m}^2$ [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 a continuous or categorical variable) is used as a central surrogate marker influenced by anti-obesity measures, but the application of widely used risk equations (e.g., UK Prospective Diabetes Study (UKPDS) and 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 this article, as the approach of simulating events is usually categorized as a structural health economic modelling component, for example 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

✉ Björn Schwander
bjoern.schwander@ahead-net.de

Extended author information available on the last page of the article

Key Points for Decision Makers

Health economic modelling is frequently applied in obesity to simulate the long-term consequences of the disease. 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.

To our knowledge, this is the first published research that investigated the impact of different commonly applied structural event simulation approaches in severe obesity modelling on event prediction and on health economic results.

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 that much, as incremental 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.

(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 to predict 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 modelling) literature lacks structural sensitivity analyses and provides only limited information on external validation [8].

Up to now it has not been investigated what 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 to offer a better guidance for outcome researchers, health economists, and decision makers on choosing and rating the structural approaches applied in health economic obesity models.

2 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 and 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 compared key health economic outcomes between the different structural approaches. The details and methodology of these different research steps are described below.

2.1 External Validation Study

As the 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 T2D, 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 who were treated with bariatric surgery intervention in the surgery arm. We 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.

2.2 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 reflects another structural approach for the obesity-associated

event simulation and were hence referred to according to the underlying structural event simulation approach as the 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 using the Framingham risk equations [25–28] in non-diabetics and 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]—i.e., 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 moving between the BMI categories.

Mortality was simulated by disease state-specific mortality risks and by a 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 T2D 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 to be flexible enough to self-adjust the risk for changing population characteristics. The details on the influencing factors considered for the different event simulation approaches, as well as the applied event rates, are presented in Online Supplemental Material (OSM) Table 1. A further calibration of the models was not performed.

2.3 Input Data and 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

and categorical BMI approaches) in order to simulate the events over time. The related SOS study data applied in the models are presented in detail in OSM Table 2 (baseline values) and OSM Table 3 (risk factor development over time).

The cost and health utility data for each model were informed by the data used in the latest UK NICE (National Institute for Health and Care Excellence) appraisal on obesity [16], which is presented in OSM Table 4. This allows a comparison of the health economic modelling results in terms of total costs, total quality-adjusted life-years (QALYs), and 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.

2.4 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° line. If the points are located over this 45° 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° line. The higher the slope is over 1 the stronger the overprediction of event rates by the model, and 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 ($\pm 25\%$ variation from the optimal slope value "1"; grade 1), moderate ($> 25\%$ and $\leq \pm 50\%$ variation, grade

2), severe ($> 50\%$ and $\leq \pm 100\%$ 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 in 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.

2.5 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% quartiles as the lower and upper ends of the box, the median as a line within the box, the mean as an “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.

3 Results

3.1 Event Validation Results

Looking at the detailed external event validation results presented in Figs. 1, 2, 3 and 4 and summarized in OSM Table 5, it can 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 (Figs. 1, 2, 3, 4) and according to the slope values, the following levels of over- and underprediction were observed: For the event mortality (Fig. 1), very severe overpredictions (grade 4) were observed for the continuous and categorical 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 (Fig. 2) presented a more diverse picture with a very severe overprediction (grade 4) observed in both study arms by the categorical 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 (Fig. 3) were very severely overpredicted (grade 4) by all approaches irrespective of the study arm, whereas here too 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 (Fig. 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 categorical 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 categorical BMI approach (overall grade 3.63; control arm 3.50; surgery arm 3.75). An overview of the grades by approach, event, and study arm, as well as the average grades, is provided in OSM Table 6.

3.2 Health Economic Results

The health economic results, comparing the control arm versus the surgery arm, related to the three structural approaches are presented in Table 1 and Fig. 5. Considering the mean results, presented in Table 1, the incremental

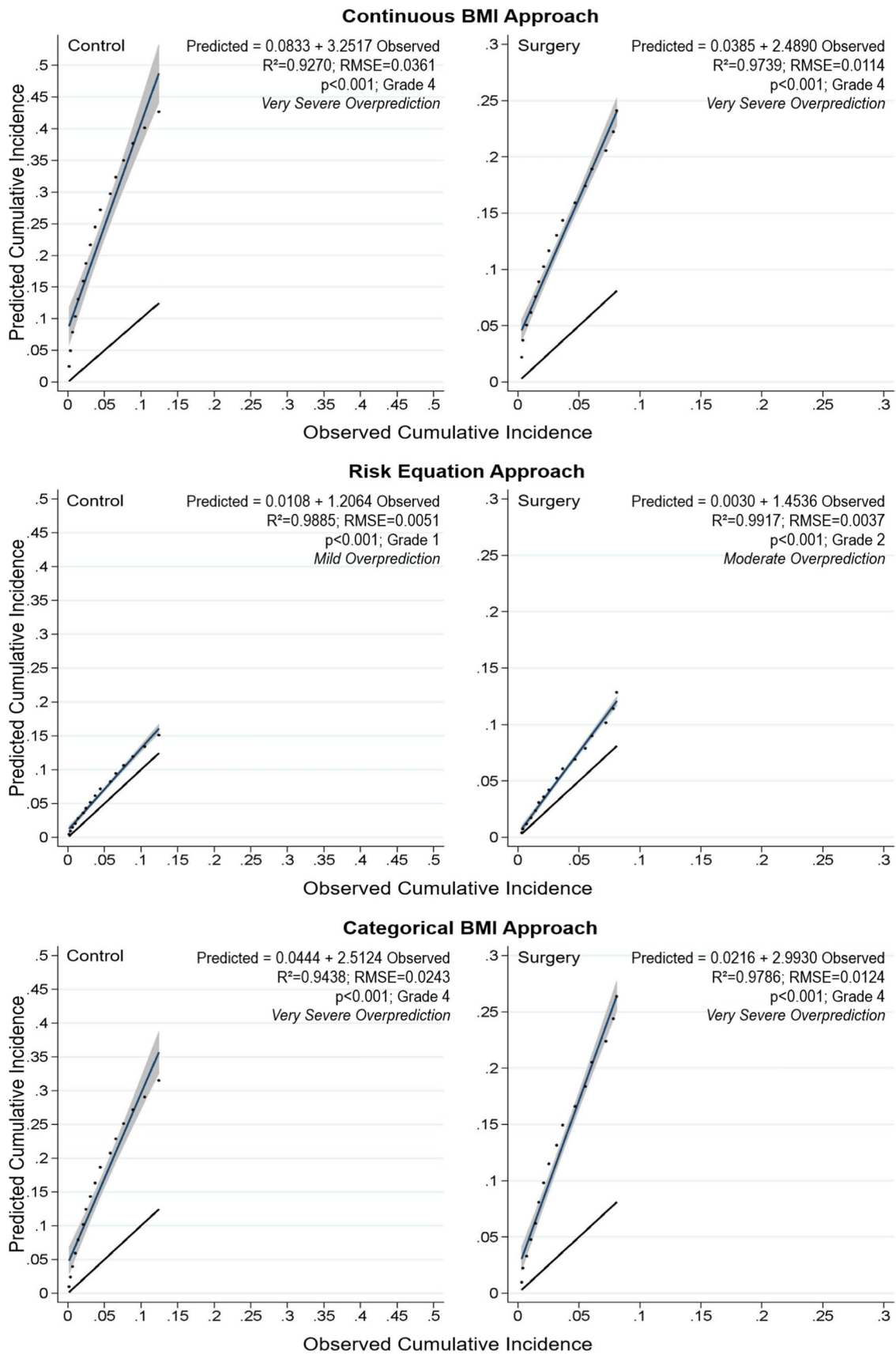


Fig. 1 Results of the external validation for overall mortality

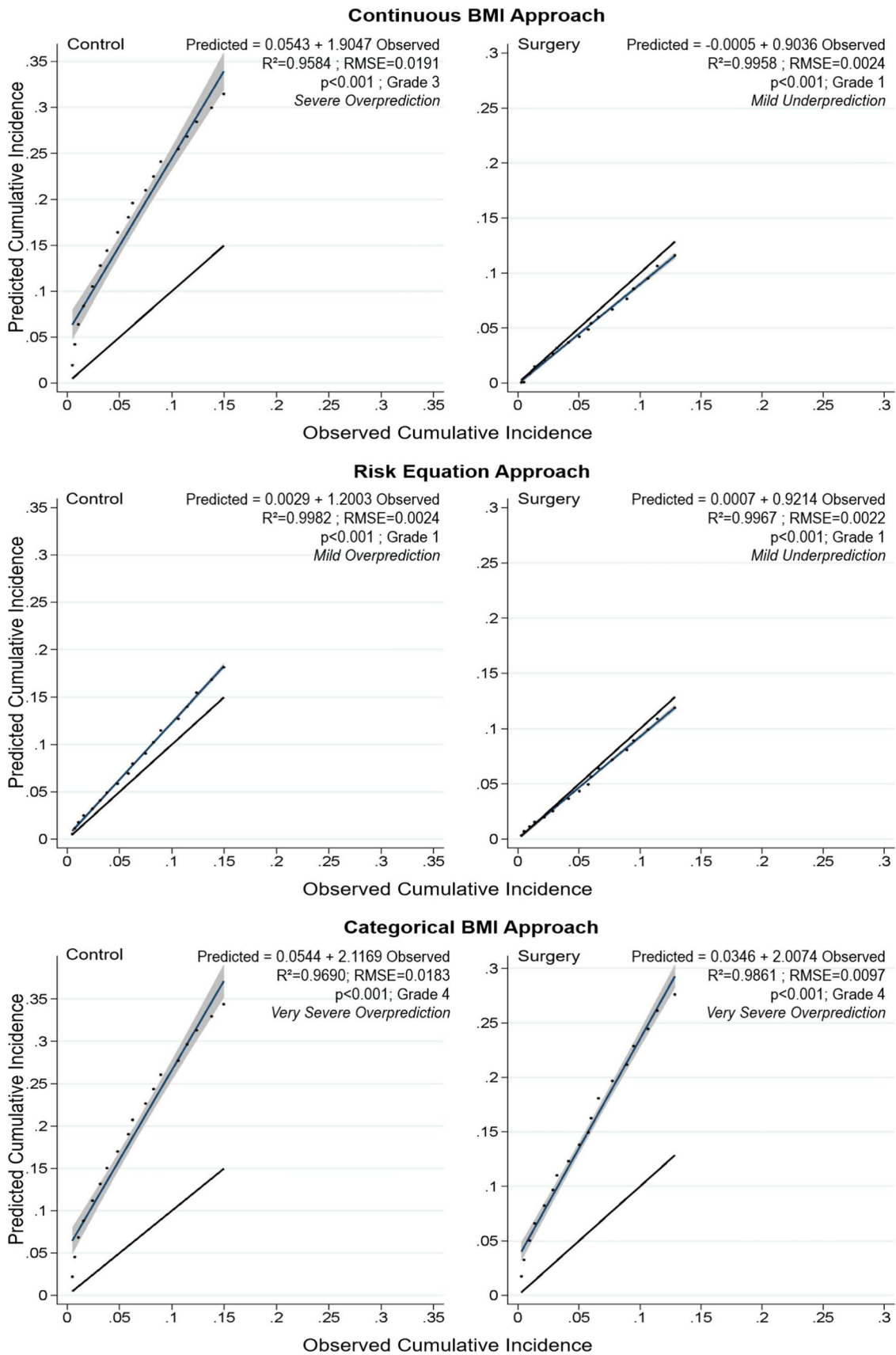


Fig. 2 Results of the external validation for total cardiovascular events

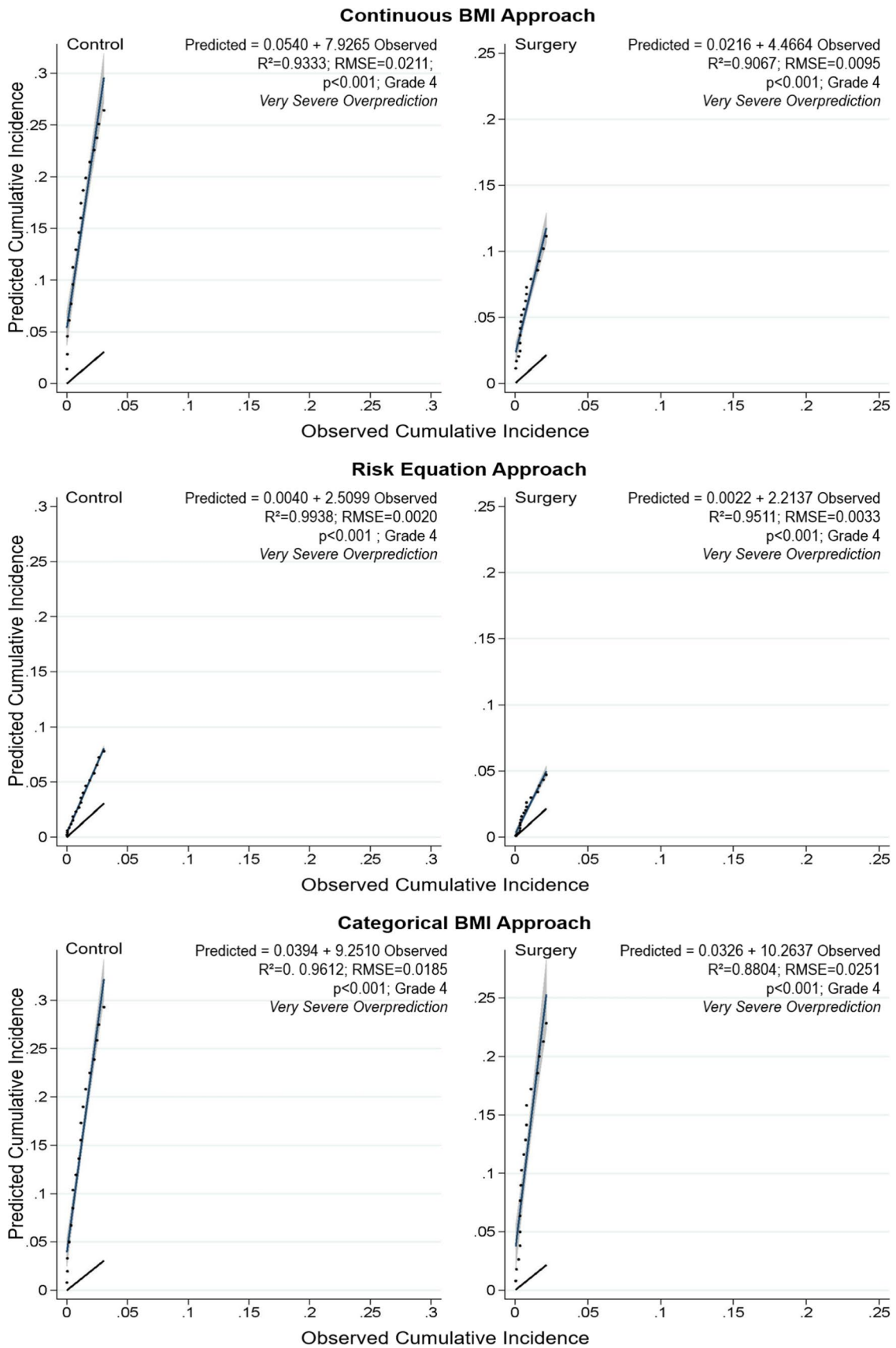


Fig. 3 Results of the external validation for fatal cardiovascular events

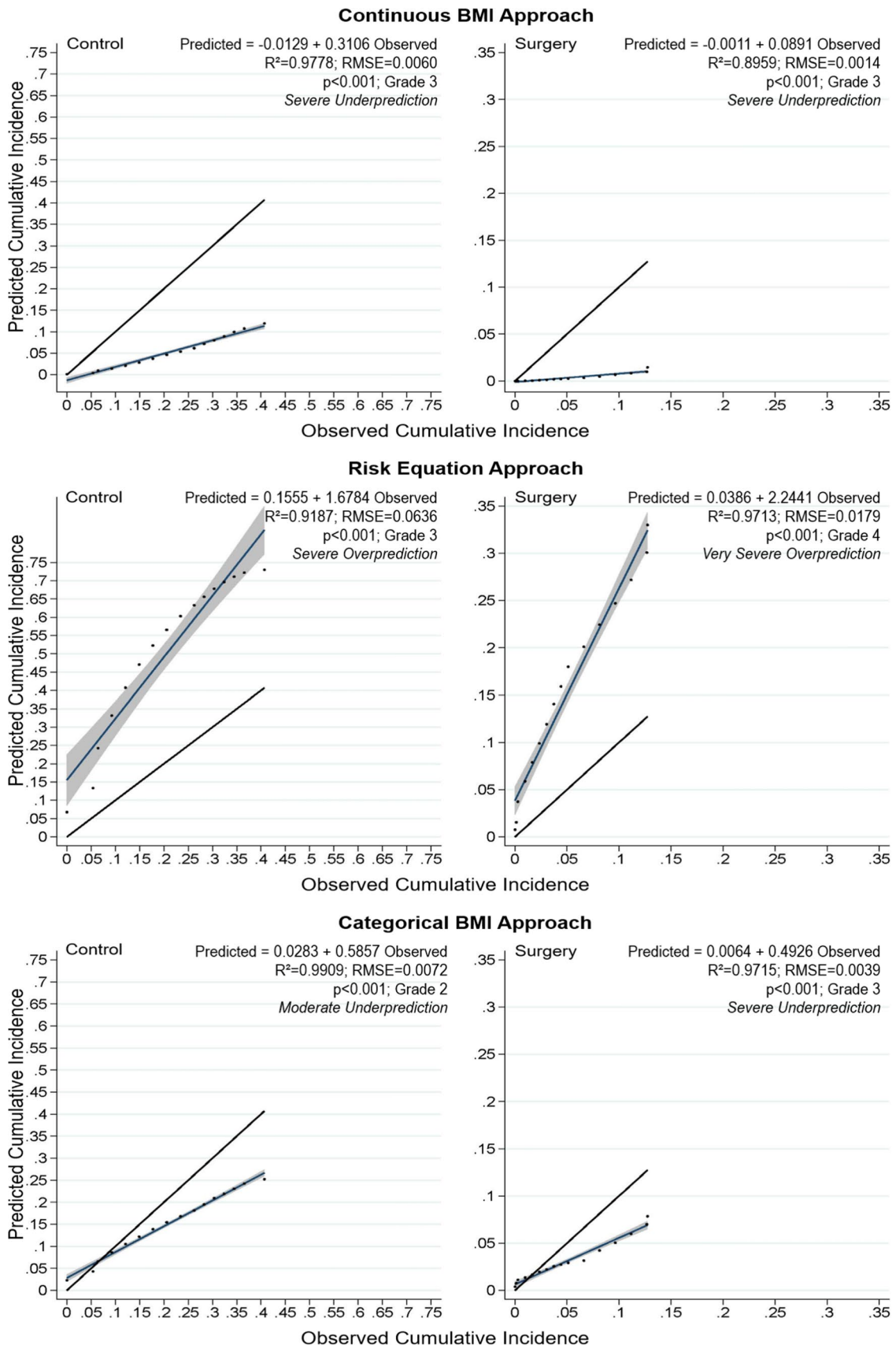


Fig. 4 Results of the external validation for type 2 diabetes

cost-effectiveness ratio (ICER) was lowest for the continuous BMI approach, followed by the risk equation approach, and was highest for the categorical BMI approach, irrespective of the model time horizon. However, looking at the distribution of the ICER values, presented in Fig. 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. The cost-effectiveness acceptability curves are visualized in Fig. 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 categorical BMI approaches.

4 Discussion

This study consisted of 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 the categorical BMI approach.

Only with regard to the prediction of T2D, 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 T2D in a Swedish cohort of severely obese patients, as according to our findings the T2D 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, for example the UK-based QDiabetes algorithm [41]. However, also here the predictive quality would still need to be investigated by an external validation.

In contrast to the risk equation approach, the external validation results of the continuous and categorical 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 BMI alone [42, 43]. Obesity is a health risk defined by abnormal or excessive fat 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 has not yet been transferred (broadly) into health-economic modelling. The reason why many health economic models still rely only on the BMI as a 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 the 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 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 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

Table 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	6598	7097	11.39	9.13	2.26	3143
	Risk Equation	14,410	7834	6576	14.57	12.60	1.97	3338
	Categorical BMI	14,873	4350	10,522	10.75	9.49	1.26	8328
Lifetime	Continuous BMI	18,126	10,162	7965	15.37	11.49	3.88	2055
	Risk Equation	26,354	19,637	6717	23.00	20.00	3.00	2241
	Categorical BMI	16,867	6599	10,268	13.92	12.27	1.65	6206

ICER incremental cost-effectiveness ratio

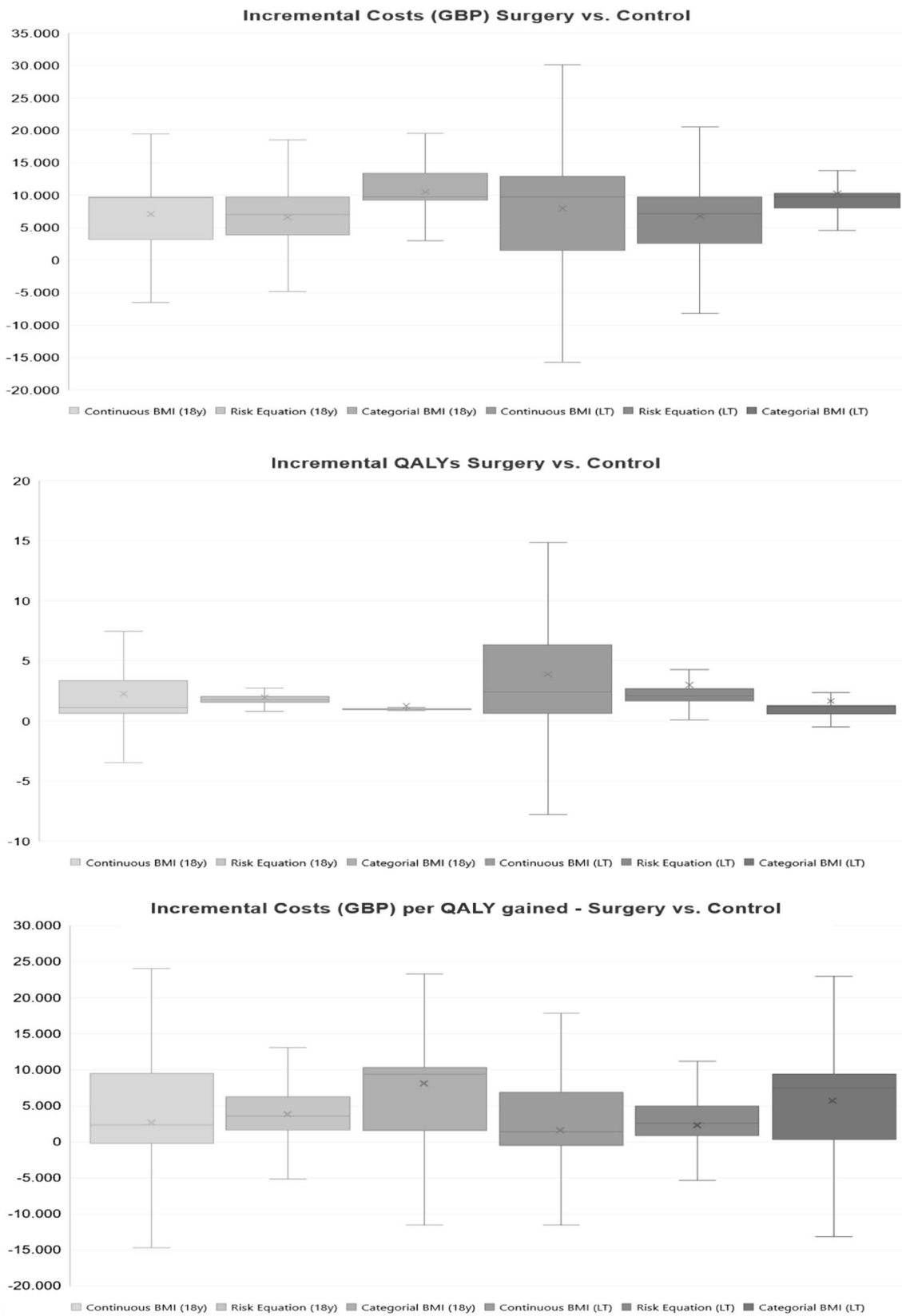


Fig. 5 Overview of incremental health economic outcomes

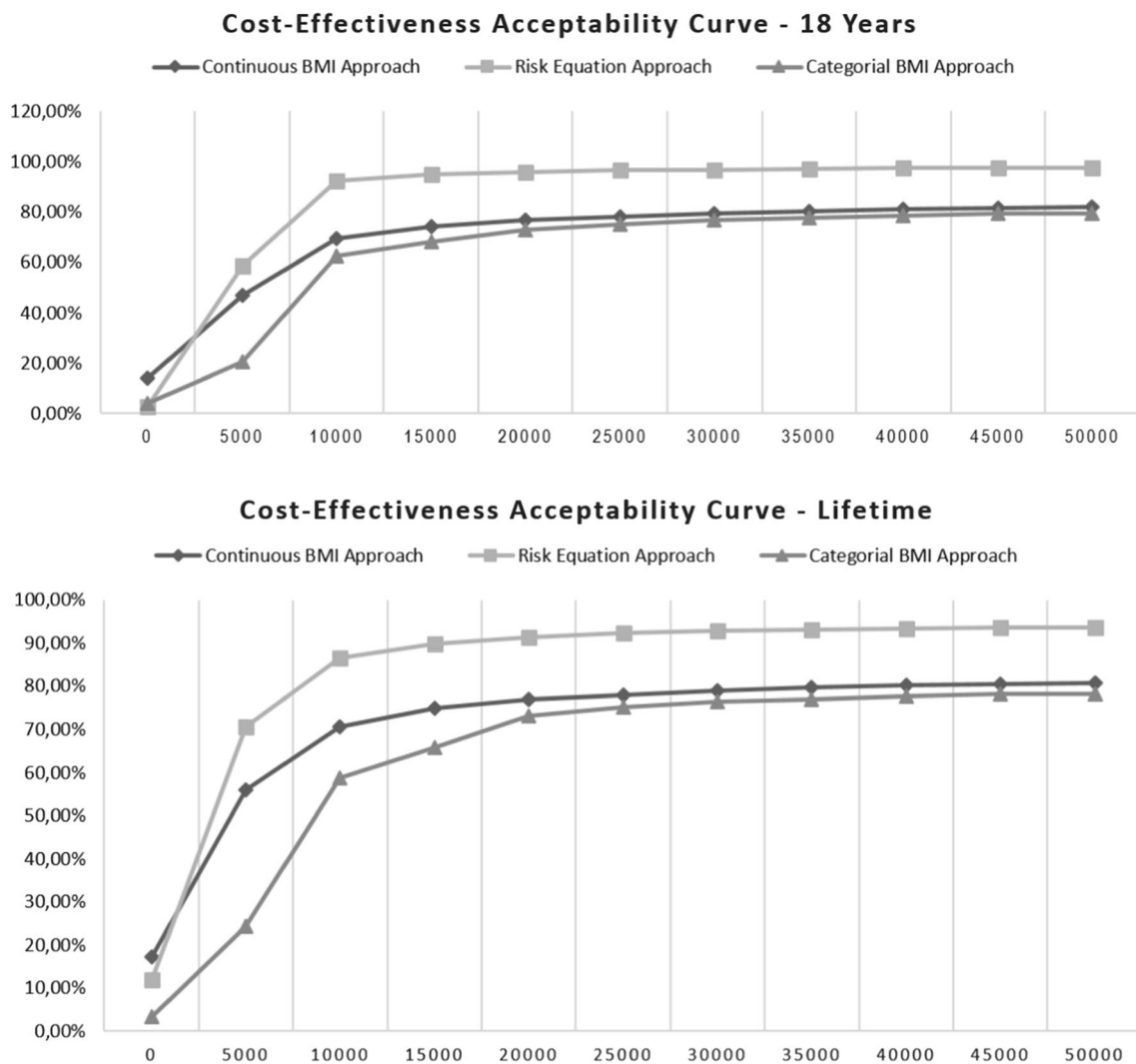


Fig. 6 Overview of cost-effectiveness acceptability curves

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 has shown a significant reduction in obesity-associated events and mortality in the bariatric surgery arm [15]. These findings support the positive reimbursement decisions on obesity surgery in many healthcare 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 article 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 categorical BMI approach) is not expected to have 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 also simulated osteoarthritis and colorectal cancer, the latter influencing survival. From both states simulated patients can 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 longer 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 explain neither the strong overprediction of mortality (indeed the SOS study also included cancer death) nor the strong underprediction of T2D 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, especially targeting 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 translated into disease risk changes and so 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 reflects 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 us 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 key health economic outcomes between the three structural approaches. Our main focus was on the comparison of the incremental cost per QALY gained, comparing the surgery versus the control arm, as this is observed as a central cost-effectiveness outcome by most cost-effectiveness driven payers and decision makers. Considering this key health economic 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 utmost importance for the healthcare 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 1), but if looking at the incremental results these differences are almost “absorbed”/“no more identifiable.”

However, if the mean ICER is to be presented and seen as the “main health economic result,” the categorical 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 at the cost-effectiveness acceptability curves, again the categorical BMI approach is the most conservative one, presenting the lowest probabilities of being cost-effective. The continuous 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 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 translated into a change in event risks. Hence, the hurdles for positive intervention effects are higher for the categorical 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 key health economic 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.

5 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 key health economic 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 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.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-022-01162-6>.

Declarations

Funding No funding was provided to assist in the preparation of this research.

Conflicts of interest The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors have reviewed the final manuscript and gave their consent for publication.

Data availability The data and models used during the presented study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Author contributions All authors contributed to the study conception and design. BS performed the model analysis and interpretation of data. KK performed the statistical analysis. All authors were involved in writing the manuscript, critically reviewed the manuscript, and have approved the final version submitted.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. World Health Organization. Fact Sheet on Obesity. 2003. https://www.who.int/dietphysicalactivity/media/en/gsf_s_obesity.pdf. Accessed 09 Feb 2019.
2. Wharton S, Lau DCW, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D, et al. Obesity in adults: a clinical practice guideline. *Can Med Assoc J*. 2020;192(31):E875–91. <https://doi.org/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. <https://doi.org/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 Feb 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 Rev Pharmacoecon*

- Outcomes Res. 2016;16(5):561–70. <https://doi.org/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. <https://doi.org/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 Health*. 2012;15(6):843–50. <https://doi.org/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 Rev Pharmacoecon Outcomes Res*. 2018;18(5):529–41. <https://doi.org/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. <https://doi.org/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. *J Med Econ*. 2007;10(3):239–54. <https://doi.org/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. <https://doi.org/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 Technol*. 2020;9(3):314–22. <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. <https://doi.org/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. *N Engl J Med*. 2004;351(26):2683–93. <https://doi.org/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. <https://doi.org/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/nice/appraisal-consultation-committee-papers-pdf-8952596893>. Accessed 19 Jan 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. <https://doi.org/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. *Popul Health Metrics*. 2016;14(1):17. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. [https://doi.org/10.1016/0002-8703\(91\)90861-b](https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1136/bmj.b349>.
 34. Mohan KM, Wolfe CD, Rudd AG, Heuschmann PU, Kolomin-sky-Rabas PL, Grieve AP. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. *Stroke*. 2011;42(5):1489–94. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1161/circoutcomes.111.964700>.
 37. Warren E, Brennan A, Akehurst R. Cost-effectiveness of sibutramine in the treatment of obesity. *Med Decision Making*. 2004;24(1):9–19. <https://doi.org/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 May 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1136/bmj.j5019>.
 42. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. *Nutr Today*. 2015;50(3):117–28. <https://doi.org/10.1097/nt.000000000000092>.
 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. <https://doi.org/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. <https://doi.org/10.1002/oby.20662>.

Authors and Affiliations

Björn Schwander^{1,2}  · Klaus Kaier³  · Mickaël Hiligsmann¹  · Silvia Evers^{1,4}  · Mark Nuijten⁵

Klaus Kaier
kaier@imbi.uni-freiburg.de

Mickaël Hiligsmann
m.hiligsmann@maastrichtuniversity.nl

Silvia Evers
s.evers@maastrichtuniversity.nl

Mark Nuijten
mark@a2m.nl

² AHEAD GmbH-Agency for Health Economic Assessment and Dissemination, Wilhelm-Leibl-Str. 7, D-74321 Bietigheim-Bissingen, Germany

³ Institute of Medical Biometry and Statistics (IMBI), University of Freiburg, Freiburg im Breisgau, Germany

⁴ Trimbos Institute-Netherlands Institute of Mental Health and Addiction, Utrecht, the Netherlands

⁵ a2m-Ars Accessus Medica, Amsterdam, the Netherlands

¹ Department of Health Services Research, CAPHRI-Care and Public Health Research Institute, Maastricht University, Maastricht, the Netherlands