A Systematic Review of methodologies used in models of the treatment of Diabetes Mellitus

Diabetes Cost-Effectiveness Systematic Review

Marina Antoniou, MSc

Division of Health Research

Lancaster University

Bailrigg, Lancaster

United Kingdom

m.antoniou2@lancaster.ac.uk

Ceu Mateus, PhD

Division of Health Research

Lancaster University

Bailrigg, Lancaster

United Kingdom

Bruce Hollingsworth, PhD

Division of Health Research

Lancaster University

Bailrigg, Lancaster

United Kingdom

Andrew Titman, PhD

Department of Mathematics and Statistics

Lancaster University

Bailrigg, Lancaster

United Kingdom

Statements and Declarations

Authors' contributions

MA: conceptualization and design, literature search and selection, quality assessment of included studies, data extraction, statistical analysis and interpretation of data, drafted and reviewed the manuscript. CM: conceptualization and design, reviewed the data, drafted and reviewed the manuscript. BH: conceptualization and design, reviewed and revised the manuscript. AT: reviewed the data generated, reviewed and revised the manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sector.

Data availability: All data analysed as part of this study are included in this published article (and its supplementary information files).

Competing of Interest: The authors declare that they have no competing interests.

Acknowledgements: We would like to thank the Editor and two anonymous reviewers for their comments that have much contributed to improve the final version of this paper.

Abstract

Background: Diabetes Mellitus is a chronic and complex disease, increasing in prevalence and consequent health expenditure. Cost-effectiveness models with long time horizons are commonly used to perform economic evaluations of diabetes' treatments. As such, prediction accuracy and structural uncertainty are important features in cost-effectiveness models of chronic conditions.

Objectives: The aim of this systematic review is to identify and review published cost-effectiveness models of diabetes treatments developed between 2011 and 2022 regarding their methodological characteristics. Further, it also appraises the quality of the methods used, and discuss opportunities for further methodological research.

Methods: A systematic literature review was conducted in MEDLINE and Embase to identify peer-reviewed papers reporting cost-effectiveness models of diabetes treatments, with time horizons of more than 5 years, published in English between 1st January 2011 and 31st of December 2022. Screening, full-text inclusion, data extraction, quality assessment and data synthesis, using narrative synthesis, were performed. The Philips checklist was used for quality assessment of the included studies. The study was registered in PROSPERO (CRD42021248999).

Results: The literature search identified 30 studies presenting 29 unique cost-effectiveness models of type 1 and/or type 2 diabetes treatments. The review identified 26 T2DM models, 3 T1DM models and one model for both types of diabetes. Fifteen models were patient-level models whereas 14 were at cohort level. Parameter uncertainty was assessed thoroughly in most of the models, whereas structural uncertainty was seldom addressed. All the models where validation was conducted performed well. The methodological quality of the models with respect to structure was high, whereas with respect to data modelling it was moderate.

Conclusions: Models developed in the past twelve years for health economic evaluations of diabetes treatments are of high-quality and make use of advanced methods. However, further developments are needed to improve the statistical modelling component of cost-effectiveness models and to provide better assessment of structural uncertainty.

Key Points for Decision Makers

i. Structural uncertainty in cost-effectiveness models is generated from all the decisions and assumptions being made during the development process of a model and can have a significant impact on the results.

ii. This systematic review identified cost-effectiveness models of diabetes treatments and discusses their methodological characteristics including uncertainty analysis methods.

iii. The diabetes models identified do not address structural uncertainty thoroughly and do not make use of the methods reported in the methodological literature to address and measure this type of uncertainty.

1. Introduction

Diabetes Mellitus is a chronic and complex disease, with increasing prevalence, across all age groups, and an increasing economic burden, making it a major public health problem worldwide [1, 2]. The prevalence of diabetes in 2019 was estimated to be 463 million people, 9.3% of the global population. More specifically, diabetes prevalence was 10.4% in high-income countries, 9.5% in middle-income countries and 4.0% in low-income countries; and it is expected to rise to 700 million people (10.9% of the global population) by 2045[3]. The global health expenditure associated with diabetes in 2019 was estimated to be USD 760 billion and by 2045 it is expected to reach USD 845 billion[4].

Pharmaceutical innovations are crucial to tackle diabetes and, in many jurisdictions, the submission of a health economic evaluation is required before a reimbursement decision is taken. Due to the chronic nature of the disease a lifetime horizon is commonly used in cost-effectiveness models. Randomised clinical trials are an important source of data used in models, however, as those have much shorter durations than the life of patients, extrapolation beyond the end of the follow-up period is unavoidable. Extrapolation adds uncertainty to the estimates obtained, which needs to be accounted for in model assessment. The five sources of uncertainty generated in cost-effectiveness (CE) models are: 1) stochastic uncertainty - the random variability in the outcomes between identical patients[5]; 2) heterogeneity - the variability, between patients, which can be attributed to the characteristics of the patients[5]; 3) parameter uncertainty - generated in the estimation of parameters of interest[5]; 4) structural uncertainty - generated from the assumptions inherent to the decision model[5]; 5) methodological uncertainty - generated from differences in the methodology that can be used in economic evaluation, such as the type of analysis, the perspective, valuation technique, discount rate and time horizon[6]. Structural uncertainty in a CE model can arise from various decisions and assumptions made during the development of the model, such as regarding comparators, relevant events, modelling on parameters and clinical uncertainty or lack of clinical evidence[7]. Multiple studies examining the impact of structural uncertainty on CE models found that choices regarding health states or specification of the transition rates led to significant differences in estimated outcomes[8-11]. Structural uncertainty is at least as important as parameter uncertainty[12] and may have a much greater impact on results than parameter uncertainty[6].

The Mount Hood Diabetes Challenge Network (https://www.mthooddiabeteschallenge.com/registry) runs diabetes computer simulation modelling conferences and has a registry of T1DM and T2DM simulation models. During the

Fourth Mount Hood Challenge[13] and the Fifth Mount Hood Challenge[14] multiple participating diabetes modellers performed validation exercises and the results were discussed with respect to the challenges and future models' improvements. These two challenges asked the participating modellers to perform simulations based on published clinical trials to compare each model against the clinical data to assess predictive accuracy regarding clinical trial data (i.e., external validation); and predictive performance of the models (i.e., cross-validation). In the Ninth Mount Hood Challenge[15] participants were tasked with assessing the magnitude of impact of structural uncertainty on life-years (LYs) and quality-adjusted life-years (QALYs) of 11 type 2 diabetes models. The findings pointed towards substantial cross-model variability in QALY predictions for a standardised set of simulation scenarios which was considerably larger than within-model variability to alternative health state utility values, emphasizing the need to address structural uncertainty in model-based analyses[15].

There are several relevant published systematic reviews that identified health economic models for the treatment of type 1 and/or type 2 diabetes mellitus and discussed their methods [16, 17] [18] [19] [20] [21] [22] [23]. The results of this systematic review will add to existing work by looking at uncertainty analysis methods and model validation applied in pharmaceutical interventions, in the past twelve years, considering only models with time horizons of more than 5 years.

The aims of this systematic review are (1) to identify CE models and model-based health economic evaluations of pharmaceutical treatments of T1DM and T2DM, published between 2011 and 2022 using time horizons above 5 years; (2) to describe structural characteristics of the decision models such as the type of the model, simulation method, diabetes related complications, model outcomes, discounting rate and time horizon; (3) to describe and discuss the methods used for addressing all five types of uncertainty and identify opportunities for future methodological research; (4) to summarize the methods and results of internal and external validation of the decision models; and (5) to assess the methodological quality of the models using best practice guidelines.

2. Methods

This systematic review is based on the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement[24]. The protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42021248999).

2.1. Data Sources and Searches

Literature searches were conducted, on 11 March 2023 in two biomedical databases, Ovid MEDLINE (Table S2) and Ovid Embase (Table S3), to identify relevant studies published between 1 January 2011 and 31 December 2022. The search terms for diabetes are given in the Supplementary Material and were based on a published systematic literature review[18] and the search terms for economic modelling were based on a published systematic review protocol[25]. Hand-searching the reference lists of the included studies to identify additional literature was used to supplement the electronic database searches. The search was limited to papers published in English.

2.2. Study Selection Criteria and Study Selection

The review included studies performing a model-based economic evaluation with a description of the model development; and modelling studies describing a health economic model for a time horizon above 5 years, for cost-effectiveness or cost-utility analysis, and including validation of the model. Only pharmacological interventions for the treatment of diabetes were considered to allow for comparison of the same type of intervention. As pharmacological interventions are commonly evaluated using RCTs, it allows for a discussion of statistical modelling along with the associated uncertainty of extrapolating clinical outcomes. A table with the detailed inclusion and exclusion criteria can be found in Table S4.

For models associated with more than a single publication, the publication included here was the one that described the model in the most detail and any subsequent publications were included only if they reported updates to the methods used to devise the model.

All the references produced from the searches were downloaded and managed in EndNote and duplicates were removed. All the titles and abstracts were screened against the eligibility criteria by a single reviewer and a second reviewer assessed 20% of the references. The titles/abstracts that did not satisfy the eligibility criteria were excluded whereas those that fully or partially satisfied the criteria proceeded to full-text screening. After reviewing the full-text references, only the studies that satisfied the inclusion criteria were considered.

2.3. Data Extraction

Data were extracted from the included studies using a data extraction form in Microsoft Excel (Table S5) developed based on the data extraction form of a published systematic review[19]. Data extraction was performed by a single reviewer while a second reviewer also assessed 20% of the included studies.

2.4. Quality Assessment

The Phillips checklist[26] was used to assess the methodological quality of the studies included in the systematic review. Furthermore, the recommendations and description for best practices provided in the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making (ISPOR-SMDM) Modelling Good Research Practices[27] were considered for the general critical appraisal of the identified decision models throughout this systematic review. It is also worth mentioning the checklist that Palmer et al. developed for transparency of input data specific for diabetes models that can be used alongside general health economic modelling guidelines[28]. A single reviewer assessed the quality of the included studies with a second reviewer assessing 20% of the included studies.

2.5. Data Synthesis

The extracted data were synthesised following the narrative synthesis framework developed by Popay and colleagues[29]. This framework involves (1) developing a theoretical model of how the interventions work, why and for whom; (2) developing a preliminary synthesis; (3) exploring the relationships in the data; and (4) assessing the

robustness of the synthesis output. T1DM and T2DM models are discussed and summarised together as the differences in the disease do not impact on uncertainty assessment or on the methods of model validation.

3. Results

3.1. Systematic Review

The systematic literature search identified 30 peer-reviewed studies published between 2011 and 2022 that developed 29 cost-effectiveness (CE) models for pharmacologic treatment of T1DM and/or T2DM with a time horizon above 5 years. The selection process and results of the systematic review are presented on Figure 1. Briefly, the initial database searching in MEDLINE and Embase identified 1795 studies. After removing 498 duplicates, the titles and abstracts of the remaining 1297 studies were screened. At this stage, 1251 studies were excluded as it was clear from the title or abstract that the inclusion criteria were not met. In total, 46 studies were eligible for full text screening out of which 28 satisfied the eligibility criteria and were included in the systematic review. From searching the reference lists of the included studies, 2 additional studies were identified and included in the systematic review. From the 30 identified studies, 17 studies are model-based economic evaluations developing a model to perform a cost-effectiveness or cost-utility analysis and 13 studies are modelling studies that developed and validated a decision model.

Table 1. Main characteristics of the 29 cost-effectiveness models included in the systematic review.

Author/Year	Diabetes Type	Type of Model	Simulation method	Time Horizon (years)	Model Outcomes	Diabetes related complication	Software
Hayes et al. 2013 (UKPDS-OM2) [30]	Type 2	Microsimulation	Patient-level	Lifetime	Life expectancy; QALYs; annual incidence of death or complications	first MI; second MI; first stroke; second stroke; CHF; IHD; first amputation; second amputation; blindness; renal failure; ulcer	Stata (for the statistical analysis)
Lundqvist et al. 2014 (IHE) [31]	Type 2	Markov	Cohort	40	Survival; LYs; QALYs; costs	BDR; PR; ME; ME and PDR; SVL; symptomatic neuropathy; PVD; LEA; Post LEA; microalbuminuria; macroalbuminuria; ESRD; IHD; MI; stroke; CHF	Microsoft Excel
Viriato et al. 2014 [32]	Type 2	Microsimulation	Patient-level	40	Costs; QALYs; Cumulative incidence; LYs	IHD; MI; CHF; renal failure; stroke; lower limp amputation; blindness in one eye; symptomatic hypoglycemia; severe hypoglycemia; weight gain; death following an event in the first year; diabetes-related death following a diabetes-related event; death due to other causes	Microsoft Excel and Visual Basic
Van der Heijden et al. 2015 (MICADO) [33]	Type 1 and Type 2	Markov	Cohort		Incidence and prevalence of complications; costs; utilities	CHD; stroke; chronic heart failure; MI; diabetic foot; foot ulcer; foot abscess; amputation; microalbuminuria; macroalbuminuria; ESRD; background retinopathy; macular oedemia; proliferative retinopathy; blindness	Mathematica
Wolowacz et al. 2015 [34]	Type 1	Microsimulation	Patient-level	Lifetime	Costs; LYs; QALYs	angina; MI; revascularization; stroke; cataract surgery; peripheral neuropathy; foot ulcer; amputation; microalbuminuria; ESRD; PDR; blind; hypoglycemia; diabetic ketoacidosis	Microsoft Excel
Ye et al. 2015 (Michigan) [35]	Type 2	Microsimulation	Patient-level	Lifetime	Costs; utilities	CAD; CHD; MI; CHF; repeat MI; short term survival following MI; CHD death	Python

Author/Year	Diabetes Type	Type of Model	Simulation method	Time Horizon (years)	Model Outcomes	Diabetes related complication	Software
Valentine et al. 2016 (PRIME) [36]	Type 1	Microsimulation	Patient-level	Lifetime	Life expectancy; QALE; cumulative incidence of all modelled diabetes complications; mean hypoglycemia and ketoacidosis rate; costs	MI; angina; stroke; HF; microalbuminuria; overt nephropathy; ESRD; neuropathy onset; amputation; mild NPDR; moderate NPDR; severe NPDR; proliferative diabetic retinopathy, blindness, macular edema; hypoglycemia; ketoacidosis	Java
Willis et al. 2013 & 2017 (ECHO-T2DM) [37, 38]	Type 2	Microsimulation	Patient-level	User definable	Mean survival; LYs; QALYs; costs;	IHD; MI; CHF; stroke; BDR; PDR; PDR&Blind; ME; ME&PDR; ME&Blind; ME&PDR&Blindness in 1 eye; blindness in both eyes; MA; GPR; ESRD; symptomatic neuropathy; PVD; Symptomatic/PVD; foot ulcer; LEA; subsequent LEA	R with Excel Interface
Kwon et al. 2018 [39]	Type 2	Markov	Cohort	25	LYs; costs	Hypoglycemia; MI; HF; stroke; weight gain; death	Microsoft Excel
Laiteerapong et al. 2018 [40]	Type 2	Microsimulation	Patient-level	Lifetime	Life expectancy; QALYs; costs	Amputation; second amputation; blindness; CHF; ESRD; IHD; MI, second MI; stroke; second stroke; foot ulcer; second foot ulcer; hypoglycemia	Excel and SAS
Nguyen et al. 2018 [41]	Type 2	Markov	Cohort	38	Costs; QALYs	NYHA I/II HF; NYHA III/IV HF; HF; MI; unstable angina; TIA; vascular disease; stroke; ESRD; severe hypoglycemia; fatal stroke; fatal MI; death from ESRD; CV death; NYHA I/II HF death; NYHA III/IV HF death; all-cause mortality	TreeAge Pro 2009
Shao et al. 2018 (BRAVO) [42]	Type 2	Microsimulation	Patient-level	Lifetime	Life expectancy; risks of different events; costs; QALYs	MI; CHF; stroke; angina; revascularization; blindness; ESRD; SPLS	Visual Basic and C++

Author/Year	Diabetes Type	Type of Model	Simulation method	Time Horizon (years)	Model Outcomes	Diabetes related complication	Software
Wu et al. 2018 (COMIT) [43]	Type 2	Microsimulation	Patient-level	Lifetime	LYs; QALYs; DALYs; Cumulative incidences of complications; Costs	Stroke; MI; CHF; CVD; CVD death; Blindness; ESRD; Clinical Neuropathy; Uncomplicated Diabetic Foot; Complicated Diabetic Foot; Minor Amputation; Major Amputation, ASCVD	R
Abramson et al. 2019 [44]	Type 2	Microsimulation	Patient-level	Lifetime	Costs; QALE	Background mortality; diabetes-related mortality	
Chin et al. 2019 [45]	Type 2	Markov	Cohort	20	Costs; LYs; QALYs	MI non fatal; stroke; HF; CVD death; non-CVD death;	Microsoft Excel and @Risk
Kansal et al. 2019 [46]	Type 2	Discrete Event	Patient-level	Lifetime	Cumulative events per 100 patient-years; Life expectancy QALYs; costs	MI; stroke; angina; HF; TIA; revascularization; macroalbuminuria; renal injury; renal failure	R
Kazemian et al. 2019 (PREDICT-DM) [47]	Type 2	Microsimulation	Patient-level	10	5/10 year survival; Cardiovascular outcomes; Renal outcomes	MI; stroke; CHF; mortality from CVD; nephropathy	Python
Pollock et al. 2019 [48]	Type 2	Microsimulation	Patient-level	40	QALE; costs	CHF; IHD; renal failure; ulcer; blindness; MI; stroke; amputation	Java 8
Su et al. 2020 (Cornerstone) [49]	Type 2	Microsimulation	Patient-level	User specified (up to 100 years)	LYs; QALYs; ICER	CHF; IHD; MI; stroke; blindness; ulcer; amputation; renal failure; mortality	Microsoft Excel
Tran-Duy et al. 2020 [50]	Type 1	Microsimulation	Patient-level	Lifetime	Annual incidence of complications and death; time to events; changes in risk factors over the simulation time	Fatal MI; nonfatal MI; fatal stroke; nonfatal stroke; HF; PVD; severe hypoglycemia; severe hyperglycemia; amputation; ESRD; PCI; CABG	Stata

Author/Year	Diabetes Type	Type of Model	Simulation method	Time Horizon (years)	Model Outcomes	Diabetes related complication	Software
Wu et al. 2020 [51]	Type 2	Markov	Cohort	Lifetime	Costs; probability of MI; probability of IS; probability of TIMI major bleeding; LYs; QALYs	MI; IS; ICH; TIMI major; ECH; death	
Bagepally et al. 2021 [52]	Type 2	Markov	Cohort	Lifetime	LYs; QALYs; Costs	Hypoglycemia; MI; HF; Stroke; Genital Infection	Excel
Bekele et al. 2021 [53]	Type 2	Markov	Cohort	40	DALYs; Costs	Uncontrolled T2DM; Complicated T2DM; Death form T2DM	TreeAge Pro 2020
Deerochanawong et al. 2021 [54]	Type 2	Markov	Cohort	Lifetime	QALYs; Costs	HF; Death from HF; Normoalbuminuria; Microalbuminuria; Macroalbuminuria; Elevated Serum Creatinine; ESRD; Death from CKD; Death	Excel
Tanaka et al. 2021 (JJCEM) [55]	Type 2	Markov	Cohort	Lifetime	QALYs; costs	Retinopathy; Retinopathy progression; Coronary heart disease; Stroke; Overt nephropathy; Non-CVD mortality; Amputation; Event-related mortality; ESRD	
Abushanab et al. 2022 [56]	Type 2	Markov	Cohort	Lifetime	QALYs; YoLS; Costs	Non-fatal MI; Non-fatal stroke; HF; Unstable angina; CVD death; Non-CVD death	Excel
Huang et al. 2022 [57]	Type 2	Markov	Cohort	30	QALYs; Costs	MI; IS; Unstable Angina; HF; CVD death; Non-CVD death	TreeAge 2019
Peng et al. 2022 [58]	Type 2	Markov	Cohort	10	QALYs; Costs	HF; MI; Stroke; all-cause mortality	
Steg et al. 2022 [59]	Type 2	Markov	Cohort	Lifetime	QALYs; Costs	Non-fatal MI; Non-fatal Stroke; TIMI major bleeding; TIMI minor bleeding; Bleeding requiring medical attention; Dyspnea; All-cause mortality; Amputation	

ASCVD indicates Arteriosclerotic Cardiovascular Disease; BDR Background Diabetic Retinopathy; CABG Coronary Artery Bypass Graft; CAD Coronary Artery Disease; CHD Coronary Heart Disease; CHF Congestive Heart Failure; CKD Chronic Kidney Disease; CVD Cardiovascular Disease; DALYs Disability-Adjusted Life Years; ECH Extracranial Haemorrhage; ESRD End-stage Renal Disease; GPR Gross Preoteinuria; HF Heart Failure; ICH Intracranial Haemorrhage; IHD Ischemic Heart Disease; IS Ischemic Stroke; LEA Lower Extremity Amputation; MA Microalbuminuria; ME Macular Edema; MI indicates Myocardial Infarction; NPDR Non-proliferative Diabetic Retinopathy; NYHA I/II or NYHA III/IV New York Heart Association; PCI Percutaneous Coronary Intervention; PDR Proliferative Diabetic Retinopathy; PR Proliferative Retinopathy; PVD Peripheral Vascular Disease; QALYs Quality-Adjusted Life Years; SPLS Severe Pressure Sensation Loss; SVL Severe Vision Loss; TIA Transient Ischemic Attach; TIMI Thrombolysis In Myocardial Infarction; YoLS Years of Life Saved

1.1. Model Structure

The main characteristics of the 29 identified decision models are summarised in Table 1. All three T1DM models identified in the systematic review are microsimulation models [34, 36, 50]. From the 26 T2DM models, 14 are cohort Markov models [31, 33, 39, 41, 45, 51-59], 11 are patient-level microsimulation models [30, 32, 35, 37, 38, 40, 42-44, 47-49], and 1 is a patient-level discrete-event models [46]. One model developed for T1DM and T2DM is a Markov model [33]. Microsimulation models and discrete-event simulations are more flexible modelling methods that incorporate individual patients' disease histories and characteristics. Even though patient-level simulation is more computationally demanding than cohort Markov models, they are widely used in diabetes modelling because they have the capability of generating more accurate results for a wider range of patient subgroups and hence provide a more robust evaluation of overall uncertainty.

The diabetes complications included in each model are presented on Table 1. One model[44] included only background mortality and diabetes related mortality as events and another model consisted uncontrolled T2DM, complicated T2DM and death from T2DM as states with diabetes-related complications[53]. All the other models included macrovascular complications and 19 models [30-34, 36-38, 40-43, 46-50, 54, 55, 59] also included a combination of microvascular complications (i.e., nephropathy and/or neuropathy and/or retinopathy). The most frequently included retinopathy complication was blindness; for neuropathy complications the most frequently included was amputation; while end-stage renal disease (ESRD) and microalbuminuria were the most frequently included nephropathy complications. The models included various macrovascular complications and it is also worth noting that myocardial infarction (MI) was included in all the models apart from Deerochanawong et al. 2021[54] and Tanaka et al. 2021[55].

Apart from QALYs/LYs there is a variety of other clinical model outcomes considered in the studies reviewed, such as the cumulative incidence of events including MI, IS, TIMI major bleeding, or other complications such as hypoglycaemia or ketoacidosis. Therefore, many CE models allow for the computation of various clinical outcomes. Two studies ([33, 35]) report utilities that can be used to compute QALYs by their multiplication by the years spent in certain health states[60].

A given discount rate is specified in a number of CE models [32, 34, 39-41, 43-46, 48, 51-54, 56-59, 61], but not in the remaining models, allowing the users to set it up to meet existing guidance.

1.2. Uncertainty Analysis

Information related to uncertainty analysis is summarised in Table 2a and Table 2b [30-59]. Stochastic uncertainty is addressed in seven (out of the 15) patient-level CE models[30, 37, 38, 42, 46-48, 50]. The approach for minimising this type of uncertainty was by using large numbers of Monte Carlo replications until the mean value of the outcomes of interest changes less than a pre-specified threshold. Heterogeneity is addressed by measuring the impact that different patient characteristics can have on model outcomes; and it is examined with sensitivity analyses by running the CE model for different subgroups of patients with certain common characteristics. From the 14 cohort Markov models, three CE models[51, 56, 59] performed subgroup analysis to assess the impact of specific patient characteristics on model results. From the 15 patient-level models, two models[34, 40] conducted subgroup analysis to deal with heterogeneity. However, it is worth mentioning that the UKPDS-OM2 model reports that the detailed modelling at a patient level has the capacity to inform individualised medicine and analyses for patient subgroups[30]; and the BRAVO model mentions that heterogeneity in the model was dealt with by using patient-level microsimulation, in which each patient has different characteristics and simulation was carried out one person at a time[42].

All the CE models assessed parameter uncertainty, except the Cornerstone model[49] and the COMIT model[43]. Twenty-three CE models performed Probabilistic Sensitivity Analysis (PSA) [31-39, 41, 44-48, 51-59], whereas five CE models used a bootstrapping technique instead [30, 36, 42, 48, 50]. The non-parametric bootstrapping approach involved repeatedly resampling the patients and re-estimating all the risk equations in order to derive sets of fully correlated regression coefficients for each risk equation, and hence generating a distribution for the CE model outcome(s) which is used to capture parameter uncertainty [30, 42, 50]. Furthermore, eighteen CE models performed Deterministic Sensitivity Analysis (DSA)[30, 32-34, 39-41, 44, 45, 50-54, 56-59], i.e., one-way sensitivity analyses and/or two-way sensitivity analyses to examine the effect of varying the values of selected parameters one by one on CE model results.

Structural uncertainty is examined in five CE models by using scenario analyses. For example, one of the scenarios that Pollock et al. 2019 [48] and Kansal et al. 2019 [46] tested setting the treatment effect to zero for several diabetes-related events to examine the impact on the results. Furthermore, Kwon et al. 2018 examined the impact of no difference in cardiovascular event rates after 2 years from initiations of the two interventions[39]. Nguyen et al. 2018 tested a scenario in which the overall 3-month stroke rate in the trial was set as the rate of stroke for both groups in

the model[41]. Lastly, Peng et al. 2022[58] tested several different scenarios by varying assumptions based on patient cohorts risk score for mortality, range of CVD risks and adopting the hazard ratios from a network meta-analysis of clinical trials[58]. Structural uncertainty associated with statistical modelling is discussed in Section 3.4.

Methodological uncertainty was examined in nine CE models [32, 39-41, 46, 48, 53, 56, 58] by performing one-way sensitivity analyses of varying methodological assumptions, such as running different scenarios of alternative discount rates and/or time horizons than the ones used in the CE model.

In summary, Table 2a shows that (1) almost all the CE models addressed parameter uncertainty, and specifically from Table 2b, PSA is the most widely used method among the CE models; and (2) the impact of structural uncertainties is not thoroughly examined in most of the CE models.

1.3. Statistical Modelling and Structural Uncertainty

In this section the methods used in the patient-level decision models for predicting the risk of each event are described and a summary can be found in Table 4.

Five out of twelve T2DM patient-level decision models used the UKPDS-OM risk equations [30, 32, 40, 48, 49]. The UKPDS-OM risk equations were developed using data from the UKPDS study population [62, 63]. Three decision models used the UKPDS-OM risk equations in combination with other sources [35, 37, 38, 44]. For instance, the latest version of the ECHO-T2DM model enables the user to choose between the UKPDS-OM, or the ADVANCE [64], or the Swedish NDR [65] risk equations for calculating the hazard of macrovascular complications, and for the microvascular complications other sources from the literature were used [66, 67]. Also, the Michigan model[35] used the UKPDS-OM risk equations to model[30] all its events apart from congestive heart failure (CHF) for which the authors created a new prediction equation based on the CHS data [68]. In one model-based economic evaluation the data from a clinical trial (EMPA-REG OUTCOME) were used to develop the risk equations[46]. Lastly, the PREDICT-DM model[47] and the COMIT model[43] used the RECODe risk equations[69]; and the BRAVO decision model developed its own risk equations[42]. Both the RECODe and BRAVO risk equations were developed based on data from the ACCORD trial[70].

Risk equations for complications and mortality calculate the event probability in a particular period for microsimulation models or the time-to-event for DES models. The UKPDS-OM risk equations [30] and the BRAVO risk equations[42] estimate/predict the hazard of each of the diabetes-related complications and all-cause mortality by using multivariate parametric proportional hazards (PH) models (Weibull PHs models for the complications and Gompertz PHs models for all-cause mortality); and diabetes-related mortality (whether mortality is a CVD death or not, for example) by using logistic regression models. Similarly, the CHF risk equation in the Michigan model [35] and the risk equations of the study by Kansal and colleagues [46] were developed by using parametric proportional hazards models. More specifically, this modelling approach assumes multivariate semi-parametric (Cox) proportional hazards models, for selecting and estimating the risk factors, but with specifying a parametric baseline hazard for enabling lifetime prediction of event times. This approach is explained in more detail in the literature[71]. On the other hand, the RECODe risk equations were developed by using Cox proportional hazards models, and therefore, without further assumptions the hazard and event times can only be predicted to a maximum of 10 years, that is the duration of the ACCORD data used to devise the equations, rather than to a longer term[72, 73].

Two out of the three T1DM patient-level decision models used risk equations to model the hazard of complications and mortality[34, 50]. In the other T1DM model (PRIME [36]) the event probabilities were derived using multiple sources from the literature.

Risk equations varied inclusion of time-varying risk factors and use of intermediate or final endpoints.

Structural uncertainty in survival modelling defined as the choice between multiple plausible statistical models[74, 75] and data-driven model selection was the predominant method used for their choice when estimating survival probabilities. Model selection was performed by comparing model fits from multiple parametric forms, such as exponential, Weibull, Gompertz log normal, and then picking the model to be implemented based on a measure of model performance (such as AIC, BIC, and log-cumulative hazards plot). Similarly, model selection in the development of the BRAVO risk equations[42] also involved testing multiple parametric functional forms for the baseline hazard, however, the statistical model was selected based on the c-statistic – a measure of discrimination power – and the Brier score – a measure of prediction accuracy. The RECODe risk equations in the PREDICT-DM model[47] also used measures that assess both model discrimination and model calibration. Finally, it is important to note that the PRIME model[36] for T1DM was the only model that used a model averaging approach to combine

multiple data sources from the literature to calculate the risk of cardiovascular disease. There was no evidence of the use of any of the other methods, introduced in the methodological literature, for accounting structural uncertainty, with regards to extrapolation models, such as scenario analysis, parameterisation or the discrepancy approach.

In contrast with the uncertainty analysis reviewed in the previous section, structural uncertainty was not precisely addressed in the papers included and no methods were mentioned to specifically account for structural uncertainty.

1.4. *Model Validation*

In fourteen CE models internal validation was performed [30, 31, 34-38, 40, 42, 47-50]. It was found that, overall, the predicted model values were in close agreement with the observed/published values. More specifically, five CE models stated that the predicted values were generally within the 95% confidence intervals of the observed cumulative curves for diabetes-related events and mortality [30, 40, 42, 49, 50]. In other CE models the authors calculated coefficients of determination and slopes close to 1, indicating a good match between predicted and observed values [31, 35, 37, 38]; and / or calculated measures of error, which were less than 2%-3% again indicating a close match between the predicted and observed values [36, 47, 48]. Lastly, Wolowacz et al. model[34] compared the predicted values with the expected values; and Kansal et al.[46] and Abramson et al. [44]verified that the model estimates matched the data.

Eleven studies conducted external validation of the CE model (Table 4)[31, 33-38, 42, 43, 47, 49]. External validation is carried out to assess the performance of the CE model with independent/external data, i.e., data that were not used to develop the model. All the models' predictive performance was satisfactory and obtained good agreement in the comparisons between the predicted outcomes and the actual observed outcomes. From the models that used the scatterplot and linear regression line method for evaluating concordance between observed and predicted values, the ones that had slopes and coefficients of determination close to perfect concordance are ECHO-T2DM 2013[37], ECHO-T2DM 2017[38], and IHE[31]. It is also worth noting that the ECHO-T2DM (2017) model[38] and the IHE Cohort model[31] performed external validations for multiple sets of risk equations (UKPDS-OM1, UKPDS-OM2, and the Swedish NDR) where it was found that the results were slightly different for each set of risk equations. The other models that used alternative methods of evaluating the accuracy of predicted results, such as the MICADO[33], PRIME[36], PREDICT-DM[47] and the model by Wolowacz et al.[34], all could simulate model outcomes well.

1.5. Quality Assessment

Results of quality assessment are summarized on Figures 2 and 3. Overall, 90% of the models had a medium score [30-45, 47, 49-55, 57, 59]and 10% had a high score[46, 48, 56, 58]. All 29 models scored high in the structure section of the Philips checklist. In the data section of the checklist 17% of the models scored low[35, 44, 45, 47, 49], 70% of the models scored medium[30-34, 36-43, 50-55, 57, 59] and 13% of the models scored high [46, 48, 56, 58]. Four models had the highest scores in the data section ([48], [46],[56] [58]) and highest scores overall. Three of these models performed scenario analyses to investigate the impact of structural uncertainties on model results ([48], [46], [58]). In the consistency section of the checklist 63% of the models scored low[30, 32, 33, 39-41, 45, 46, 48, 50-59], 23% of the models scored medium [31, 34-36, 42, 43, 47] and 13% of the models scored high[37, 38, 44, 49]. Low scores in the consistency section report to models where authors did not conduct internal and/or external validation. Two models performed and reported the results of internal validation, external validation, and cross-validation (Cornerstone[49], ECHO-T2DM[37, 38]); and therefore, these models satisfied all the points in the consistency section.

2. Discussion

2.1. Findings

Diabetes models with a time horizon above 5 years developed since 2011 used advanced statistical methods regarding their structure. Patient-level simulation is more computationally demanding and more complex than cohort models, however, they are commonly used in diabetes modelling [30], [13, 14]. Therefore, in a complex disease such as diabetes the additional modelling flexibilities of a microsimulation model and discrete-event simulation may be useful for representing the disease more precisely. Furthermore, many decision models include both microvascular and macrovascular diabetes-related events, and allow for the computation of varied model outcomes. From the Philips checklist scores it can be concluded that most of the publications are transparent in their description of the model development process with respect to its structure.

Uncertainty analysis methods, such as PSA, were used in most of the models reviewed to address parameter uncertainty in a similar manner. Stochastic uncertainty was addressed in seven out of 14 patient-level models ([30, 37, 38, 42, 46-48, 50]). Structural uncertainty was addressed with sensitivity analyses/scenario analyses in only five models ([48], [46], [39], [41], [58]), indicating the need for guidelines on the methods for addressing structural uncertainty so they can be more broadly and consistently applied.

One important source of structural uncertainty is generated by choices made for survival extrapolation. Apart from the PRIME model[36] where a model averaging method was used for calculating the risk of MI, angina and stroke, no other methods were used in any of the other models for addressing structural uncertainty from survival modelling.

In studies undertaking CEAs - model-based economic evaluations – the authors did not perform any external validation, and in only four of them performed internal validation ([48], [46], [40], [44]). Internal and external validation tests were mainly carried out in the other modelling studies. Most of the papers that conducted internal and/or external validation, did so in a similar manner, and provided a description of the methods used coupled with a thorough interpretation of their results. The methods or measures used for comparing the predicted values, simulated by the models, with the observed values, from the external data, varied. Different studies arrived at often quite different model structures for the same, or similar patients. This is despite all the models performing well in their internal and external validations, implying the models generate accurate results and are of high quality. Potentially, this highlights the importance of addressing structural uncertainty as it indicates that there may have been competing model structures which gave similar, or better, fits to the data.

"Structural uncertainty" was not widely used in the publications reviewed, which makes it harder to understand if it was assessed and by which methods. Also to be noted is the inexistence of a single measure of prediction accuracy throughout the publications to compare the internal and external validation results of all the models that conducted validation.

Quality assessment of the 29 models was carried out by using the Philips checklist. In line with the findings in model validation, all the models scored reasonably well in the methodological quality assessment. Overall, the majority of the models (90% of the models) satisfied between 51%-74% of the points in the Philips checklist. However, after distinguishing the section about the "structure" and the section about the "data", it was apparent that all the models achieved significantly higher scores in the structure section than in the data section. Most of the models did not address

structural uncertainty and did not document or performed sensitivity analysis regarding the continuing effect of treatment or regarding the methods used for extrapolating data.

2.2. Limitations

This systematic review has some limitations. Firstly, only journal articles written in English were included, and therefore, models presented in conference abstracts or models published in languages other than English were not included. Secondly, the Philips checklist for assessing the methodological quality of the models has some limitations. The responses to many criteria were subjective and may also be characterised as overly general to assess models in Diabetes Mellitus which is a complex disease[28]. Furthermore, this systematic review discussed structural uncertainty mainly with regards to survival extrapolation. However, it is important to recognise that structural uncertainty in a cost-effectiveness model is generated from various sources, such as the set of adverse events included within the overall model or the cycle length used for certain events for instance, and needs to be addressed[7].

2.3. Recommendations for Further Research

This systematic review found that guidance on methods to assess uncertainty related to statistical modelling could improve the quality of the models being developed and contribute to an improved decision making process. Structural uncertainty seems to have received little attention from researchers and it would benefit from further methodological and empirical research. Some authors have argued that model selection does not address structural uncertainty regarding the statistical model for extrapolation and hence is seen as an unsuitable method[7, 76]. However, the use and applicability of other more recent methods, such as model averaging, has not been explored or widely applied in diabetes health economic evaluations yet. There are other authors suggesting the use of the Bayesian framework [74, 77, 78] to improve current methods but too little has been done specifically in diabetes.

3. Conclusions

Accuracy and reliability are key goals in modelling in health economic evaluations and are highly dependent on the methodologies used for developing a model. Diabetes is a complex and chronic disease which mostly requires the use of a model to capture the various disease paths, calculate long-term outcomes from clinical trials data or final outcomes

from intermediate endpoints, and so on. The accuracy of the results of the models and the confidence in reimbursement decisions for efficient allocation of scarce resources depend on the structure of the model, the methods used for estimating and extrapolating the model parameters, and the uncertainty analysis methods used.

4. References

1. Mayer-Davis, E.J., et al., *Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002–2012.* N Engl J Med, 2017. **376**(15): p. 1419-1429.
2. Xie, J., et al., *Global burden of type 2 diabetes in adolescents and young adults, 1990-2019: systematic analysis of the Global Burden of Disease Study 2019.* BMJ, 2022. **379**: p. e072385.
3. Saeedi, P., et al., *Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition.* Diabetes research and clinical practice, 2019. **157**: p. 107843-107843.
4. Williams, R., et al., *Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th edition.* Diabetes Res Clin Pract, 2020. **162**: p. 108072-108072.
5. Briggs, A.H.D., et al., *Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6.* Value Health, 2012. **15**(6): p. 835-842.
6. Brisson, M. and W.J. Edmunds, *Impact of Model, Methodological, and Parameter Uncertainty in the Economic Analysis of Vaccination Programs.* Med Decis Making, 2006. **26**(5): p. 434-446.
7. Bojke, L.M., et al., *Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods.* Value Health, 2009. **12**(5): p. 739-749.
8. Frederix, G., et al., *The impact of structural uncertainty on cost-effectiveness models for adjuvant endocrine breast cancer treatments: The need for disease-specific model standardization and improved guidance.* Pharmacoeconomics, 2014. **32**(1): p. 47-61.
9. Le, Q.A., *Structural Uncertainty of Markov Models for Advanced Breast Cancer: A Simulation Study of Lapatinib.* Med Decis Making, 2016. **36**(5): p. 629-640.
10. Silva-Illanes, N. and M. Espinoza, *Critical Analysis of Markov Models Used for the Economic Evaluation of Colorectal Cancer Screening: A Systematic Review.* Value Health, 2018. **21**(7): p. 858-873.
11. von Schéele, B., et al., *Relationship between modeling technique and reported outcomes: case studies in models for the treatment of schizophrenia.* Expert review of pharmacoeconomics & outcomes research, 2014. **14**(2): p. 235-257.
12. Kim, L.G. and S.G. Thompson, *Uncertainty and validation of health economic decision models.* Health Econ, 2010. **19**(1): p. 43-55.
13. Palmer, A.J., et al., *Computer modeling of diabetes and its complications: A report on the Fourth Mount Hood Challenge Meeting.* Diabetes Care, 2007. **30**(6): p. 1638-1646.
14. Palmer, A.J.B.M., *Computer Modeling of Diabetes and Its Complications: A Report on the Fifth Mount Hood Challenge Meeting.* Value Health, 2013. **16**(4): p. 670-685.
15. Tew, M., et al., *Exploring Structural Uncertainty and Impact of Health State Utility Values on Lifetime Outcomes in Diabetes Economic Simulation Models: Findings from the Ninth Mount Hood Diabetes Quality-of-Life Challenge.* Med Decis Making, 2022. **42**(5): p. 599-611.
16. Tarride, J.-E., et al., *A Review of Methods Used in Long-Term Cost-Effectiveness Models of Diabetes Mellitus Treatment.* PharmacoEconomics, 2010. **28**(4): p. 255-277.
17. Yi, Y., et al., *Economic models in type 2 diabetes.* Curr Med Res Opin, 2010. **26**(9): p. 2105-2118.
18. Charokopou, M., et al., *Methods applied in cost-effectiveness models for treatment strategies in type 2 diabetes mellitus and their use in Health Technology Assessments: a systematic review of the literature from 2008 to 2013.* Curr Med Res Opin, 2016. **32**(2): p. 207-218.

19. Henriksson, M., et al., *A Systematic Review of Cost-Effectiveness Models in Type 1 Diabetes Mellitus.* PharmacoEconomics, 2016. **34**(6): p. 569-585.
20. Li, J., et al., *Decision models in type 2 diabetes mellitus: A systematic review.* Acta Diabetol, 2021. **58**(11): p. 1451-1469.
21. Usman, M., et al., *Cost-effectiveness of intensive interventions compared to standard care in individuals with type 2 diabetes: A systematic review and critical appraisal of decision-analytic models.* Diabetes Res Clin Pract, 2020. **161**: p. 108073-108073.
22. Asche, C.V., S.E. Hippler, and D.T. Eurich, *Review of Models Used in Economic Analyses of New Oral Treatments for Type 2 Diabetes Mellitus.* Pharmacoeconomics, 2014. **32**(1): p. 15-27.
23. Ogurtsova, K., et al., *External validation of type 2 diabetes computer simulation models: Definitions, approaches, implications and room for improvement-a protocol for a systematic review.* Syst Rev, 2017. **6**(1): p. 267-267.
24. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* BMJ, 2009. **339**(jul21 1): p. b2535-b2535.
25. Karagiannidou, M., et al., *Systematic literature review of methodologies and data sources of existing economic models across the full spectrum of Alzheimer's disease and dementia from apparently healthy through disease progression to end of life care: a systematic review protocol.* BMJ Open, 2018. **8**(6): p. e020638-e020638.
26. Philips, Z., et al., *Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment.* PharmacoEconomics, 2006. **24**(4): p. 355-371.
27. Caro, J.J.M.F.F., et al., *Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1.* Value Health, 2012. **15**(6): p. 796-803.
28. Palmer, A.J., et al., *Computer Modeling of Diabetes and Its Transparency: A Report on the Eighth Mount Hood Challenge.* Value Health, 2018. **21**(6): p. 724-731.
29. Popay, J., et al., *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme*. 2006.
30. Hayes, A.J., et al., *UKPDS Outcomes Model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82.* Diabetologia, 2013. **56**(9): p. 1925-1933.
31. Lundqvist, A., et al., *Validation of the IHE cohort model of type 2 diabetes and the impact of choice of macrovascular risk equations.* PLoS One, 2014. **9**(10): p. e110235-e110235.
32. Viriato, D., et al., *Cost-effectiveness of metformin plus vildagliptin compared with metformin plus sulphonylurea for the treatment of patients with type 2 diabetes mellitus: a Portuguese healthcare system perspective.* J Med Econ, 2014. **17**(7): p. 499-507.
33. van der Heijden, A.A.W.A., et al., *Policy evaluation in diabetes prevention and treatment using a population-based macro simulation model: the MICADO model.* Diabet. Med, 2015. **32**(12): p. 1580-1587.
34. Wolowacz, S., et al., *Development and validation of a cost-utility model for Type 1 diabetes mellitus.* Diabet. Med, 2015. **32**(8): p. 1023-1035.
35. Ye, W., et al., *The Michigan Model for Coronary Heart Disease in Type 2 Diabetes: Development and Validation.* Diabetes technology & therapeutics, 2015. **17**(10): p. 701-711.
36. Valentine, W.J.P., et al., *The Prime Diabetes Model: Novel Methods for Estimating Long-Term Clinical and Cost Outcomes in Type 1 Diabetes Mellitus.* Value Health, 2016. **20**(7): p. 985-991.
37. Willis, M., C. Asseburg, and J. He, *Validation of economic and health outcomes simulation model of type 2 diabetes mellitus (ECHO-T2DM).* J Med Econ, 2013. **16**(8): p. 1007-1021.
38. Willis, M., et al., *Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM).* PharmacoEconomics, 2017. **35**(3): p. 375-396.

39. Kwon, C.S., E. Seoane-Vazquez, and R. Rodriguez-Monguio, *Cost-effectiveness analysis of metformin+dipeptidyl peptidase-4 inhibitors compared to metformin+sulfonylureas for treatment of type 2 diabetes.* BMC Health Serv Res, 2018. **18**(1): p. 78-78.
40. Laiteerapong, N., et al., *Individualized glycemic control for U.S. Adults with type 2 diabetes a cost-effectiveness analysis.* Ann Intern Med, 2018. **168**(3): p. 170-178.
41. Nguyen, E., et al., *Cost-utility of empagliflozin in patients with type 2 diabetes at high cardiovascular risk.* J Diabetes Complications, 2018. **32**(2): p. 210-215.
42. Shao, H., et al., *Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO).* PharmacoEconomics, 2018. **36**(9): p. 1125-1134.
43. Wu, B., et al., *Development and validation of a Health Policy Model of Type 2 diabetes in Chinese setting.* J Comp Eff Res, 2018. **7**(8): p. 749-763.
44. Abramson, A., et al., *Quantifying the Value of Orally Delivered Biologic Therapies: A Cost-Effectiveness Analysis of Oral Semaglutide.* J Pharm Sci, 2019. **108**(9): p. 3138-3145.
45. Chin, K.L., et al., *Cost-effectiveness of first-line versus delayed use of combination dapagliflozin and metformin in patients with type 2 diabetes.* Sci Rep, 2019. **9**(1): p. 3256-3256.
46. Kansal, A., et al., *Cost-effectiveness analysis of empagliflozin treatment in people with Type 2 diabetes and established cardiovascular disease in the EMPA-REG OUTCOME trial.* Diabet Med, 2019. **36**(11): p. 1494-1502.
47. Kazemian, P., et al., *Development and Validation of PREDICT-DM: A New Microsimulation Model to Project and Evaluate Complications and Treatments of Type 2 Diabetes Mellitus.* Diabetes Technol Ther, 2019. **21**(6): p. 344-355.
48. Pollock, R.F., et al., *Long-term Cost-effectiveness of Insulin Degludec Versus Insulin Glargine U100 in the UK: Evidence from the Basal-bolus Subgroup of the DEVOTE Trial (DEVOTE 16).* Applied health economics and health policy, 2019. **17**(5): p. 615-627.
49. Su, Z.T., et al., *The Use of Computer Simulation Modeling to Estimate Complications in Patients with Type 2 Diabetes Mellitus: Comparative Validation of the Cornerstone Diabetes Simulation Model.* Pharmacoecon Open, 2020. **4**(1): p. 37-44.
50. Tran-Duy, A., et al., *A patient-level model to estimate lifetime health outcomes of patients with type 1 diabetes.* Diabetes Care, 2020. **43**(8): p. 1741-1749.
51. Wu, B. and L. Shi, *Cost-utility of ticagrelor plus aspirin in diabetic patients with stable coronary artery disease.* European heart journal. Cardiovascular pharmacotherapy, 2020.
52. Bagepally, B.S., et al., *Cost-utility analysis of dapagliflozin compared to sulfonylureas for type 2 diabetes as second-line treatment in indian healthcare payer's perspective.* Clinicoecon Outcomes Res, 2021. **13**: p. 897-907.
53. Bekele, M., O.F. Norheim, and A. Hailu, *Cost-Effectiveness of Saxagliptin Compared With Glibenclamide as a Second-Line Therapy Added to Metformin for Type 2 Diabetes Mellitus in Ethiopia.* MDM policy & practice, 2021. **6**(1): p. 238146832110057-23814683211005771.
54. Deerochanawong, C., et al., *Cost–Utility Analysis of Dapagliflozin as an Add-On to Standard Treatment for Patients with Type 2 Diabetes and High Risk of Cardiovascular Disease in Thailand.* Diabetes therapy, 2021. **12**(7): p. 1947-1963.
55. Tanaka, S., et al., *Developing a health economic model for Asians with type 2 diabetes based on the Japan Diabetes Complications Study and the Japanese Elderly Diabetes Intervention Trial.* BMJ Open Diabetes Research & Care, 2021. **9**(1): p. e002177.
56. Abushanab, D., et al., *First-line Treatment with Empagliflozin and Metformin Combination Versus Standard Care for Patients with Type 2 Diabetes Mellitus and Cardiovascular Disease in Qatar. A Cost-Effectiveness Analysis.* Curr Probl Cardiol, 2022. **47**(6): p. 100852-100852.

57. Huang, K., et al., *Cost–Effectiveness Analysis of Dapagliflozin Plus Standard Treatment for Patients With Type 2 Diabetes and High Risk of Cardiovascular Disease in China.* Frontiers in public health, 2022. **10**: p. 936703-936703.
58. Peng, Z.Y., et al., *Cost-effectiveness of sodium-glucose cotransporter-2 inhibitors versus dipeptidyl peptidase-4 inhibitors among patients with type 2 diabetes with and without established cardiovascular diseases: A model-based simulation analysis using 10-year real-world data and targeted literature review.* Diabetes Obes Metab, 2022. **24**(7): p. 1328-1337.
59. Steg, P.G., et al., *Cost-effectiveness of ticagrelor in patients with type 2 diabetes and coronary artery disease: a European economic evaluation of the THEMIS trial.* Eur Heart J Cardiovasc Pharmacother, 2022. **8**(8): p. 777-785.
60. Virgili, G., et al., *Utilities and QALYs in health economic evaluations: glossary and introduction.* Internal and Emergency Medicine, 2010. **5**(4): p. 349-352.
61. Tanaka, S., et al., *Developing a Health Economic Model for Asians with Type 2 Diabetes Based on the Japan Diabetes Complications Study and the Japanese Elderly Diabetes Intervention Trial.* Diabetes (New York, N.Y.), 2018. **67**(Supplement_1).
62. Holman, R.R., et al., *10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes.* N Engl J Med, 2008. **359**(15): p. 1577-1589.
63. Turner, R., *Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33).* The Lancet (British edition), 1998. **352**(9131): p. 837-853.
64. Kengne, A.P., et al., *Contemporary model for cardiovascular risk prediction in people with type 2 diabetes.* Eur J Cardiovasc Prev Rehabil, 2011. **18**(3): p. 393-398.
65. Zethelius, B., et al., *A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish National Diabetes Register (NDR).* Diabetes Res Clin Pract, 2011. **93**(2): p. 276-284.
66. Eastman, R.C., et al., *Model of complications of NIDDM. I. Model construction and assumptions.* Diabetes Care, 1997. **20**(5): p. 725-734.
67. Bagust, A., et al., *An economic model of the long-term health care burden of Type II diabetes.* Diabetologia, 2001. **44**(12): p. 2140-2155.
68. Fried, L.P., et al., *The Cardiovascular Health Study: design and rationale.* Ann Epidemiol, 1991. **1**(3): p. 263-76.
69. Basu, S., et al., *Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials.* Lancet Diabetes Endocrinol, 2017. **5**(10): p. 788-798.
70. Buse, J.B.M.D.P., *Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial: Design and Methods.* Am J Cardiol, 2007. **99**(12): p. S21-S33.
71. Bender, R., T. Augustin, and M. Blettner, *Generating survival times to simulate Cox proportional hazards models.* Statist. Med, 2005. **24**(11): p. 1713-1723.
72. Briggs, A., M. Sculpher, and K. Claxton, *Decision Modelling for Health Economic Evaluation*. 2006: Oxford University Press.
73. Davies, C., et al., *The "Hazards" of Extrapolating Survival Curves.* Med Decis Making, 2013. **33**(3): p. 369-380.
74. Hardern, C., et al., *EX2 Structural Uncertainty in Survival Extrapolation: Exploring the IMPACT of FOUR MODEL Averaging Methods and Adjusting for DATA Maturity.* Value in health, 2020. **23**: p. S402-S402.
75. Dakin, H.A., et al., *Accurately Reflecting Uncertainty When Using Patient-Level Simulation Models to Extrapolate Clinical Trial Data.* Med Decis Making, 2020. **40**(4): p. 460-473.
76. Haji Ali Afzali, H. and J. Karnon, *Exploring Structural Uncertainty in Model-Based Economic Evaluations.* PharmacoEconomics, 2015. **33**(5): p. 435-443.

77. Kearns, B., et al., *How Uncertain is the Survival Extrapolation? A Study of the Impact of Different Parametric Survival Models on Extrapolated Uncertainty About Hazard Functions, Lifetime Mean Survival and Cost Effectiveness.* PharmacoEconomics, 2020. **38**(2): p. 193-204.
78. Negrín, M.A., J. Nam, and A.H. Briggs, *Bayesian Solutions for Handling Uncertainty in Survival Extrapolation.* Med Decis Making, 2017. **37**(4): p. 367-376.

Table 2a. Uncertainty section of Philips checklist.

Author/Year	Methodological Uncertainty	Structural Uncertainty	Heterogeneity	Parameter Uncertainty		
	Have the methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	Has heterogeneity been dealt with by running the model separately for different subgroups?	Are the methods of assessment of parameter uncertainty appropriate?	Has probabilistic sensitivity analysis been done, if not has this been justified?	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
Hayes et al. 2013 (UKPDS-OM2) [30]	No	No	No	Yes	Yes*	No
Lundqvist et al. 2014 (IHE) [31]	No	No	No	Yes	Yes	NA
Viriato et al. 2014 [32]	Yes	No	No	Yes	Yes	No
Van der Heijden et al. 2015 (MICADO) [33]	No	No	No	Yes	Yes	Yes
Wolowacz et al. 2015 [34]	No	No	No	Yes	Yes	No
Ye et al. 2015 (Michigan) [35]	No	No	No	Yes	Yes	NA
Valentine et al. 2016 (PRIME) [36]	No	Partial	No	Yes	Yes	NA
Willis et al. 2013 & 2017 (ECHO-T2DM) [37, 38]	No	No	No	Yes	Yes	NA
Kwon et al. 2018 [39]	Yes	Yes	No	Yes	Yes	Yes
Laiteerapong et al. 2018 [40]	Yes	No	No	No	No	No

Author/Year	Methodological Uncertainty	Structural Uncertainty	Heterogeneity	Parameter Uncertainty		
	Have the methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	Has heterogeneity been dealt with by running the model separately for different subgroups?	Are the methods of assessment of parameter uncertainty appropriate?	Has probabilistic sensitivity analysis been done, if not has this been justified?	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
Nguyen et al. 2018 [41]	Yes	Yes	No	Yes	Yes	Yes
Shao et al. 2018 (BRAVO) [42]	No	No	No	Yes	Yes*	NA
Wu et al. 2018 (COMIT) [43]	No	No	No	No	No	NA
Abramson et al. 2019 [44]	No	No	No	Yes	Yes	No
Chin et al. 2019 [45]	No	No	No	Yes	Yes	Yes
Kansal et al. 2019 [46]	Yes	Yes	No	Yes	Yes	Yes
Kazemian et al. 2019 (PREDICT-DM) [47]	No	No	No	Yes	Yes	NA
Pollock et al. 2019 [48]	Yes	Yes	No	Yes	Yes	No
Su et al. 2020 (Cornerstone) [49]	No	No	No	No	No	NA
Tran-Duy et al. 2020 [50]	No	No	No	Yes	Yes*	Yes
Wu et al. 2020 [51]	No	No	Yes	Yes	Yes	Yes
Bagepally et al. 2021 [52]	No	No	No	Yes	Yes	Yes
Bekele et al. 2021 [53]	Yes	No	No	Yes	Yes	Yes
Deerochanawong et al. 2021 [54]	No	No	No	Yes	Yes	Yes
Tanaka et al. 2021 (JJCEM) [55]	No	No	No	Yes	Yes	NA

Author/Year	Methodological Uncertainty	Structural Uncertainty	Heterogeneity	Parameter Uncertainty		
	Have the methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	Has heterogeneity been dealt with by running the model separately for different subgroups?	Are the methods of assessment of parameter uncertainty appropriate?	Has probabilistic sensitivity analysis been done, if not has this been justified?	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
Abushanab et al. 2022 [56]	Yes	No	Yes	Yes	Yes	Yes
Huang et al. 2022 [57]	No	No	No	Yes	Yes	Yes
Peng et al. 2022 [58]	Yes	Yes	No	Yes	Yes	Yes
Steg et al. 2022 [59]	No	No	Yes	Yes	Yes	No

*Bootstrapping approach

Table 2b. Method of parameter uncertainty analysis used in each model.

Author/Year	Deterministic Sensitivity Analysis	Probabilistic Sensitivity Analysis	Non-parametric Bootstrapping
Hayes et al. 2013 (UKPDS-OM2) [30]	Yes	No	Yes
Lundqvist et al. 2014 (IHE) [31]	No	Yes	No
Viriato et al. 2014 [32]	Yes	Yes	No
Van der Heijden et al. 2015 (MICADO) [33]	Yes	Yes	No
Wolowacz et al. 2015 [34]	Yes	Yes	No
Ye et al. 2015 (Michigan) [35]	No	Yes	No
Valentine et al. 2016 (PRIME) [36]	No	Yes	Yes
Willis et al. 2013 & 2017 (ECHO-T2DM) [37, 38]	No	Yes	No
Kwon et al. 2018 [39]	Yes	Yes	No
Laiteerapong et al. 2018 [40]	Yes	No	No
Nguyen et al. 2018 [41]	Yes	Yes	No
Shao et al. 2018 (BRAVO) [42]	No	No	Yes
Wu et al. 2018 (COMIT) [43]	NA	NA	NA
Abramson et al. 2019 [44]	Yes	Yes	No
Chin et al. 2019 [45]	Yes	Yes	No
Kansal et al. 2019 [46]	No	Yes	No
Kazemian et al. 2019 (PREDICT-DM) [47]	No	Yes	No
Pollock et al. 2019 [48]	No	Yes	Yes
Su et al. 2020 (Cornerstone) [49]	NA	NA	NA
Tran-Duy et al. 2020 [50]	Yes	No	Yes
Wu et al. 2020 [51]	Yes	Yes	No
Bagepally et al. 2021 [52]	Yes	Yes	No
Bekele et al. 2021 [53]	Yes	Yes	No
Deerochanawong et al. 2021 [54]	Yes	Yes	No
Tanaka et al. 2021 (JJCEM) [55]	No	Yes	No

Author/Year	Deterministic Sensitivity Analysis	Probabilistic Sensitivity Analysis	Non-parametric Bootstrapping
Abushanab et al. 2022 [56]	Yes	Yes	No
Huang et al. 2022 [57]	Yes	Yes	No
Peng et al. 2022 [58]	Yes	Yes	No
Steg et al. 2022 [59]	Yes	Yes	No

Table 3. Type of risk equations used in the patient-level models.

Method	Number of models
T2DM	12
UKPDS-OM1 / UKPDS-OM2 risk equations	5
UKPDS-OM1 / UKPDS-OM2 risk equations & other sources	3
RECODe risk equations	2
BRAVO risk equations	1
Other (study-specific) risk equations	1
T1DM	3
Risk equations / Risk functions	2
Sources from the literature	1

BRAVO indicates Building, Relating, Assessing, and Validating Outcomes; RECODe, Risk Equations for Complications Of type 2 Diabetes; UKPDS-OM1, United Kingdom Prospective Diabetes Study – Outcomes Model 1; UKPDS-OM2, United Kingdom Prospective Diabetes Study – Outcomes Model 2.

Table 4. Internal and External Validation results.

Author/Year	Internal Validation	External Validation
Hayes et al. 2013 (UKPDS-OM2) [30]	Predicted curves were within the 95% CIs of the actual cumulative failure curves for all events and death	NA
Lundqvist et al. 2014 (IHE) [31]	NDR risk equations Intercept = 0.849 Slope = 0.918 Coef. Determination = 0.971 UKPD-OM1 risk equations Intercept = -0.644 Slope = 0.944 Coef. Determination = 0.980 UKPDS-OM2 risk equations Intercept = 0.0828 Slope = 0.896 Coef. Determination = 0.967	NDR risk equations Intercept = 0.286 Slope = 0.985 Coef. Determination = 0.960 UKPDS-OM1 risk equations Intercept = -0.754 Slope = 1.049 Coef. Determination = 0.963 UKPDS-OM2 risk equations Intercept = -0.300 Slope = 0.899 Coef. Determination = 0.968
Viriato et al. 2014 [32]	NA	NA
Van der Heijden et al. 2015 (MICADO) [33]	NA	Incidence of amputation estimated by MICADO = 592 compared to the observed incidence 728. Incidence of ESRD estimated by MICADO = 247 compared to the observed incidence 277. For macrovascular outcomes the MICADO model appeared visually to overestimate the incidence in the Kaiser Permanente dataset and match the Swedish dataset.
Wolowacz et al. 2015 [34]	Model predictions were within 2% of expected values	Model predictions were within 8% of expected values
Ye et al. 2015 (Michigan) [35]	Slope = 0.98 Coef. Determination = 0.99	Slope = 0.84 Coef. Determination = 0.81
Valentine et al. 2016 (PRIME) [36]	RMSD < 3%	Cardiovascular events validation Four RMSD values < 1% and two RMSD values < 3%. All six standardized AUC < 1%.

Author/Year	Internal Validation	External Validation
		Microalbuminuria validation RMSD = 8.5% and AUC = 0.1% Overt nephropathy validation RMSD = 11.5% and AUC = 0.6% Retinopathy events validation RMSD = 2.8% and AUC = 0.7% Lower extremity amputation validation RMSD (males/females) = 0.8% / 0.6% AUC = 0.03%
Willis et al. 2013 (ECHO-T2DM) [37]	Intercept = -0.002 Slope = 1.024 Coef. determination = 0.95	Intercept = 0.015 Slope = 1.072 Coefficient of determination = 0.97
Willis et al. 2017 (ECHO-T2DM) [38]	UKPDS 82 Intercept = 0.023 Slope = 0.974 Coef. Determination = 0.86 MAE = 0.051 RMSE = 0.073 MSLAR = 0.266 MSLE = 0.354 UKPDS 68 Intercept = 0.006 Slope = 0.957 Coef. Determination = 0.86 MAE = 0.046 RMSE = 0.071 MSLAR = 0.242 MSLE = 0.322 ADVANCE Intercept =-0.001 Slope = 1.007 Coef. Determination = 0.89 MAE = 0.039 RMSE = 0.064 MSLAR = 0.238 MSLE = 0.307	UKPDS 82 Intercept = 0.028 Slope = 1.029 Coef. Determination = 0.91 MAE = 0.043 RMSE = 0.062 MSLAR = 0.346 MSLE = 0.421 UKPDS 68 Intercept = 0.001 Slope = 1.038 Coef. Determination = 0.90 MAE = 0.038 RMSE = 0.056 MSLAR = 0.342 MSLE = 0.425 ADVANCE Intercept = 0.020 Slope = 0.983 Coef. Determination = 0.90 MAE = 0.054 RMSE = 0.071 MSLAR = 0.505 MSLE = 0.602

Author/Year	Internal Validation	External Validation
	NDR Intercept = 0.002 Slope = 1.002 Coef. Determination = 0.89 MAE = 0.034 RMSE = 0.056 MSLAR = 0.274 MSLE = 0.328	NDR Intercept = 0.024 Slope = 1.028 Coef. Determination = 0.91 MAE = 0.048 RMSE = 0.070 MSLAR = 0.420 MSLE = 0.513
Kwon et al. 2018 [39]	NA	NA
Laiteerapong et al. 2018 [40]	Predicted values in all cases were generally within the 95% CIs of the observed cumulative incidence	NA
Nguyen et al. 2018 [41]	NA	NA
Shao et al. 2018 (BRAVO) [42]	All the predicted incidence curves fit close to the observed curves and all predicted curves were within the 95% CI.	Intercept = 0.001 Slope = 1.071 Coef. Determination = 0.86
Wu et al. 2018 (COMIT) [43]	Overall coefficient of determination = 0.8701 Slope = 0.9631 MAPE = 32.93%	
Abramson et al. 2019 [44]	Verified that the adherence, HbA1c levels, and disutility of patient cohorts passing through the model matched the data used to estimate the model parameters.	NA
Chin et al. 2019 [45]	NA	NA
Kansal et al. 2019 [46]	Reproduced the overall event rates in EMPA-REG OUTCOME when treated as competing events	NA
Kazemian et al. 2019 (PREDICT-DM) [47]	MAPE (intensive/standard)= 19% / 25% MEDAPE (intensive/standard) = 20% / 16% RMSPE (intensive/standard) = 23% / 35% 95% limit of agreement = 0.020 Mean difference = 0.0025 ICC (intensive/standard) = 0.94 / 0.90	VADT MAPE (intensive/standard) = 29% / 20% MEDAPE (intensive/standard) = 29% / 21% RMSPE (intensive/standard) = 32% / 24% 95% limit of agreement = 0.032 Mean difference = -0.0067 ICC (intensive/standard) = 0.89 / 0.96 Look AHEAD MAPE (lifestyle/support&education) = 42% / 10% MEDAPE (lifestyle/support&education) = 21% / 5%

Author/Year	Internal Validation	External Validation
		RMSPE (lifestyle/support&education) = 70% / 14% 95% limit of agreement = 0.011 Mean difference = -0.0033 ICC (lifestyle/support&education) = 0.95 / 0.99
Pollock et al. 2019 [48]	MAPE = 0.159%	NA
Su et al. 2020 (Cornerstone) [49]	Predicted incidence curves closely matched the reported curves for all first-time events including mortality in the source publication	Coefficient of determination = 0.637
Tran-Duy et al. 2020 [50]	The predicted values in all cases were generally within the 95% CIs of the observed cumulative incidence	NA
Wu et al. 2020 [51]	NA	NA
Bagepally et al. 2021 [52]	NA	NA
Bekele et al. 2021 [53]	NA	NA
Deerochanawong et al. 2021 [54]	NA	NA
Tanaka et al. 2021 (JJCEM) [55]	NA	NA
Abushanab et al. 2022 [56]	NA	NA
Huang et al. 2022 [57]	NA	NA
Peng et al. 2022 [58]	NA	NA
Steg et al. 2022 [59]	NA	NA

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) diagram

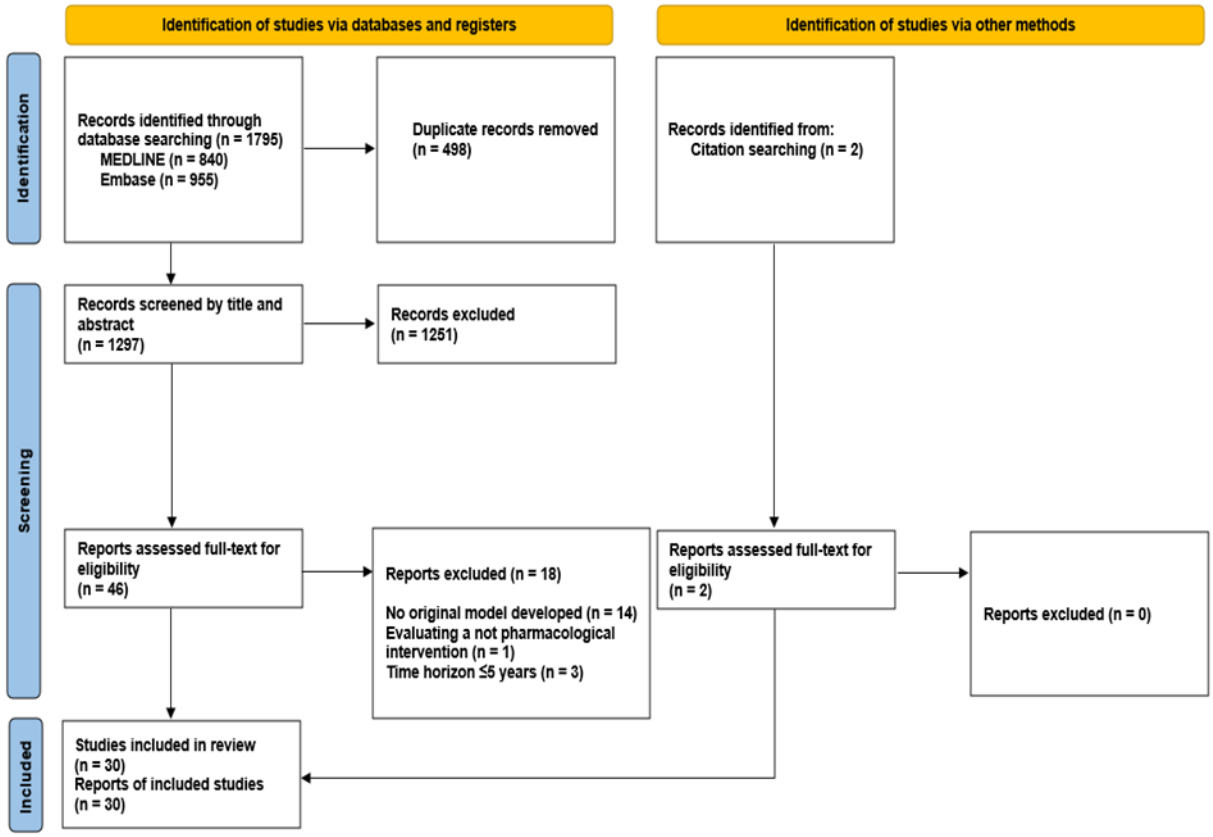

Figure 2. Illustrates the percentage of models that were classified as low, medium or high quality in each section of the Philips quality assessment checklist as well as based on the overall score throughout all the sections.

Figure 3. Spider chart that depicts the quality assessment score of each of the 20 models separately in each of the four dimensions (overall, structure, data, consistency).