Aus dem Institut für Versorgungsforschung und Gesundheitsökonomie
der Heinrich-Heine-Universität Düsseldorf
Direktorin: Univ.- Prof. Dr. Dr. Andrea Icks, MBA

Systematic literature review of external validation in type 2 diabetes computer simulation models: definitions, approaches, implications and room for improvement

Dissertation

zur Erlangung des Grades eines Doktors der Medizin der Medizinischen Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Marko Frings
2023

Als Inauguraldissertation gedruckt mit Genehmigung der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachterin: Prof. Dr. Dr. Andrea Icks, MBA

Zweitgutachter: Prof. Dr. Rolf Holle

Widmung

Für meine Frau Julia

Auflistung eigener Publikationen

Es wurden keine Teile der Arbeit vorab veröffentlicht.

Zusammenfassung

Computersimulationsmodelle des Typ-2-Diabetes mellitus (T2DM) werden in der gesundheitsökonomischen Evaluation zur Beurteilung der Kosteneffektivität von therapeutischen und präventiven Maßnahmen eingesetzt. Damit diese Modelle von Entscheidungsträgern allgemein eingesetzt werden können, muss die Validität dieser Modelle überprüft werden. Externe Validierung kann zur Beurteilung einer generellen Anwendbarkeit verwendet werden, indem ein Modell die beobachteten Ergebnisse einer neuen und unabhängigen (klinischen) Studie simulieren sollte. Es gibt mehrere Leitlinien und Empfehlungen zur Durchführung von externer Validierung, jedoch gibt es keine systematische Übersichtsarbeit über die derzeitige Praxis. Das Ziel dieser Arbeit war, eine systematische Literaturübersicht durchzuführen, um die externen Validierungsansätze bei Simulationsmodellen von T2DM zu beschreiben und zu bewerten.

Der systematische Literaturübersicht (SLR) wurde in Anlehnung an die Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) durchgeführt. Dazu wurden im September 2017 und im Mai 2020 15 Datenbanken durchsucht und es wurde ein Titel- und Abstract- sowie ein Volltextscreening durchgeführt, um Modelle nach zuvor festgelegten Kriterien auszuwählen. Nur Modelle mit realisierter externer Validierung wurden einer Datenextraktion und -synthese unterzogen. Die Analyse wird aufgrund der Vielfalt der Modelle und der externen Validierungsmethoden in einer narrativen und deskriptiven Weise präsentiert.

Die Recherche lieferte insgesamt 21.737 Treffer. Nach dem Titel- und Abstractscreening wurden 79 T2DM-Modelle identifiziert und 24 von ihnen ausgewählt, für die eine externe Validierung durchgeführt oder erwähnt wurde (in 43 Veröffentlichungen). Weiterhin wurden 6 Modelle in die Analyse aufgenommen, für die eine externe Validierung nur auf Mt Hood Challenges, Tagungen des Fachbereichs, durchgeführt wurde. Somit wurden insgesamt 30 Modelle mit durchgeführter externer Validierung ausgewählt. Im Allgemeinen waren die Detailtiefe und die Qualität der externen Validierung unter den Studien sehr unterschiedlich. 18 Modelle enthielten keine Erklärung der Herkunft ihrer Datenquellen. Für 9 Modelle (30 % von 30 Modellen) wurde eine Regressionsanalyse der simulierten gegenüber den beobachteten Ergebnissen durchgeführt und für 12 Modelle (40 % von 30 Modellen) wurde die Vorhersagegenauigkeit berechnet. Dieser SLR gibt einen Überblick über die derzeitige Praxis der externen Validierung in der Literatur: Die Mehrheit der Modellierer führt keine externe Validierung durch oder es werden keine ausreichenden Ergebnisse hinsichtlich der Leitlinien angeben. Gleichzeitig finden sich in der Literatur "State of the Art"-Beispiele für externe Validierung. Für die Zukunft wäre es wünschenswert, dass Autoren klar angeben, ob sie eine externe Validierung durchführen und den Prozess so genau wie möglich beschreiben.

Summary

Computer simulation models of type 2 diabetes mellitus (T2DM) are used in health care decision-making to assess the cost-effectiveness of therapeutic and preventive interventions. To be generally applicable for decision makers, the models need to prove their validity. External validation can be used to assess the general applicability by testing whether a model can simulate the results observed in new independent studies or clinical trials. There are several guidelines and recommendations for conducting external validation. However, no systematic review of current practice exists. Thus, I aimed to conduct a systematic literature review to describe and appraise the external validation approaches employed in computer simulation models of T2DM.

The systematic literature review (SLR) was conducted referring to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The literature search was conducted in 15 databases in September 2017 and in May 2020. A title and abstract and full text screening were conducted to select models that met predefined criteria. Only those models in which external validation was preformed were subject to data extraction and synthesis. The analysis is presented in a narrative and descriptive way due to diversity of the models and the external validation methods.

A total of 21,737 records were found in literature. After the title and abstract screening, I identify 79 T2DM models and selected 24 of them for which the external validation was performed or mentioned (in 43 papers). In addition, six models were added to the analysis for which external validation was only performed at Mt Hood Challenges, meetings to discuss current modeling topics in diabetes. Thus, in total, I selected 30 models in which external validation was applied. In general, the level of detail and quality of the external validation varied considerably between studies. 18 models did not provide an explanation for the selection of their data sources. Regression analysis on predicted vs. observed outcomes was performed for 9 models (30 % of 30 models) and precision of prediction was calculated in 12 models (40 % of 30 models). This SLR gives an overview of the current practices of external validation in the literature: the majority of modelers did not assess the external validity of their diabetes simulation models at all, or did not provide sufficient details required in the guidelines abovementioned. At the same time, “state of art” examples of external validation performed in accordance with the guidelines can be found in the literature. For the future, it would be desirable for authors of models to clearly state whether they perform external validation and then describe the process as precisely as possible.

Abkürzungsverzeichnis

CBA	Cost-benefit analysis
CDSR	Cochrane Database of Systematic Reviews
CHD	Coronary heart disease
CHF	Congestive heart failure
CORE	Center for Outcomes Research
CVD	Cardiovascular disease
DARE	Database of Abstracts of Reviews of Effects
DALYs	Disability-adjusted life years
DCCT	Diabetes Control and Complications Trial
DM	Diabetes Mellitus
DPP	Diabetes Prevention Program
DSME/S	Diabetes self-management education and support
EQ-5D	EuroQol five-dimensional
EV	External validation
FPG	Fasting plasma glucose
GDM	Gestational diabetes mellitus
GFRIFG	Glomerular filtration rateImpaired fasting glucose
IGT	Impaired glucose tolerance
ISPOR	The Professional Society for Health Economics and Outcomes Research
MI	Myocardial infarction
MtHCh	Mount Hood Challenge
NPH	Neutral Protamine Hagedorn
OGTT	Oral glucose tolerance test
OpenDOAR	Directory of Open Access Repository

PORSPERO	International Prospective Register of Systematic Review
PQDT	ProQuest Dissertations & Theses Database
QALY	Quality-adjusted life year
RCT	Randomized controlled trial
SLR	Systematic literature review
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
USD	United States Dollar

Contents

1 Introduction

Diabetes Mellitus (DM) is one of the largest health challenges of the 21st century reaching prevalence of 9.3% worldwide among adults 20-79 years old, and an important challenge for global health care systems and societies. Global health expenditures due to diabetes were estimated to be around 760 billion USD in 2019. Prevalence and associated costs were expected rising in the following years [1]. Thus, on the one hand, large amount of care and financial resources need to be provided for health services; on the other hand, the disease produces socioeconomic losses [2, 3]. Complications associated with diabetes can be serious like cardiac infarction, stroke, kidney failure and blindness [1]. Under these circumstances, it is very important for health care providers and decision makers to deal with DM and its complications efficiently and effectively.

Decision makers and health care providers need to assess cost-effectiveness of various treatments and interventions in diabetes care and prevention and to make long-term predictions and to allocate sparse resources effectively. These tasks can be performed only by operating with up-to-date scientific evidence [4, 5]. Computer simulation models are useful tools for this issue. In the past, it was always explicitly mentioned that simulations were carried out using computers. Today, the term simulation model implies this. Various models simulate the diabetes mellitus progression and its complications, and are used in health economic evaluations. They can be used to evaluate medical and public health interventions in a lifelong perspective by predicting the outcomes of diabetes treatment and consequent health care costs [4].

Nevertheless, such models should prove their validity to fulfil their purpose, for example, being tested with external or new data sources, which were not used to set up an original model (named “external validation”) [4]. Mount Hood Challenge (MtHCh) is a meeting organized since 1999 where simulation modelers of diabetes models meet and discuss actual topics in this field regarding model structure and outcomes [6]. The 4th, 5th, and 9th MtHChs dealt with external validation exercises and reports were published [7–9]. Also, there are several guidelines and recommendations such as the guideline of the American Diabetes Association (ADA), recommendations from the Professional Society for Health Economics and Outcomes Research (ISPOR) Task Force and reports and materials of the Mt Hood 4 Modeling Group of how the external validation should be performed [10–13]. However, there is no systematic overview of current practices in external validation approaches in diabetes modelling and whether these attempts are made in accordance with the guidelines.

The main objective of this thesis is to conduct a SLR to describe and appraise external validation approaches employed in the simulation models of diabetes.

1.1 Diabetes

1.1.1 Definition and pathophysiology

Diabetes is a group of chronic and debilitating conditions that are defined by a hyperglycaemia, elevated blood glucose. This umbrella term for several diseases varying by pathophysiology and causes can be classified into five groups: 1. Type 1 diabetes mellitus (T1DM), 2. Type 2 diabetes mellitus (T2DM), 3. Prediabetes, 4. Gestational diabetes mellitus (GDM), 5. Other types of diabetes like neonatal diabetes mellitus, maturity onset diabetes of the young etc. [1].

The above-mentioned hyperglycaemia is caused by an insulin resistance, or an underproduction and/or a loss of production of insulin. Insulin is hormone produced by the beta cells of the pancreas and is released in the bloodstream as a response to blood glucose elevating. Through its binding to receptor on somatic cells it effects the uptake of glucose from the bloodstream to the cells. Further, insulin has a relevant impact on the fat and protein metabolism. The most common forms are T1DM and T2DM. T1DM is characterized by an autoimmune reaction which affects the beta cells of the pancreas and leads to a destruction of these cells. Consequently, insulin is produced less or not at all and somatic cells cannot uptake glucose. In contrast, an insulin resistance of somatic cells is the main characteristic of T2DM. More insulin is produced through the beta cells to compensate this condition. In the long term, beta cells cannot provide this high supply and they lose their function to produce insulin. The outcomes are the same as in T1DM [1].

Prediabetes, or intermediate hyperglycaemia, is a state characterized by higher risk of diabetes and its earlier microvascular and macrovascular complications, and defined by glycaemia higher than normal, but lower than diabetes threshold. The term covers several conditions: impaired fasting glucose (IFG), that can be diagnosed by fasting plasma glucose (FPG), and impaired glucose tolerance (IGT), diagnosed by oral glucose tolerance test (OGTT). Recently, the evaluated HbA1c level starts being classified as a definition of prediabetes as well [14, 15].

Further test criteria a presented in the following section.

Diagnostic criteria

For diagnosis of hyperglycemia ADA and WHO criteria are most relevant. In general, three groups can be distinguished: 1. Diabetes, 2. IGT, 3. IFG. Blood test for measuring the

hyperglycaemia are FPG, two-hour plasma glucose in OGTT. HbA1c and random plasma glucose can also be used but random plasma glucose is only diagnostic criteria for diabetes. ADA and WHO criteria differ in the diagnosis for prediabetes. The following tables illustrates the criteria for diagnosis of hyperglycaemia [1, 16]:

		Types of hyperglycaemia				
		Diabetes	IGT		IFG	
			WHO	ADA	WHO	ADA
Tests	Fasting glucose*	≥7.0 mmol/l 126 mg/dl	<7.0 mmol/l 126 mg/dl		6.1-6.9 mmol/l 110-125 mg/dl	5.6-6.9 mmol/l 100-125 mg/dl
	2-h plasma glucose**	≥11.1 mmol/l 200 mg/dl	7.8-11.1 mmol/l 140-200 mg/dl	7.8-11.0 mmol/l 140-199 mg/dl	<7.8 mmol/l 140 mg/dl	
	Hb1Ac	≥48 mmol/mol ≥6,5%	ADA: 39-47 mmol/mol, 5.7-6.4%			
	Random plasma glucose***	≥11.1 mmol/l 200 mg/dl				

Table 1: Diagnostic criteria of hyperglycaemia, adapted from [1, 16, 17]

* No caloric intake for 8 hours.

**75 mg oral glucose load which should be dissolved in water. The glucose level should be measured from venous blood 2-h postprandial. The test is also called OGGT.

***Can be used when symptoms of a hyperglycaemia are present.

For a diagnosis of diabetes one of the four above-mentioned tests is required. According to the WHO criteria, for an IGT diagnosis a fasting plasma glucose and an OGTT are necessary. An IFG can be diagnosed with a fasting plasma glucose and if measured, with an OGTT. According to ADA for an IGT only an OGTT and for an IFG a FPG is necessary. A prediabetes can be diagnosed with a Hb1Ac according to ADA [1, 16]. In the following, information about T2DM will be presented if available. Otherwise, diabetes will be dealt in general.

1.1.2 Epidemiology

Following hypertension and tobacco use, hyperglycaemia is the third leading risk factor for mortality [18]. In 2019, around 463 million people 20-79 years old were suffering from diabetes, diagnosed and undiagnosed, and 4.2 million died due to this disease. These figures include both type 1 and 2 diabetes. The global prevalence of diabetes among adults 20-79 years old is

estimated to be around 9.3%. Furthermore, the projected total number of persons with diabetes will increase to 578 million in 2030 and 700 million in 2045. The related prevalence is estimated to be 10.2% by 2030 and 10.9% by 2045 [1]. In contrast to the increasing prevalence, a recent review has shown that in most of the studies from developed countries between 2006-2014 diabetes incidence is stable or decreases in these years. This can be a first sign of successful diabetes prevention programs. Whereas in the developing world and in low income countries, the diabetes incidence is increasing between 2006-2014 [19]. The increase in the prevalence indicates that diabetes is one of the biggest growing health problems today. Predominantly, diabetes is a problem of higher age, urban regions and middle- and low-income countries [1].

Around 90% of all cases of diabetes are classified as T2DM [1]. Overweight, sedentary life-styles, diet, aging, family history, environmental factors, urbanization, low income and tobacco, and alcohol use are relevant risk factors for developing T2DM [20].

1.1.3 Complications and Comorbidities of Diabetes

Diabetes is a life-long chronic condition that is developing over years, accompanied by comorbidities, and often leads to devastating and costly late complications.

Usually, due to diabetes, patients suffer from a lot of complications. They can be classified in diabetes-specific microvascular complications (retinopathy, nephropathy and neuropathy) and macrovascular complications (coronary heart disease (CHD), periphery artery disease, cerebrovascular disease (CVD)) [21–23]. The main function of macrovessels is to provide organs with blood and in particular oxygen. Macrovascular complications are classified as damage through atherosclerosis which is induced through inflammatory process like oxidative stress caused by hyperglyceamia. The atherosclerosis of large arteries leads to a reduced inner diameter of the arteries. This is the cause for hypoxia and inflammation of subsequent tissue [24, 25]. Microvessels comprises arterioles, capillaries and venules which main function is the regulation of blood pressure and a nutrient supply. Furthermore, the microcirculation between microvessels and the surrounding tissue is responsible for vasomotion, permeability, and myogenic response. Microvascular complications are characterized by a hyperglycaemia which leads to a thickening of capillary basement membrane of microvessels including glomeruli, retina, myocardium, skin, and muscles resulting in elevated blood pressure, delayed wound healing, and tissue hypoxia [24, 26].

Macrovascular complications comprise coronary heart disease (CHD), peripheral artery disease and cerebrovascular disease that can be combined together as cardiovascular disease (CVD) [21, 22]. The prevalence for CVD in people with T2DM is estimated to be 32.2%. Moreover, the CVD and CVD acute events are the main cause of death among people with T2DM: around 10% of diabetes patients die due to CVD [27].

CHD includes angina pectoris, myocardial infarction (MI) and congestive heart failure (CHF) [28, 29] with worldwide prevalence of 14.6% for angina, 10% for myocardial infarction and 14.9% for CHF in people with diabetes [27]. Peripheral artery disease is defined as circulatory dysfunction of blood vessels supplying legs and arms which can lead ulceration and neuropathy. Revascularization and extremity amputations are late complications [1, 21]. Cerebrovascular disease is in this topic mainly associated with stroke which occurs worldwide in 7.6% of all diabetes [27, 30]. The relative risk for an CVD events is doubled for patients with diabetes compared to those without [31]. Prediabetes, including IFG and IGT, is as well a risk factor for a CVD event and is associated with a higher risk for mortality [32–34].

Microvascular complications include retinopathy, nephropathy and neuropathy [25]. 35% of all diabetes patients suffer from a diabetic retinopathy [35]. In a systematic review including eight studies (five from Asia, one from North America, one from Caribbean, and one from sub-Saharan) the annual incidence of diabetic retinopathy ranged between 2.2 % to 12.7 % [36]. It can cause blindness and in the 20-74 year age group it is the leading cause for blindness in the U.S. and thus socioeconomic problem [37, 38].

Chronic kidney disease (CKD) is the manifestation of a diabetic nephropathy. In general, CKD can be induced by abnormal hyperglycaemia, hypertension, polyneuropathic bladder with urine retention, recurrent infections of the urinary tract and macrovascular angiopathy [1]. End-stage renal disease is the most severe manifestation of CKD. Approximately 50% of patients with T2DM have a chronic kidney disease. Incidence rates of ESRD for patients with diabetes vary from 5.4 to 804.0 per 100,000 person-years and relative risk is between 6.2 to 62.0 times higher compared to patients without diabetes. The best way to prevent a CKD and an end-stage renal disease is a screening for a proteinuria [1, 39–41].

Diabetic neuropathy is the third microvascular comorbidity of diabetes. It consists of sensory, focal or multifocal and autonomic neuropathies. The most common form is a peripheral neuropathy which exhibits as chronic sensorimotor distal symmetric polyneuropathy [1, 25]. Peripheral diabetic neuropathy is often associated with peripheral vascular disease. Both can cause ulcers and necroses which finally lead to lower extremity amputation [1]. The range of

incidences for lower extremity amputations is between 78 to 704 per 100,000 person-years for patients with T2DM and the relative is between 7.4 to 41.3 times higher compared to patients without diabetes [42].

In 2019, every eight seconds a patient with diabetes died. It is estimated that 11.3% of all deaths globally relate to diabetes[1]. A study from the UK shows that mortality is almost doubled in people with T2DM compared to those without (hazard ratio (HR) 1.93 ,95% CI 1.89-1.97) [43]. The main cause of death in people with diabetes are CVD with a proportion of one third to one-half of all deaths. Around 46 % of these deaths are under 60 years old whereby relevant welfare losses can be explained [1].

1.1.4 Therapy principles of T2DM

The ADA recommends for the treatment of prediabetes and T2DM a combination of lifestyle intervention and pharmacological treatment [44, 45]. Lifestyle intervention includes medical nutrition therapy, physical activity, diabetes self-management education and support (DSME/S), smoking cessation counseling, and psychosocial care [46]. DSME refers to the imparting of knowledge, skills, and abilities to succeed with diabetes self-care. Needs, goals, and life experience are combined with evidence-based knowledge in this process. DSMS contains activities that help patients in implementing and sustaining the behaviors to succeed in a long-term perspective with the management of their disease. Medical nutrition therapy involves eating patterns, macronutrient distribution, meal planning, and weight management [47]. Pharmacological therapy is graded in different steps whereby intensity/ aggressiveness of therapy can be increased. Used therapy steps should be evaluated after 3 to 6 months by HbA1c testing, and in case of persistent hyperglycaemia an additional therapy should be initiated otherwise the therapy is adequate [44].

Firstline of glycaemia control in patients with T2DM should a pharmacological therapy with metformin combined with a lifestyle intervention including at least physical activity and weight management. When hyperglycemia persist, second line therapy consists of first line therapy and an addition of one of the 6 classes of drugs: sulfonylurea, thiazolidinedione, dipeptidyl peptidase 4 (DDP-4) inhibitors, GLP-1 receptor agonists, SGLT2 inhibitors or insulin. The choice of a drug depends on comorbidities and risk factors of a patients. For example, for patients with CVD, CKD or HF GLP-1 receptor agonists or SGLT2 inhibitor should be included in combination therapy because the cardiovascular risk is reduced with these agents. If hyperglycaemia is severe, insulin should be part of a combination therapy because it lowers the blood glucose level effectively. Last step is a combination of insulin with one of the above-mentioned agents

or combination of short- and long-acting insulin analogues. Sole insulin therapy is required if nearly all beta cells are destroyed [44].

In general, patients with prediabetes should be offered a lifestyle intervention which includes at least a body weight reduction and physical activity. A pharmacological therapy with metformin can be added in patients with body mass index (BMI) > 35 kg/m², with an age < 60 years, and women with prior gestational diabetes mellitus [45].

In the following paragraph some examples for the efficacy of the above-mentioned interventions are given. Systematic reviews of intervention program for lifestyle modifications have shown a relative risk reduction of the cumulative incidence for T2DM of 29%. Only small losses of 1,5 kg body weight are necessary to reduce the risk for diabetes [48]. A pharmacological example is a treatment with empagliflozin 5 or 10 mg daily versus placebo in patients with T2DM. It is beneficial for preventing composite endpoint of death from cardiovascular causes, nonfatal MI or nonfatal stroke (HR 0.86; p= 0.04) [49]. Economic models simulate disease progression, complications along with prevention and treatment options and effects and might be helpful in decision making by supporting the most cost- and clinical-effective therapies.

1.1.5 Economic aspects of diabetes

All forms of diabetes are an important issue for global health care and societies. On the one hand, large amount of care and financial resources need to be provided for prevention, treatment and supporting medical services to manage diabetes epidemic. On the other hand, the disease produces socioeconomic losses as a result of productivity loss, earlier retirement, disability, and earlier death [2, 3].

The financial burden associated with diabetes care can be categorized as direct and indirect costs. Direct costs are expenditures for treating diabetes and its complications including health care expenses, medication costs, and out-of-pocket payments. Hospitalisation due to diabetes complications and comorbidities is the main influencing factor for direct costs. Indirect costs are linked to productivity losses and welfare payments. Productivity losses are subgrouped as losses due to presenteeism, absenteeism, reduced productivity of people with diabetes and early retirement due to disability or earlier death [1, 3, 50]. The global health expenditures (direct costs) in the 20-79 year age group were 232 billion USD in 2007 and increased to 760 billion USD in 2019 which is an increase of approximate 210%. A further increase is projected to 825 billion USD (8.6%) by 2030 and 845 billion USD (11.2%) by 2045 compared with 2019. In

developed countries like Switzerland, United States of America and Austria the mean health expenditures per person in 2019 varied from 5,259 to 11,916 USD. Indirect costs are difficult to estimate due to their interaction with economy and welfare payments, lacking of data, and methodological issues [1]. Their estimated share of total global costs are 34.7% with an absolute amount of 455 billion USD [51].

For the U.S. total costs for health care expenditures were 3.5 trillion USD in 2017 [52]. Direct costs for diabetes were 237 billion USD (6.8%) and indirect costs 90 billion USD (2.6%) in 2017 [2]. On patient basis, average health care expenditures account for 9,600 USD per person per year for diabetes treatment in the USA. Indirect costs contain 3.3 billion USD for absenteeism, 26.9 billion USD for reduced productivity of working population and 2.3 billion USD for not working population, 37.5 billion USD for inability to work and 19.9 billion USD for lost productivity due to premature death [2].

In total, 85,200 USD in the USA are spent as medical costs for a patient with T2DM during his or her life. The treatment of diabetic complications has a share of 53% of these costs where 57% of the complication costs are used to treat macrovascular complications [53]. Kähm et al. estimated costs of diabetes complications for different age groups with an regression analysis based on German nationwide statutory health insurance data [54]. Costs for complications for a woman of 50-59 years in the first quarter of diagnosis are estimated as follows: fatal MI 8,710€, fatal ischemic heart disease 20,952 €, angina pectoris 2,705 €, fatal stroke 11,186 €, diabetic foot 1,303 €, amputation 14,294 €, retinopathy 681 €, blindness 2,943 €, nephropathy 3,363 € and ESRD 22,701 €. In the following quarters of the year of diagnosis the costs vary from 691 € for retinopathy to 6,140 € for ESRD in total [54].

1.2 Health economic evaluation with RCTs

Medical decisions regarding health care services affect patients' health, lead to consuming health care resources, and have an effect on other economical and societal sectors besides health care [5]. The health care resources in any system are scarce by definition and since there is an interaction between clinical, financial and societal outcomes of any medical intervention, decision makers need comprehensible and objective information about cost (inputs) and consequences (outputs) respectively to benefits to choose the best alternative among several [5, 55]. In general, an economic evaluation is defined as a comparison of one or several alternatives with a default one. The four main types of economic evaluation are cost analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis. For all these analyses the costs are measured in monetary units and therefore easy to compare [5]. The corresponding outcomes depend on the purpose and the study used for the health economic evaluation. To understand modelling, cost-effectiveness analysis and cost-utility analysis are relevant.

In cost-effectiveness analysis relevant output is one single effect or outcome which is equal in all (treatment) alternatives. Outcomes are measured in natural units e.g. life-years gained, symptom-free days, complications avoided. Cost-utility analysis is characterized by the measurement of patients' preferences as utility. The preferences are expressed through a utility function. Hence, single and multiple effects can be expressed by the output. Usual measure is quality adjusted life-years (QALYs) [5, 55].

The gold standard for clinicians to assess (new) therapy respectively treatment options and effects is an evaluation in a randomized controlled trial (RCT). It is mostly used to deliver the relative effectiveness of a new intervention over standard therapy [56]. For economic evaluation with RCTs, the outcomes can be compared with costs for the intervention. The main advantage to use RCTs in economic evaluation is that the relative effectiveness is reliable due to structural balance between study arms through randomization in RCTs. One of disadvantages is firstly that only several options are compared whereby in reality all options of a disease are available. Furthermore, in chronic diseases health care, different treatments might be necessary at the different disease stages. Second, RCTs have a limited time horizon and especially chronic disease have a lifelong impact on patients. Costs of an intervention fall into the time horizon of the RCT but benefits of it can occur after the follow-up. Third, the purpose of RCT can be different to the decision making purpose. RCTs are applied by pharmaceutical companies to show effects in relevant markets. The effects can be biased from pragmatic health care conditions and for different health care systems and countries. However, pivotal trials, multi-national

and/or multi-center trials try to overcome this imperfection by a strict protocol or very complex study design. Fourth, RCTs reveal evidence for the special purpose they are performed but for decision making all currently available evidence should be taken into account to evaluate given strategies or allocate resources efficiently. Hence, economic evaluation based on clinical trials implying RCTs will only deliver a partial analysis [56].

1.3 Decision-analytic modeling

1.3.1 Decision analytic modeling in general

The economic evaluation, usually performed in a RCT, also can be realized or extended through modelling technics, a way of representing the complexity of the real world in a more simple and comprehensible form given in mathematical terms. Modelling involves a theoretical description and a mathematical framework and allows to perform decision making under uncertainty in complex systems. In such a model, natural way of a disease progression is combined with various interventions to simulate, predict and compare outcomes [5]. Clinical and economic features can be defined based on known and estimated inputs and interactions. The accuracy of the prediction is dependent on the assumptions and inputs used. Decision analysis works best if is based on solid evidence. However, often the greatest value of decision analysis done with a help of modelling is when there are too few known inputs to make an easy decision in real world.

The economic and clinical evaluation based on randomized clinical trials has some disadvantages. For example, the different treatment options might be restricted to a few, the evaluation might be performed over a limited time period, the purpose of an original study might deviate from a purpose of a decision making process, and a study might show the limited evidence regarding to population at risk and specific conditions [56].

Simulation models are useful tools to overcome limitations mentioned above. They can provide reliable up-to-date information that is necessary for making a decision in health care that would serve individual and system's needs. Since a new intervention or drug therapy may be more expensive and differ by clinical effectiveness than current alternatives, decision makers from different health care systems may require an economic evaluation based on modeling whether a new therapy should be financed [5]. By using a model, they evaluate medical and public health alternatives and predict outcomes of e.g. T2DM which can be used for a cost effectiveness analysis. A model combines relevant available scientific evidence from various sources with

the corresponding economic information. Moreover, all complications of a disease can be simulated in a life-long perspective. This enables decision makers to choose between different treatment options based on the best available evidence [4]. Key elements of a decision analytic model are (1) the identification of appropriate comparators, (2) implementing of all required clinical data to estimate model parameters, (3) the decision about the extrapolation of mid-term clinical finding to long-term outcomes, and (4) the presentation of results for decision making [55].

Decision trees, Markov models and discrete event simulation models are the three main types of decision analytic models that can be distinguished. Decision trees represent a decision problem as tree diagram in which different outcomes occur with certain probabilities [57]. In contrast to decision trees, Markov models are more compact way to present decision problems. Some health states (e.g. healthy, ill, death) are defined and transition probabilities between the states are calculated or obtained from literature [57, 58]. In discrete event simulation models individuals (entities) pass different processes (events) that affects their outcome (attributes) [57]. Markov models are the most widely used models in health economic evaluation [58, 59].

An example for an economic evaluation based on RCT is the Diabetes Prevention Program (DPP) [60]. The aim of this study was to investigate how far overweight, and sedentary lifestyle are reversible in patients with an elevated plasma glucose level. The outcome was to prevent or delay the onset of T2DM. Patients were randomly assigned to one of the three groups: first, standard lifestyle recommendations plus placebo twice daily, second, standard lifestyle recommendations plus metformin at a dose of 850 mg twice daily, and third, intensive program of lifestyle modification. As result, lifestyle modification and metformin can reduce the incidence of T2DM [61]. For the economic evaluation, clinical outcomes, costs, and QALYLs were used. Over 3 years, lifestyle modification and metformin intervention cost approximately 2,250 USD more per participant than placebo. In the DPP, lifestyle and metformin intervention cost 24,400 USD and 34,500 USD, respectively, per case of diabetes delayed or prevented and 51,600 USD and 99,200 USD per QALY gained [60]. For a liftetime perspective, the analysis was repeated with modeling techniques using the CDC/ RTI Model which is a Markov model with annual transition probabilities between disease states. In this analysis, the lifestyle and the metformin intervention cost 1,124 USD respectively 31,286 USD per QALY gained [62].

An example for an economic evaluation based on modelling is a study performed by Valentine (2006) [63]. The IMS Center for Outcome Research (CORE) Diabetes Model was used to compare long-acting insulin analogue detemir with intermediate-acting Neutral Protamine

Hagedorn (NPH) insulin and long-acting insulin glargine. For the comparison of insulin detemir with NPH insulin costs and QALYs were increased. Therefore, an incremental cost effectiveness ratio of approximately 15,000 USD per QALY gained was calculated. The comparison of insulin detemir with insulin glargine revealed, that costs for detemir were reduced by 5,175 USD compared to glargine and QALYs were increased by 0.063 years. Thus, insulin detemir is dominant compared to insulin glargine [63].

1.3.2 Modeling in T2DM

Diabetes is a chronic disease and trials dealing with it have only a specific time horizon of some years. Decision makers need to choose the best alternative for patients and the health care system on a lifelong perspective. Modeling can overcome this problem [9]. Therefore, modeling in diabetes is used for about 25 years. In 1996, the first model simulated micro- and macrovascular complications for T1DM based on the Diabetes Control and Complications trial (DCCT) [64–66]. As known in a systematic review about modeling in diabetes [65], the first study about modeling in T2DM was published by Eastman et al. (1997) simulating a broad range of micro- and macrovascular outcomes based on the DCCT [66, 67]. Today, modeling is a widely used method to perform economic evaluation of diabetes and other diseases [68].

In general, diabetes models can simulate the outcomes of diabetes on a single patient (microsimulation) or at a cohort level (macrosimulation), or on a complete population level. Microsimulations regard patient characteristics of single patients and simulate their specific disease progress. After simulation of all patients' outcomes they are aggregated to determine mean costs and benefits. Cohort simulations use groups with similar characteristics and compute the consequences on a group level. For more detailed analysis subgroup analysis are possible. Microsimulations are mainly used in T2DM decision analytic modeling because it can incorporate the complexity of a disease progress [4].

Furthermore, it can be distinguished if interaction between individuals of a simulation is allowed or not. Interaction between individuals is not allowed in decision tree models and Markov models. Interaction is allowed in dynamic population models including discrete event simulation models [57]. The easiest way to present and perform a health economic analysis are decision trees. The main structure is a tree diagram. Expected costs and consequences are determined through different clinical pathways which are represented through nodes. The following description reflects this graphical depiction. Each tree begins with a clinical decision. Outgoing from this there are several randomly followed clinical pathways depending on the disease state. The end of a pathway is represented through a clinical outcome. To this outcome

expected costs and consequences can be assorted. Each clinical pathway has a specific probability to be followed. Evaluation of a decision is proceeded by roll back process in which the generated values are rolled back over the different pathways to the beginning to determine the overall cost-effectiveness ratio. Decision tree analysis has some limitations. They are not suitable for the processes when health states change often, the consequences are delayed and do not occur in a single period, or parameters change over time and when events re-occur (i.e., hypoglycemia after medication). However, these aspects are highly relevant for a chronic disease like T2DM. As consequence, decision trees analysis is used rarely in T2DM modeling.

Currently, Markov models and discrete event models are the two most established model structures in T2DM modeling [4, 55].

First, Markov models are composed by states, transitions and probabilities. Health states are mutually exclusive health conditions e.g. healthy, ill and dead. The transition between different health states is expressed through a certain probability and time cycles. Time is usually running by equal cycles in these models. In standard states, it is allowed to stay in the same state for an unlimited number of cycles. In a tunnel state, however, there is no possibility of "stopping" and the state must be abandoned at the end of the cycle. Transition probabilities between the different conditions are calculated. At the end of a cycle, pay-offs can be attributed to health conditions and the effectiveness of decisions can be estimated. Decision trees and Markov models can be combined where several Markov models represent different treatment options in a decision tree. Limitations of Markov models are when several diseases like macro- and microvascular complications are considered and they can occur in different combinations. Furthermore, interaction between individuals cannot be modelled. Then the Markov model would get very complex and unclear [4, 55, 69].

The second model type can handle these issues. Discrete event simulation models utilize entities (e.g. patients) with different attributes (e.g. age, gender, status of disease). Calculations are performed when an attribute changes and costs can be allocated to the attributes [4, 55]. Disadvantages are that discrete event simulation models can get very complex due to a lot of possibilities and interaction between individuals [57, 58].

These models are widely used nowadays since they can simulate all complications of the disease and are not focused on a specific problem or a particular sample of patients of interest, unlike trials [4, 68].

Until today, six reviews were performed about modeling in diabetes. Two reviews included papers until 2008 to give an overview of actual models used in diabetes modeling [65, 66, 70].

A following review updated the data between 2008 and 2013 [66, 71]. A review from 2011 evaluated the quality of the following three models: Archimedes Model, CDC/ RTI Model and IMS Center for Outcomes Research (CORE) Diabetes Model [66, 72]. In 2015, Henriksson et al. presented an overview about models in T1DM [66, 73]. In the same year, Kirsch analyzed Markov models which evaluated multicomponent disease management programs [66, 74]. A more detailed descriptions of the reviews concerning external validation is given in the section 'Objectives'.

Since 1999, 9 MtHCh meetings took place where actual topics about modeling in diabetes are discussed [6]. The 4th, 5th, and 9th deal with external validation of models but there is no exact consensus about what model validation means in detail [8–10, 75]. Actual topics discussed in the 9th MtHCh are the ability of models to reproduce outcome of recent cardiovascular outcome trials. Calibration of the models to improve their accuracy and the development of new risk equations to simulate cardiovascular outcome more accurate are future research questions [9].

1.4 Types of model validation

Being a useful instrument, simulation models, nevertheless, must prove their validity to fulfill their purpose. Validation is a method to appraise how well models perform. There exist several types of validation depending on the purpose and application of a model. Therefore, it does not exist a check-list what steps have to be performed for a validation exercise [11, 12].

Six kinds of validations can be identified in literature [4, 12]. Firstly, face validity means that results of a validation process are plausible to experts of the area, easily comprehensible and all relevant aspects are embraced by the model. Relevant steps of face validity include assessing the comprehension of the model structure, data sources, problem formulation and results. Concerning structure experts should assess if all relevant aspects of clinical reality are covered by the model and alternative are consistent with medical practice. Furthermore, experts can evaluate if the data used for modeling exercise is sufficient and on the actual state of evidence. For problem formulation, experts should appraise the application report if the plan of a study, assumptions, covered time period, population in focus, chosen interventions, and outcomes are suitable to answer the research question. Results should be assessed with their consistence of reality and experts' believes. Still, experts could give an assessment of the performance of the model but this process is subjective by its nature. Face validity is helpful to increase the credibility and the acceptance of results in scientific society and in public. However, models are simplifications of reality and not all aspects can be represented. Medical practice can be

inconsistent with medical evidence and face validity can enable model structures which are not consistent with available evidence [4, 12, 76, 77].

Secondly, a debugging should be performed by controlling mathematical calculations. This process consists of the evaluation of formulas and their incorporation in the source code. If a model is debugged the source code should be on the actual state. In an appraisal, a modeler can explain the equations and source code to other experts to assess mistakes. Furthermore, a double programming by two modelers could be performed to assess source code correctness. Limitations of debugging are that only the mathematical and programming aspects are evaluated and an assessment of accuracy of the model does not take place [4, 12, 76].

Thirdly, after a model has been set up, its functionality is verified using the data which have been used to build the model. This process is called internal validation and means that the model should reproduce the same results which are given in the original dataset. Similar to debugging limitations of internal validity are that only the performance of the model to reproduce the data original data source is assessed and the accuracy to make general predictions is not appraised [11, 12, 78].

Fourthly, in external validation exercises models can be tested against external or new data which was not used to set up the model [4]. This process is discussed in the following section in detail.

Fifthly, cross-validation refers to a comparison of two or more models which provide the simulated results under the same initial conditions and with a use of the same data set as an input. Ideally, the models should return the same results. The relevance of cross validation depends on methods and data source used to build the models. If models used the same data sources or transition probabilities inherited from former models, they might be considered as dependent. Thus, results are expected to be similar and the significance of cross validation is limited. In case of independent models that fundamentally differ, any variation should be explained [4, 12, 76].

Sixthly, predictive validity is used when a study is in progress and predictions of the population are simulated before the end of the study so that the results are unknown. This exercise is not regarded to be essential, but it is an independent type of validation because results of the simulated study are unknown therefore it is considered as best way of validation because it predicts future events. A difference of prediction and later results is not a big issue because important aspects cannot be known when the simulation is performed. If new evidence is available a good

model should be recalibrated. A limitation of predictive validity is that decision makers need to decide in short time periods and the end of a trial can be in distant future [4, 12, 76].

1.4.1 External validation

External validation serves as a method to assess if a model is generally applicable and reliable. Hence, simulation exercises have to be performed with data which was not used to set up a model [4]. Relevant new data sources might be trials or epidemiological studies. Ideally, models should be regularly validated against major / 'landmark' studies of the scientific field where a model is used. The best way to find suitable source is a formal search. Depending on the purpose of external validation and the data source, the validation can be performed to a whole or just to parts of a model, for example disease incidence and progression, occurrence of clinical outcomes, effects of interventions, care process or behavior of patients. The validation exercise should be as accurate as possible with regard to study population and study protocol. Requirements for trials are that they are applicable to the model and data contains information about the setting, population, treatment protocol, follow-up protocol and description of outcomes. It is important to control if these aspects correspond between data source and models because a mismatch in simulation is foreseeable. If possible, all subjects of a study should be included in the validation exercise and patient specific data should be used as possible. Otherwise, distribution of patients and their characteristics or patients with mean scores can be used. Furthermore, external validation should use the same statistical methods that were used in the respective study. For trials group differences and for all other simulations absolute values should be presented. Statistical analysis should indicate effects of sampling error with respect to sample size, metrics to appraise the statistical significance for expected values and sensitivity analysis of the results. A report of comparison should contain a detailed description of data source, simulation setup, results with evaluation of discrepancies. Papers cannot illustrate all aspects of a validation exercise so a detailed description of the process should be available on request. If multiple data sources were used for external validation it occurs that results of models do not match every time with data. It should be explained why there is a discrepancy and if the model fails it should be recalibrated to deliver more precise results. Modelers should prove if the revised model is still externally valid [11–13].

MtHCh meetings are the opportunity for modelers to perform external and cross validation and to discuss problems and actual topics with research colleagues. Therefore, they are important events to improve the accuracy of models in diabetes.

1.5 Objectives

In the literature, six reviews were identified that deal with health economic aspects of diabetes modeling and model assessments in diabetes field [66]. The first review by Tarride et al. described the methods used in health economic modeling of diabetes. Concerning external validation, it is mentioned in the review that several models perform internal and external validation and the results of the 4th MtHCh [10] are presented [65]. The second review appraised T2DM models published until 2008 and evaluated the models themselves and captured treatments through the models. For 10 models it was described if external and internal validations were performed. Furthermore, the results of the validations were outlined shortly [70]. The third review covered new models which were published between 2008 to 2013. The authors indicated if external or internal validation was performed in the models but a detailed description of methods was missing [71]. The fourth review covers the performance of three models whereby the frequency of external validation exercises was given [72]. The fifth review deals with modeling methods and their quality in T1DM models. In the fifth review, it is only mentioned if external validation was performed or not [73]. As well in the other review, an economic evaluation of multicomponent disease management programs with Markov models was presented in the sixth review. Validation was not mentioned here [74]. By my knowledge, there is no review that focused on and summarized the external validation.

There are several guidelines and recommendations for external validation such as from the American Diabetes Association (ADA), the Professional Society for Health Economics and Outcomes Research (ISPOR) Task Force and The Mt Hood 4 Modeling Group [10–13]. However, there is no systematic overview of current practices in external validation approaches and whether these attempts are made in accordance with the guidelines.

The main objective of this thesis is to conduct a systematical literature review to describe and evaluate the process of external validation and to provide an overview and recommendations for performing external validation in the future.

Papers were gathered by a formal systematic search in different literature sources and screened to their relevance. This process is described in detail in ‘Material and Methods’. The evaluation of external validation includes the description of the validation process in the identified literature and the thoroughness of this process. Especially definition, description and use, data sources used in the external validation, and results of external validation will be described.

2 Material and Methods

The SLR was conducted referring to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [79]. The review is registered on the website of international prospective register of systematic review (PORSPERO) under the no. CRD42017069983 [66].

The process of conducting the SLR contains the following steps:

1. Construction of search terms;
2. Search in databases and extract hits;
3. 1st stage screening: title and abstract screening for appropriate models;
4. 2nd stage screening: full text screening for external validation in the models found on the previous step;
5. Additional data sources selection;
6. Data extraction;
7. Data synthesis/ analysis.

Steps one and two were performed in advance and are not part of the thesis. For comprehensibility, they are described in the subsequent part. The part of the study covered in this thesis contains the repetition of the search, and the steps 3-7 above.

2.1 Construction of search terms

The search terms were developed and adapted for the MEDLINE search engine according to its search rules and translated into other engines' syntaxes (protocol available) [66]. These original search terms are given in Appendix Table 11. The sensitivity of the search strategy was checked by matching the search hits with other references known from previous SLRs and other similar studies.

2.2 Search in databases and extract hits

The literature search was conducted in 15 data bases Table 3 (in results): Central (via Wiley), CINAHL (via EBSCOhost), CDSR (via Wiley), CMR (via Wiley), DARE (via Wiley), EconLit (via EBSCOhost), EMBASE (via Ovid SP), MEDLINE (via NLM, PubMed), NHS EED (via Wiley), PsycINFO (via Ovid SP), Scopus (via Elsevier) and Web of Science (via Thomson Reuter). The ensuing Grey literature databases were also searched: ProQuest Dissertations & Theses Database (PQDT) (via ProQuest), System for Information on Grey Literature in Europe (Open Grey) (via INIST/CNRS) and The Directory of Open Access Repositories (OpenDOAR)

(via CRC) (protocol available). Also, further search was performed in Cochrane Database of Systematic Reviews (CDSR) and the Database of Abstracts of Reviews of Effects (DARE) to detect existing reviews similar to the topic of this thesis. The Grey literature search served to expand the scope of the search and to include models that are not mentioned in previous SLRs. In addition, models externally validated in MtHCh meetings were also included in the analysis [9, 10, 75]. Finally, hand search in PubMed and Google Scholar was performed.

The search was conducted twice: in September 2017 and in May 2020. First, the time frame from 1997 to September 2017 was chosen. 1997 was taken as the starting point because this was the year when the first paper on diabetes modeling was published [66, 67]. To keep the database up to date after 2017, the search was repeated once with publication dates restricted to a period from September 2017 to May 2020.

The references were stored in Mendeley, a cloud-based reference manager [80].

2.3 Screening

The screening for the data sources was conducted in two stages: title and abstract screening for models, that met the given criteria listed below, and full text screening to finalize the list of the models and to indicate if external validation was performed in the found models. Later, the data extraction and data synthesis were performed only on the subset of the models that were externally validated.

At first stage, the abstract screening was conducted on Rayyan, an abstract screening service [81]. Relevant hits were screened in two independent groups. The first group consists of Marko Frings and Katherine Ogurtsova and the second group of Ute Linnenkamp and Thomas Heise. The screening was performed according to predefined inclusion (first three items) and exclusion criteria 1-10 giving in Table 2. The aim of abstract screening was to find as many models as possible and to keep the data base as large as possible.

Therefore, the first three inclusion criteria for the search were kept broad, and the last inclusion criteria and exclusion criteria were not strictly applied on the first stage for the following three reasons. Frist, external validation is partly only a minor aspect in papers about modeling and is hence not mentioned in the abstracts. Second, external validation is sometimes only mentioned in papers. The validation process itself might be described in a technical report, which could be available only as supplementary material online or on request. Third, I aimed at evaluating how frequently external validation was performed in the literature, which is why I intended making the search broader.

Inclusion criteria	Exclusion criteria
1. The model focuses on incidence and/or progression of T2DM (model structure criteria) 2. It is based on simulation techniques 3. The simulation is built on individual units representing an organ, a person or a cohort 4. (only full text screening) External validation (EV) was performed for the relevant model	1. Unsuitable publication type: a. Editorials b. Comments c. Newspapers d. Systematic review 2. No abstract available 3. No full publications available 4. Only conference abstract given 5. Diabetes is mentioned only as a comorbidity 6. No model is mentioned or described 7. A model does not include the onset and/or the development of T2DM but focuses on a partial aspect of the disease 8. A model focuses on T1DM only 9. No simulation is performed 10. Simulation inappropriate: the simulation is based not on individual units: an organ, a person or a cohort. 11. (only full text screening) a. Models not reporting EV b. Models where EV only mentioned, no further information found c. Models where EV only in abstract

Table 2: Inclusion and exclusion criteria

The abstract screening was piloted in each group and then discussed between the groups to adapt the procedure. Inconsistencies were discussed and resolved within and between groups. Exclusion criteria were not strictly applied on conference abstracts and posters if a model or external validation was mentioned. On the later steps, the full text papers that might follow conference abstract were additionally searched or authors and modeling groups were contacted.

At the second stage, full text screening was performed to finalize the list of the models fitted to our criteria and to find if the external validation in these models was carried out. Full papers

were screened by one group of two researchers (Marko Frings and Katherine Ogurtsova). The full text papers were scrutinized in details regarding all exclusion and all inclusion criteria. The attention was drawn to additional inclusion and exclusion criteria (inclusion criteria 4 and exclusion criteria 11). Papers were excluded if any of the exclusion criteria were met or if not all of the 4 inclusion criteria were matched. Having attempted to obtain a broad database in the previous step, it was now assessed whether external validation was a topic of the full text papers.

2.4 Additional data sources selection

In the additional data selection, authors of known models were contacted to assess if they have additional information or technical documentation available. If external validation was only mentioned in the full text paper, the modelers were also contacted to find out whether they had additional information on external validation or a manual search was carried out for additional information. Also, conference abstracts where external validation was mentioned were used as a starting point to search for further publications of these authors and research groups, or authors and modeling groups were contacted.

SLRs found during the search were included to expand the scope as an additional data source. Also, reports on MtHChs were selected as an additional important data source. MtHChs are regular meetings of the research field to discuss relevant topics in diabetes modelling. The 4th, 5th, and 9th deal with external validation of models. The presentation of the results is therefore based on the data obtained through the title and abstract and full text screening and the data obtained through the additional search. If models only performed external validation on MtHChs, this is shown separately in the results. Otherwise, the collected data are presented together per model.

2.5 Data extraction

The list of all models found in the search was used to rate the widespread presence of external validation. Only those models in which external validation was preformed were subject to data extraction and synthesis.

The data was extracted in a predefined data extraction spreadsheet to identify the model characteristics and the extent and characteristics of the performed external validation. The spreadsheets were stored in the cloud storage Google docs [82]. Since there was no standardized tool for data extraction until 2016 [83], the spreadsheet was created by ourselves (Marko Frings and Katherine Ogurtsova). The spreadsheet was constructed in accordance with the relevant aspects of the guidelines of the American Diabetes Association (ADA), the Professional Society for

Health Economics and Outcomes Research (ISPOR) Task Force and The Mt Hood 4 Modeling Group [10–13, 84]. If information were not found, they were labeled as ‘unavailable’. The data was collecting according to following categories: of the screening

1. Model characteristics;

2. External validation: definition, description and use;

3. Data sources used in the external validation: description and use;

4. Results of external validation: methods and reporting formats [66].

This process was also piloted to adapt the procedure. Inconsistencies were discussed and resolved again.

The screening and paper selection process is presented as PRISMA flow diagram where included and excluded number of papers are presented [79].

2.6 Data synthesis/ analysis

The analysis is presented in a narrative and descriptive way due to diversity of the models and the external validation methods. First, the results of the search in the different databases are presented, and the results of the database search and the screenings are shown as a PRSIMA flow diagram and reasons for the exclusion of hits are given. Then an overview of the models found and what kind of validation was done for each model is presented. Additionally, an overview of the MtHChs and the external validations performed at the meetings is provided. Then, the narrative review of the definitions, descriptions and use of external validation per model are given. The processes of how the modelers found data sources and/or studies for external validation is outlined, and all relevant studies for performing external validation in the literature are presented. The results of external validation in the found literature are illustrated in a table. Thereby, methods used for external validation, risk factors necessary for setting up the initial population, relevant outcomes and the presentation of the external validation process from the research groups are described. Finally, an overview is given of the extent to which the models fulfilled the criteria of the guideline. This structured information reflects the external validation practice and performance.

3 Results

3.1 Search in databases and Screening

The search in the databases, extracting hits and the analysis were conducted in September 2017 and repeated in May 2020. Table 3 shows the number of hits found in bibliographic and grey literature databases. Overall, 21,737 studies we found through the search. The highest number of the hits was in Web of Science with 6,076 hits, followed by Scopus with 4,069 and Medline with 3,574 hits. Just two studies were found in CDSR and DARE.

Bibliographic Databases:	**No of hits**
• CENTRAL (via Wiley)	357
• CINAHL (via EBSCOhost)	893
• CDSR (via Wiley)	2
• CMR (via Wiley)	9
• DARE (via Wiley)	2
• EconLit (via EBSCOhost)	1,440
• EMBASE (via Ovid SP)	3,353
• MEDLINE (via NLM, PubMed)	3,574
• NHS EED (via Wiley)	62
• PsycINFO (via Ovid SP)	244
• Scopus (via Elsevier)	4,069
• Web of Science (via Thomson Reuters)	6,076
• **Grey literature databases**:	
• ProQuest Dissertations & Theses Database (PQDT) (via ProQuest)	1532
• System for Information on Grey Literature in Europe (OpenGrey) (via INIST/CNRS)	24
• The Directory of Open Access Repositories (OpenDOAR) (via CRC)	100
Total with duplicates	**21,737**
Total without duplicates	**12,807**

Table 3: Number of hits in searched databases

The following presentation of the results is composed of two parts. First, the selection of relevant papers from the search is shown (paper-based models). Second, the models that were only represented at MtHChs are presented.

First, the total statistics of the studies selection is given in PRISMA flow diagram in Figure 2. Additional to 21,737 references found and extracted by formal search, 214 studies were found and extracted from systematic reviews, or from reference lists of eligible papers. After reduplication, 8,930 (41%) papers were removed and 13,021 were available for the abstract screening. The four-eyes-principle was used to perform the abstract screening in two groups. With application of the mentioned inclusion (criteria 1-3) and exclusion criteria (criteria 1-10) of the methods section (Table 2), 12,261 records were excluded and in total 760 records were selected for full text screening. These number consists of full texts and relevant abstracts. Conflicts were discussed and resolved. The total number of hits comprises of 630 papers found through primary search (2017), 67 studies from secondary search (2020), and 63 studies from additional data sources were identified and included in the full text screening. A total of 79 different models were identified in all of 760 records (Table 12 in Appendix).

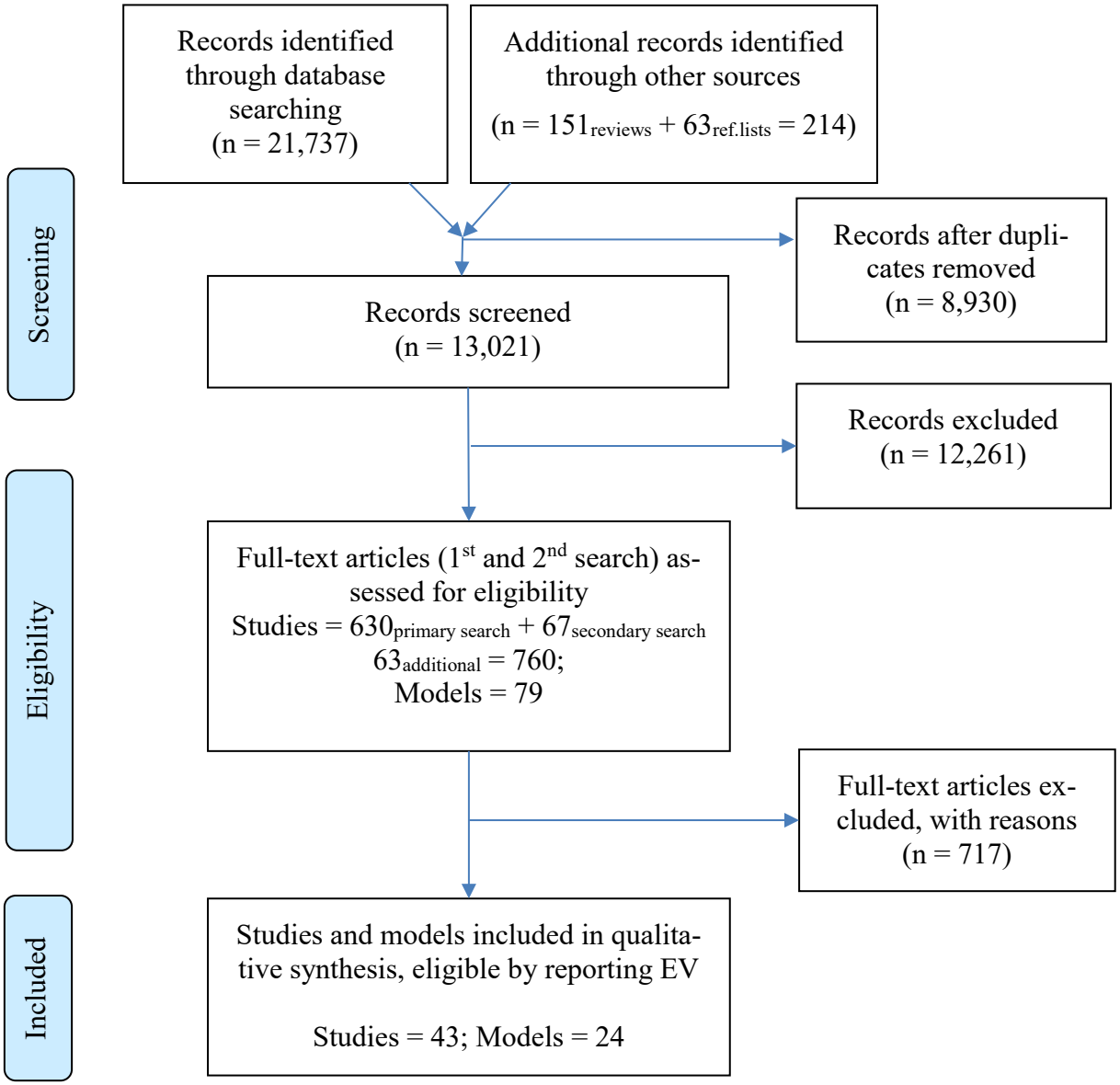

Figure 1: PRISMA flow diagram for the screening process, adapted from [79]

Among all reference eligible for full text screening, 19 papers were written in other languages than English. All these papers, except written in German, were read with help of special translation software or external translators. The papers were assessed following the general rules despite the language they are written. The results of this process are presented in Table 4. In total, 7 papers written in other languages than English were included in full text screening. The largest share of included papers was those written in German.

Language	N. of publications in full text screening	N included in data extraction
Spanish	4	1
Bulgarian	1	0
Chinese	4	0
German	7	4
Farsi	1	0
Japanese	1	1
Lithuanian	1	1
Total	19	7

Table 4: Statistics of other language hits

In case of missing information, I have contacted 59 corresponding authors to request more information. 19 researchers replied (32 %) and 19 additional publications were provided by authors.

The identified papers eligible for the full text screening were screened by two researchers (Marko Frings and Katherine Ogurtsova). The papers fitted to all inclusion and exclusion criteria **(**Table 2) were identified. At the end of the full text screening, 43 papers (referring to 24 models) were included for data extraction and appraisal.

Reason for exclusion	N of excluded
Duplication	6
Conference Abstract	296
Review	29
Inappropriate publication type	4
No full publication available	10
No model given	6
No simulation given	8
Only reference of previous publications	124
Inappropriate model/simulation type	63
T1DM only	9
No validation mentioned or described	147
Model where EV only mentioned	8
Models where EV only in abstract	7
Total excluded at the first step:	717

Table 5: Reasons for exclusion in full text screening Number of hits for searched databases

In total on a paper level, 717 papers were excluded. The most frequent reasons for exclusion (Table 5) were: only a 'conference abstract' was available (296 records), 'no validation mentioned or described' (147 records), 'only reference of previous publications' (124 records), and 'inappropriate model/ simulation type' (63 records).

On the model level, for 55 models (70 %) no external validation was mentioned or reported (Table 6). For 44 models (80 % of 55) no external validation was reported, for 3 models (5 % of 55 models) external validation exercises were mentioned in a paper but no further details were given, and for 8 models (15 % of 55 models), there was external validation only mentioned in an abstract but no publication could be found.

External validation performed on level of models		
No	55	70 %
• Models not reporting external validation	44	80 % (of 55)
• Models where external validation only mentioned	3	5 % (of 55)
• Models where external validation only in abstract	8	15 % (of 55)
Yes	24	30 %

Table 6: Statistics of the performance of external validation on level of models

For 6 models external validation (Sheffield Diabetes Model, EAGLE, Evidence-Based Medicine Integrator, PROSIT Model, TTM, and IQVIA CORE Diabetes Model) was only performed on MtHChs. In the presentation of results, paper-based models are separated from those that were only represented at MtHChs. Only for the calculation of proportions are both taken together. A total of 30 models were included in the final analysis, of which 24 are based on paper-based models and 6 were only represented on MtHChs (Figure 3).

3.2 Characteristics of models

Table 7 gives an overview of the models, with their key facts, the model types, the types of validation, the papers with external validation and the participation in the MtHChs. Of the 30 models included, 17 are Markov models (57 %) and 13 (43 %) are discrete event simulation models. For 24 models (80 %), other validations such as face validation, verification, cross, or internal validation were performed next to external validation. For 6 models (20 %, NHS model, Chicago Model, Markov model/Vijan, Dutch diabetes model, Australian Diabetes Models, and Discrete event simulation model from Jiao) [85–90] only external validation was performed. No more details are shown, as the focus of this work is on external validation.

Model	Sources	Model aims and key facts	Model type	Types of validation involved	Papers based on external validation	Participating on Mt HCh
IMS CORE Diabetes Model	[91]	Projecting long-term health outcomes and economic consequences by taking into account baseline cohort characteristics and past history of complications, current and future diabetes management and concomitant medications, screening strategies, and changes in physiological parameters over time	Markov model structure with combination of Monte Carlo simulation	face validation, verification, cross, internal, and external validation	[10, 75, 92, 93]	2nd, 3rd, 4th, 5th, 6th, 7th and 8th
Cardiff Model	[94, 95]	Cardiff Stochastic Simulation Cost-Utility Model (DiabForecaster), which evolved from the Eastman model, was used to follow a cohort of 10 000 patients over 20 years. The model was designed to simulate health related outcomes and long-term economic impact of treatments of T2DM patients.	Discrete event stochastic simulation model	face validation, verification (consistency), cross, internal, temporal, and external validation	[9, 10, 95]	4th and 9th
UKPDS Outcomes Model	[96, 97]	The first version of UKPDS Outcomes model is based on the the United Kingdom Prospective Diabetes Study (UKPDS) and was published in 2004. The aim is to simulate the occurrence of major diabetes-related complications over a lifelong perspective and to generate health economic outcomes such as quality-adjusted life expectancy. The second version was published 2013 and included additional recent data and an internal validation of UKPDS over 25 years of follow-up.	Discrete event simulation model	Cross, internal, and external validation	[9, 10, 75, 98–100]	3rd, 4th, 5th, 6th, 7th, 8th, and 9th

Archimedes Model	[101]	Detailed simulation of diabetes (including factors anatomy, pathophysiology, test, treatments, and outcomes) that can be applied to a broad variety of clinical and administrative problems	Discrete event simulation	face validation, verification, internal, cross, external, and predictive validation	[10, 102–105]	4th and 7th
IHE/ ECHO-T2DM Model	[106–108]	IHE/ ECHO-T2DM is a long-term diabetes model to simulate the cost-effectiveness of different diabetes treatment options. Micro- and macrovascular complications and death are simulated.	Stochastic microsimulation Markov model	face validation, verification, cross, external validation	[9, 75, 106–108]	5th, 6th, 7th, 8th, and 9th
CDC/ RTI Model	[109]	CDC/ RTI Model simulates the development and progression of type 2 diabetes and is used to perform cost-effectiveness analysis of different prevention and treatment options. Micro- and macrovascular complications are simulated.	Markov model	cross, internal, and external validation	[9, 75, 109]	5th, 6th, 7th, 8th, and 9th
Michigan Model for Diabetes	[110]	Model that simulates the impact of screening, prevention, and treatment of T2DM treatment options on the development, progression, quality of life, and costs of type 2 diabetes.	Markov model structure with combination of Monte Carlo simulation	verification, cross, internal, and external validation	[9, 75, 111–113]	3rd, 5th, 7th, 8th, and 9th
NHS model	[85]	Model is based on UKPDS Outcomes Model with adjusting nephropathy with biomarkers such as albuminuria and eGFR.	Discrete event (time) simulation model	external validation	[85]	none

Chicago Model	[114]	Chicago Model was developed to estimate life expectancy and the risk of complications. It accounts for demographics, functional status, comorbid illness, risk factors, and duration of diabetes.	Markov model structure with combination of Monte Carlo simulation	external validation	[86]	none
DiDACT model	[115]	DiDACT model model was developed to simulate expected health outcomes and costs attributable to T2DM for evaluation of policies and interventions in treating T2DM.	Markov model	cross, and external validation	[10, 116, 117]	3rd, and 4th
MICADO	[118]	Model simulates long-term effects of interventions in people with and without diabetes. The model takes into account a population scope and has the possibility to assess parameter uncertainty using probabilistic sensitivity analyses.	Markov model	cross, and external validation	[9, 118, 119]	7th, 8th, and 9th
Markov model/Vijan	[87]	Markov model which simulates the onset and progression of retinopathy and nephropathy in patients with T2DM.	Markov model	external validation	[87]	none
Dutch diabetes model	[88]	Model simulates diabetes and its complications over a life-long perspective in the Dutch population.	Markov model	external validation	[88]	none
SPHR Diabetes	[120]	The model was developed to simulated long-term health outcomes and healthcare costs for the evaluation of diabetes prevention strategies.	Kind of discrete event simulation model	verification, cross, internal, and external validation	[9, 120]	8th, and 9th
Microsimulation/Javanbakht	[121]	The model was constructed to simulate the progress of diabetes within the Iranian population over 22 years and the economic impact.	Markov model	face, cross, internal, and external validation	[121]	none

IHSD DPMM	[122]	Interaction between demographics, smoking status, biometrics, incidence of disease and adverse health events, and mortality were simulated.	Markov model	external, and predictive validation	[122, 123]	none
Microsimulation:model form Caro JJ /Eastman	[124]	The model simulates micro- and macrovascular complications of patient with T2DM from diagnosis to death.	Markov model structure with combination of Monte Carlo simulation	face validation, verification, predictive validation	[124]	none
Australian Diabetes Model	[89]	The model simulates the incidence and progression complications of diabetes within the Australian population.	Discrete event simulation model	external validation	[89]	none
Discrete event simulation model from Jiao	[90]	The model simulates patients over a lifelong perspective from no complication to death.	Discrete event simulation model	external validation	[90]	none
PREDICT-DM	[125]	The model simulates the disease progression and compares different treatment options.	Markov model with combination of Monte Carlo simulation	verification, internal, and external validation	[125]	none
CDOM	[126]	A Chinese specific outcome model.	Markov model with combination of Monte Carlo simulation	internal, and external validation	[126]	none
Cornerstone Diabetes Simulation	[127]	The model simulated over a lifelong perspective health outcomes and economic consequences of T2DM	Discrete event simulation model with comibation of Monte Carlo simulation	cross, internal and external validation	[127]	none
Chinese Outcomes Model for T2DM	[128]	The model simulates complications of diabetes within the Chinese poplulation.	Discrete event simulation model	face validation, verification, cross, internal, and external validation	[128]	none
BRAVO Model	[129]	The model accounts for increased risk of complications as diabetes progressed and for interactions between complications.	Discrete event simulation model	cross, internal, and external validation	[9, 129]	9th

Sheffield Diabetes Model	[130]	The Sheffield Diabetes Model simulates the natural history of diabetes and the lifetime cost effectiveness of different treatment options.	Decision tree and Markov model	cross, and external validation	[10]	3rd, and 4th
EAGLE	[131]	Model was developed to implement new features and further develop the methodology of existing models to present a more comprehensive assessment of long-term effects of diabetes treatment and related costs.	Markov model structure with combination of Monte Carlo simulation	verification, cross, internal, and external validation	[10]	3rd, and 4th
Evidence-Based Medicine Integrator	[132]	Evidence-Based Medicine Integrator combines risk equations based on the underlying population to build the model, with estimates of treatment effects taken directly from systematic reviews and randomized trial results.	Discrete event simulation model	cross, and external validation	[75]	5th
PROSIT Model	[133]	PROSIT Model is not a single model because six diabetic complication models (CHD, stroke, nephropathy, retinopathy, amputation, and hypoglcaemia) are combined.	Markov model	cross, and external validation	[9]	9th
TTM	[9]	The model estimates clinical and economic outcomes for patients with T2DM under user-specified treatment strategies	Kind of discrete event simulation model	cross, and external validation	[9]	9th
IQVIA CORE Diabetes Model	[9]	The model is a web-based diabetes policy analysis tool that performs real time simulations to predict clinical outcomes and costs for cohorts of patients with diabetes	Markov model structure with combination of Monte Carlo simulation	cross, and external validation	[9]	9th

Table 7: Overview of model characteristics and validation types, adapted from [134]

3.3 Mount Hood Challenge participation

MtHChs are meetings to discuss relevant topics of modeling in diabetes. Since 1999, 10 MtHChs took place. The last MtHCh was held in South Korea at the end of 2019 and there are no detailed information published until now [135]. So, this MtHCh was not regarded for the following results (see Table 7 and Table 8).

Model name	MtHCh1	2	3	4	5	6	7	8	9	Participates
IMIB Model	Yes									1
Global Diabetes Model (GDM)	Yes	Yes	Yes	Yes						4
CDC Diabetes Prevention Model		Yes								1
CDC/ RTI Model				Yes	Yes	Yes	Yes		Yes	5
IMS CORE Diabetes Model		Yes	Yes	Yes	Yes	Yes	Yes	Yes		7
IQVIA CORE Diabetes Model									Yes	1
IQVIA CORE Hypertension Model									Yes	1
Economic Model of Diabetes progression		Yes								1
Michigan Model for Diabetes			Yes		Yes		Yes	Yes	Yes	5
IMOR Model		Yes								1
UKPDS Risk Engine		Yes	Yes	Yes	Yes					4
UKPDS Outcomes Model			Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
DiDACT model			Yes	Yes						2
EAGLE			Yes	Yes						2
DiabForecaster Model (Same model like Cardiff Model)				Yes						1
Sheffield Diabetes Model			Yes	Yes						2
Archimedes Model				Yes			Yes			2
Cardiff Model			Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
IHE/ ECHO-T2DM Model					Yes	Yes	Yes	Yes	Yes	5
Evidence-Based Medicine Integrator					Yes					1
Reference Model						Yes	Yes	Yes		3
Diabetes Modelling and Analysis Framework (DMAF)							Yes			1
MICADO: Modelling Integrated Care for Diabetes based on Observational Data							Yes	Yes	Yes	3

Ontario Diabetes Economic Model							Yes			1
Medical Decision Modelling Group (MDM) - Treatment Transitions Model (TTM)								Yes	Yes	2
MMUs Diabetes Model								Yes		1
SPHR Diabetes								Yes	Yes	1
SPHR CVD Prevention Model									Yes	1
BRAVO Model									Yes	1
PROSIT Model									Yes	1
Scottish CVD Policy Model									Yes	1
SHARP-CKD-CVD Model									Yes	1
	2	6	9	11	8	6	11	10	15	

Table 8: Participating models in Mount Hood Challenges [135]

The MtHChs with the most participating models was 9th MtHCh with 15 models, 4th and 7th MtHCh with 11 models and 8th MtHCh with 10 models. On 1st MtHCh, only two models participated. The most often participating models were IMS CORE Diabetes Model, UKPDS Outcomes Model and Cardiff Model with 7 participations. IHE/ ECHO-T2DM Model and Michigan Model for Diabetes participated both 5 times. Only one time participated 17 models (IMIB Model, CDC Diabetes Prevention Model, IQVIA CORE Diabetes Model, IQVIA CORE Hypertension Model, Economic Model of Diabetes progression, IMOR Model, DiabForecaster Model, Evidence-Based Medicine Integrator, Diabetes Modelling and Analysis Framework, Ontario Diabetes Economic Model, MMUs Diabetes Model, SPHR Diabetes, SPHR CVD Prevention Model, Bravo Model, PROSIT Model, Scottish CVD Policy Model and SHARP-CKD-CVD-Model).

Of the 24 paper-based models, only 11 models out of 24 (46 %) have participated in a MtHCh at least once. 13 models out of 24 (54 %) have published external validation results only via papers (Figure 3).

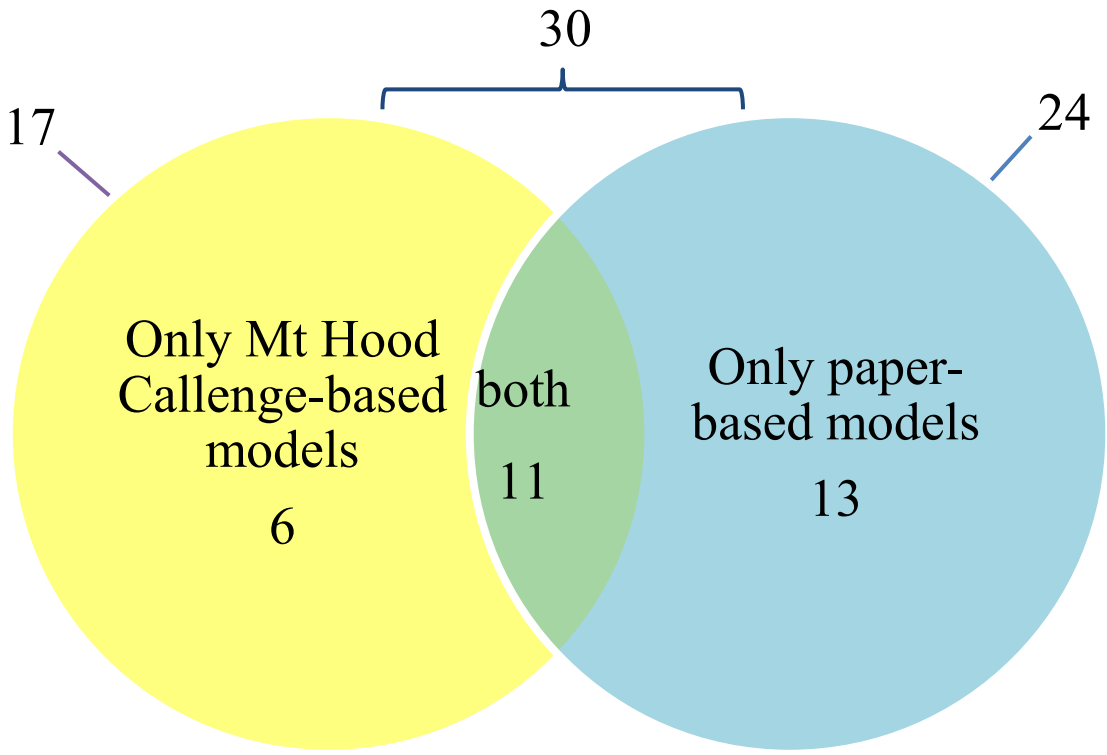

Figure 2: MtHCh and paper-based models

External validation exercises were performed on 4th, 5th and 9th MtHChs [9, 10, 75, 135]. At the 4th MtHCh, the validation of 8 models was carried out against CARDS (Table 7 and Table 8). Baseline characteristics and risk factors were given to start the simulation exercises. Simulated endpoints were cumulative incidence of cardiovascular events, all-cause mortality, and quality-adjusted life expectancy. The result was an overestimation of the incidence of macrovascular complications [10].

At the 5th MtHCh, 8 models carried out external validation with the ACCORD, ADANCE and ASPEN studies (Table 7 and Table 8). Baseline characteristics and risk factors were again provided as the basis for simulation. The summed outcomes are shown in Table 18 in Appendix, among others. The models performed well in simulating the relative risk between the intervention and control arms, but did not perform as well in simulating the absolute risk of diabetes complications [75].

External validation of 12 models was carried out at the 9th MtHCh using the EMPA-REG and CANVAS studies (Table 7 and Table 8). Baseline characteristics and risk factors were provided as the basis for simulation. The summed outcomes are again shown in Table 18 in Appendix, among others. In general, treatment effects were underestimated for both studies. The results

could be improved by recalibrations of the models [9]. The detailed results of the MtHChs are presented for the corresponding models in Table 15 to Table 18 in Appendix, among others.

3.4 External validation

The participation in the MtHChs has already been treated in the previous section and will not be discussed further in this section.

In the following, established models are considered for which external validation has been described in at least 2 papers and which have been externally validated on minimum 2 MtHChs. These include 5 models (IMS CORE Diabetes Model respectively IQVIA CORE Diabetes Model, UKPDS Outcomes Model, IHE/ ECHO-T2DM Model, and Michigan Model for Diabetes). The other 25 models are therefore referred to as "less established" models. Mainly, besides external validation other types of validations (face validation, verification, cross, temporal, internal, and predictive validation) were performed for the models, but for 6 models (NHS model, Chicago Model, Markov model/Vijan, Dutch diabetes model, Australian Diabetes Model, Discrete event simulation model from Jiao) only external validation was performed.

The results are presented in the several tables. Table 13 in Appendix provides an overview of the definition, description and use of external validation. Table 9 gives an overview of the data source selection. Table 14 in Appendix summarizes the results assessment in external validation. Table 15 to Table 18 in Appendix show the extracted data: these four tables display the used outcomes, risk factors and studies in external validation exercises. These results are grouped by models. Table 10 lists the criteria of the guidelines and the extent to which they were met in the external validation per model.

3.4.1 Definition

Table 13 in Appendix provides an overview of the definition, description and use of external validation. For 8 models a definition of external validation is given, for 10 models the process of external validation is described but no exact definition is given, for 3 models a mixture of definition and description is given and for 9 models no definition is given at all. An example of the definition of external validation is the validation of the Cardiff Model "External validation compares output from the model with data not specifically used to construct the disease progression algorithms"[95]. An example of the description of the process is given in the validation of the Markov model/Vijan "One way to test the validity of this model is to determine whether

it can accurately predict the rates of microvascular disease reported in actual patient populations" [87]. An example of a mixture of definition and description is the validation of the IMS CORE Diabetes Model "External validation compares clinical events predicted by the CDM with observed clinical outcomes using studies not directly used to inform disease progression within the model" [93].

3.4.2 Data sources

Table 9 gives an overview the data source selection. Only for the CDC/ RTI Model was a formal search for external validation studies conducted. For 3 models (UKPDS Outcomes Model, Archimedes Model, and PREDICT-DM), validation was performed for a specific population. In 4 models (IMS CORE Diabetes Model, Archimedes Model, PREDICT-DM, and CDOM), validation studies were chosen by baseline characteristics and/or outcomes. For 3 models (IHE/ ECHO-T2DM Model, CDC/ RTI Model, Michigan Model for Diabetes), data sources already used in other validation studies were used for external validation. IHE/ ECHO-T2DM Model, Cornerstone Diabetes Simulation and the BRAVO model used studies from MtHChs for external validation. Markov model/Vijan and Evidence-Based Medicine Integrator stated that recent studies were used for external validation. For the Cardiff Model and DiDACT model, it was reported that key trials/landmark trials were used for external validation. 18 models did not provide an explanation for the selection of their data source.

Table 15 and Table 16 in Appendix present the studies used for external validation process. In total 86 studies were used for external validation and the process was performed at least 172 times. Landmark studies are those that have been used by at least 4 models in external validation (UKPDS, ACCORD, ADVANCE, ASPEN, VADT, ADDITION, CARDS, Look AHEAD, EMPA-REG, CANVAS).

The most often used studies for external validation are EMPA-REG and CANVAS, which were both only utilized on the 9th MtHCh, and ADVANCE and CARDS each of them employed 11 times. 64 studies were only used once for external validation.

EMPA-REG is a randomized controlled trial on cardiovascular morbidity and mortality in patients with T2DM at high CVD risk performed from September 2010 to April 2013. 7,020 patients were treated with 10 mg or 25 mg empagliflozion (sodium-glucose cotransporter 2 inhibitor) or placebo once daily in addition to their standard treatment. Primary composite outcome was death from CVD causes, nonfatal myocardial infarction or nonfatal stroke. The risk for

CVD events was lower in the treatment group compared to placebo group (HR 0.86 P= 0.04) [49].

CANVAS is a similar randomized controlled trial on same primary composite outcome as EMPA-REG in patients with T2DM with high CVD risk. 10,020 patients from two trials (CANVAS and CANVAS-R) were treated with canagliflozin (sodium-glucose cotransporter 2 inhibitor). Patients from CANVAS received canagliflozin at a daily dose of 300mg, 100mg or placebo and patients from CANVAS-R a daily dose of initially 100mg with a possible increase to 300 mg or placebo. The risk for CVD events was lower in the treatment group than in control group (HR 0.86 P= 0.02) [136].

Model	Formal search	Specific population	Selection based on baseline characteristics and/or outcomes	Used in other validation studies	MtHChs	Recent studies	Key trial/ landmark trial	no explanation given
IMS CORE Diabetes Model			x					
Cardiff Model							x	
UKPDS Outcomes Model		x						
Archimedes Model		x	x					
IHE/ ECHO-T2DM Model				x	x			
CDC/ RTI Model	x			x				
Michigan Model for Diabetes				x				x
NHS model								x
Chicago Model								x
DiDACT model							x	
MICADO								x
Markov model/Vijan						x		
Dutch diabetes model								x
SPHR Diabetes								x
Microsimulation/Javanbakht								x
IHSD DPMM								x
Microsimulation:model form Caro JJ /Eastman								x
Australian Diabetes Model								x
Discrete event simulation model from Jiao								x
PREDICT-DM		x	x					
CDOM			x					
Cornerstone Diabetes Simulation					x			

Chinese Outcomes Model for T2DM								x
BRAVO Model					x			x
Sheffield Diabetes Model								x
EAGLE								x
Evidence-Based Medicine Integrator						x		
PROSIT Model								x
TTM								x
IQVIA CORE Diabetes Model								x
Sum	1	3	4	3	3	2	2	18

Table 9: Description of data source selection

3.4.3 External validation results

Table 14 in Appendix summarizes the results of external validation in the found literature. In Table 17 and Table 18 in Appendix risk factors and outcomes used for external validation were shown.

3.4.3.1 External validation methods implemented

The extent of external validation varies between the models. Outcomes were compared in 23 of the 30 models (77 %). Often, further statistical procedures were applied, such as regression analysis for 9 models (30 %) and confidence intervals also for 12 models (40 %). In addition to the regression analysis, the coefficient of determination (R^2) was determined for 7 models (23 %). The mean absolute percentage error (MAPE) was calculated for 8 models (27 %). Hazard ratios were given for 5 models (17 %), a sensitivity analysis for 7 models (33 %), and an F-test was carried out for 2 models (7 %). For 20 models (67 %), it is stated in the literature that overall accurate/adequate predictions were calculated by the models. For 8 models (27 %), it is reported that they over- or underestimated the outcomes. If parts of the results do not provide accurate/adequate predictions, this was explained in more detail by 17 of 30 models (57 %). Microsimulation/Javanbakht and Microsimulation:model form Caro JJ /Eastman did not give detailed results.

3.4.3.2 Risk factors

Table 17 in Appendix shows the used risk factors which are grouped socioeconomic risk factors, laboratory results with subgroups of lipids and other laboratory results like HbA1c, T2DM characteristics, cardiological parameters, weight, comorbidities, behavior, and other risk factors. In total at least 216 validation exercises were performed in 30 models. The most often used risk factor categories are laboratory results with 63, socioeconomic with 54 and cardiological with 25. Age was used in 23 models and is the most frequent used risk factor. It is followed by sex used in 21 models and HbA1c used in 20 models. 18 risk factors were only used in one model. The three models with the broadest range of risk factors used are Chicago Model with 18, and Cardiff Model, and Chinese Outcomes Model for T2DM with both 14 risk factors, respectively. No risk factors were used or available in Evidence-Based Medicine Integrator, PROSIT Model and TTM.

3.4.3.3 Outcomes

In Table 18 in Appendix, the outcomes are displayed grouped by general outcomes, macrovascular diseases, microvascular diseases, and other complications. Incidence respectively cumulative incidence, prevalence, mortality/ survival, life expectancy, and prevalence of complications belong to general outcomes. Macrovascular diseases are classified as cardiovascular disease which can be subclassified in coronary heart disease, peripheral artery disease and cerebrovascular disease. Microvascular diseases are divided in retinopathy, nephropathy and neuropathy. When external validation was performed mainly the incidence of diabetes, clinical outcomes/ incidence of complications and comorbidities and life expectancy/survival were validated. In 30 models external validation was performed at least 197 times. The most often validated outcome is mortality/ survival which was used in 20 models. Afterwards stroke was used in 19 models and myocardial infarction in 18 models. Other often used outcomes are cardiovascular disease and cardiovascular death both 14 times used, and congestive heart failure, and (lower extremity) amputation both 11 times used. Least often validated outcomes which were only used once are cumulative incidence, diabetes prevalence, prevalence of complications, transient ischemic attack (TIA), cataract, death from end-stage renal disease, annual per patient cost estimates for diabetes, disability-adjusted life years (DALYs), cancer, and progression of HbA1c. The model with the widest range of outcomes that were validated is IHE/ ECHO-T2DM Model with 15 validated outcomes. IMS CORE Diabetes Model and Cornerstone Diabetes Simulation are validated both with 13 outcomes, and Cardiff Model, and CDC/ RTI Model are both validated with 12 times.

To sum up, in paper-based models validation exercises are more often performed than in models only used on MtHChs.

3.4.3.4 Presenting results

The presentation of results varied among the models (Table 14 in Appendix). For 22 models (73 %), tables were used to present the results. Summary measures of goodness of fit were tabulated for 2 models (7 %). For 8 models (27%), there was no tabular presentation of results.

Scatterplots were mainly used for 11 models (37 %). For 4 models (13 %), the predictions were plotted over time. Kaplan Meier-curves, receiver operating curves, boxplots, histograms, and validity curves were occasionally given. For 13 models (43 %), no graphs or figures were used to illustrate the results.

3.4.4 Summary of guidelines criteria and further specifications per model

Table 10 summarizes the results in relation to the guidelines. The criteria of the guidelines are listed and the extent to which a model has fulfilled these criteria is assessed. The criteria are listed in the following and explained where needed. In general, the guidelines use terms such as 'frequent', 'many', and 'landmark trial' but do not define them precisely. Therefore, I made specifications based on the results, which can be considered arbitrary, but were necessary for the evaluation of the corresponding criteria.

First, the guidelines advise that frequent validation should be performed, but do not specify this. Therefore, I have considered frequent validation when there has been at least external validation in one paper and on one MtHCh. If only one of the conditions was fulfilled, the criterion was assessed as partially fulfilled. Second, a definition, description of the process or both should be given. Third, a formal search should be performed to find relevant and sufficient data sources to perform the external validation. Fourth, external validation should be performed against many landmark trials. However, again there is no quantification or definition of ‘many’ and ‘landmark trial’ in the guidelines. I have therefore defined a ‘landmark trial’ that have been used at least 4 times in external validation. Moreover, I assumed as a limit for ‘many’ that external validation should be performed against more than 3 landmark trials. If a model has been externally validated against only 1 to 3 landmark trials, the criterion was considered partially fulfilled. Fifth, the criterion 'frequent validation of standard outcomes' is not further classified in the guidelines. I have therefore defined the threshold to be more than 8 standard outcomes. If a model had 4 to 8 standard outcomes externally validated, the criterion was considered partially met. Sixth, the methods should be in accordance with the study used. Seventh, methods to assess sampling error, statistical significance, and/or sensitivity analysis should be applied. Eighth, the description of the external validation should be detailed. Ninth, authors of external validation studies should indicate whether their external validation provides accurate/adequate predictions and they should explain deviations.

Model	Validation in at least one paper and on one MtHCh	Definition, description of process or both	Formal search	Validation against more than 3 landmark trials	Validation of more than 8 standard outcomes	Methods in accordance with study used	Application of methods to assess sampling error, statistical significance, and/or sensitivity analysis	Description given in details	Accurate/ adequate predictions by author's suggestion and explanation of deviations
IMS CORE Diabetes Model	x	x	-	x	x	x	x	x	x
Cardiff Model	x	x	-	x	x	x	x	x	x
UKPDS Outcomes Model	x	x	-	x	x	x	x	x	x
Archimedes Model	x	x	-	x	(x)	x	x	x	x
IHE/ ECHO-T2DM Model	x	x	-	x	x	x	x	x	x
CDC/ RTI Model	x	x	x	x	x	x	x	x	x
Michigan Model for Diabetes	x	x	-	x	(x)	x	x	x	x
NHS model	(x)	-	-	-	-	x	x	(x)	x
Chicago Model	-	x	-	-	-	x	x	(x)	x
DiDACT model	-	-	-	-	-	-	-	-	-
MICADO	x	x	-	(x)	x	x	x	x	x
Markov model/Vijan	-	x	-	(x)	-	x	x	x	x
Dutch diabetes model	-	-	-	-	-	-	-	-	(x)
SPHR Diabetes	x	x	-	(x)	(x)	x	-	x	(x)
Microsimulation/ Javanbakht	-	-	-	-	-	-	-	-	-
IHSD DPMM	(x)	x	-	(x)	-	(x)	-	(x)	(x)
Microsimulation:model form Caro JJ /Eastman	-	-	-	-	-	-	-	-	-
Australian Diabetes Model	-	-	-	-	-	x	-	-	(x)

Discrete event simulation model from Jiao	-	-	-	-	-	x	x	x	(x)
PREDICT-DM	-	x	-	(x)	(x)	x	x	x	(x)
CDOM	-	x	-	-	(x)	x	x	x	x
Cornerstone Diabetes Simulation	-	x	-	(x)	x	x	x	x	x
Chinese Outcomes Model for T2DM	-	x	-	x	x	x	x	x	x
BRAVO Model	(x)	-	-	x	x	x	x	x	(x)
Sheffield Diabetes Model	-	x	-	(x)	(x)	x	x	x	-
EAGLE	-	x	-	(x)	(x)	x	x	x	-
Evidence-Based Medicine Integrator	-	-	-	(x)	(x)	x	-	-	-
PROSIT Model	-	-	-	(x)	(x)	x	x	-	-
TTM	-	-	-	(x)	(x)	x	x	-	-
IQVIA CORE Diabetes Model	-	-	-	(x)	(x)	x	x	-	-

x Criterion fulfilled, (x) Criterion partially fulfilled, - Criterion not fulfilled

Table 10: Criteria of the guidelines per models

4 Discussion

The main objective of this thesis was to conduct a SLR on external validation practices of T2DM and to describe and evaluate the process of external validation and to provide an overview and recommendations for performing external validation in the future.

First of all, in the following, an overview and a general discussion of validation criteria proposed in the guidelines is given.

Criterion 1. Definition. The basis for the process evaluation is the definition of external validation and whether modelling groups all understand it in the same way. This criterion should be easy to follow.

Criterion 2. Data sources. In all guidelines, this means simulation against external data that has not been used for the construction of the respective model [11–13]. This criterion should also be easy to follow.

Criterion 3. Frequency. Regarding the frequency of a validation, I have taken as a limit that external validation must be performed for a model in at least one paper and on one MtHCh. Furthermore, according to the guidelines external validation should be performed sufficiently regularly and against "many landmark trials". The background for this condition is that there should be no pre-selection of "appropriate" studies for the models [11–13]. The challenge with this condition given in the guidelines is that the frequency and the term "landmark trial" are no further defined in the guidelines. In reality, studies are often selected because of specific characteristics. For example, on the 9th MtHCh external validation was performed with the EMPA-REG and CANVAS studies, both of them showed a survival benefit with the use of SLGT-2 inhibitors for patients with T2DM [9, 49, 136]. Therefore, in this thesis, studies that have been used at least 4 times in external validation are considered landmark studies. Furthermore, I assumed as a limit for 'many' that external validation should be performed against more than 3 landmark trials. This can be criticized as arbitrary, however, I decided to do this in order to enable reproducibility of our results.

Criterion 4. Formal search of studies source. The next criterion is that a formal search for external validation should be undertaken. A formal search is certainly desirable to ensure the largest possible data base for external validation, but it does not guarantee that a model will be externally validated against sufficient studies. Other methods are therefore also necessary to achieve the widest possible range of studies [12, 13].

Criterion 5. Standard outcomes. Additionally, a frequent validation of standard outcomes should be performed. I have therefore defined the threshold to be more than 8 standard outcomes. These standard outcomes were exclusive. Again, this can be considered as arbitrary, but it is important for the reproducibility of the results.

Criterion 6 and 7. Population characteristics and statistical methods. There should be an agreement of population characteristics and statistical methods between study and model. This point is not controversial in the guidelines, as discrepancies would predictably lead to deviations in the results or results would not be comparable [11–13]. To increase the reliability of the results, they should be tested for statistical significance and the sampling error should be taken into account. This is not further defined in the guidelines, since the models are often very different and no concrete recommendation on specific methods may be given.

Criterion 8 and 9. Reporting results. The condition that the results of external validation should be described in sufficient detail, the extent to which accurate/adequate predictions have been made by author's suggestion and the extent to which deviations have been explained are uncontroversial.

A detailed discussion of each model is proposed in the following section.

4.1 Performance of the individual models

In the following, the results especially from Table 10 are used to assess the extent to which a model has performed external validation in accordance with the guidelines. The models are approximately arranged according to the frequency of their use in the literature. However, there is no fixed order.

4.1.1 IMS CORE Diabetes Model

The IMS CORE Diabetes Model has been externally validated sufficiently frequently in papers and on MtHChs. A definition or a description of the process was given. Studies for external validation were selected using baseline characteristics and/or outcomes rather than a formal search. However, it can be stated that IMS CORE Diabetes Model has been validated against many landmark studies (≥4). A lot of standard outcomes (>8) have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. External validation has been described in adequate detail, and the model produces accurate/ adequate predictions by author's suggestion.

Deviations were reasonably explained. Overall, external validation was performed in accordance with the guidelines.

4.1.2 Cardiff Model

The Model has been externally validated once in a paper and on the 4th and 9th MtHCh. A definition of the external validation was provided. For the external validation, mainly landmark studies were used and provide a robust data base for the validation process. Many standard outcomes (>8) have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. Limitations of the model were indicated. External validation has been described in adequate detail, and model produces accurate/ adequate predictions by author's suggestion. In conclusion, external validation was performed in accordance with the guidelines.

4.1.3 UKPDS Outcomes Model

The UKDPS Outcomes Model has been externally validated sufficiently frequently in papers and on MtHChs. A definition of external validation was presented. Studies were selected on the basis that a specific population matched the model characteristics. However, it can be stated that UKPDS Outcomes Model has been validated against many landmark studies. A lot of standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error, statistical significance, and sensitivity analysis. External validation has been described in adequate detail, and the model produces accurate/ adequate predictions by author's suggestion. Deviations, and limitations were sufficiently explained. In general, external validation was performed in accordance with the guidelines.

4.1.4 Archimedes Model

The Archimedes Model has been externally validated sufficiently frequently in papers and it has been externally validated on the 4th MtHCh. A definition of external validation was given. Studies were selected on the basis that a specific population matched the model characteristics. However, it can be stated that Archimedes Model has been validated against many landmark studies. Six standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and

statistical significance. Limitation of the model were indicated. External validation has been described in adequate detail, and the model produces accurate/ adequate predictions by author's suggestion. To conclude, external validation was performed in accordance with the guidelines for the most part.

4.1.5 IHE/ ECHO-T2DM Model

The IHE/ ECHO-T2DM has been externally validated sufficiently frequently in papers and participated on five MtHChs, but it has only been externally validated on the 9th MtHCh. A definition or a description of the process of external validation was provided. Studies were selected that were used in either other validation studies or on MtHChs. Therefore, it can be stated that IHE/ ECHO-T2DM Model has been validated against many landmark studies. Many standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. External validation has been described in adequate detail, and the model produces accurate/ adequate predictions by author's suggestion. Limitations were reasonably explained. To sum up, external validation was performed in accordance with the guidelines.

4.1.6 CDC/ RTI Model

The Model has been externally validated once in a paper and it has been externally validated sufficiently frequently on MtHChs. The external validation has been described in adequate detail. No definition but a description of the process was provided. Studies were selected using a formal search or because they had already been used in other studies. Therefore, it can be stated that CDC/ RTI Model has been validated against many landmark studies. A lot of standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. The model produces accurate/ adequate predictions by author's suggestion and deviations, and limitations are sufficiently explained. Overall, external validation was performed in accordance with the guidelines for the most part.

4.1.7 Michigan Model for Diabetes

The Michigan Model for Diabetes has been externally validated sufficiently frequently in papers and on MtHChs. The external validation has been described in adequate detail. No

definition but a description of the process was presented. Studies were selected which were already used in other validation studies or no explanation was given. However, it can be stated that Michigan Model for Diabetes has been validated against many landmark studies. Eight standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. The model produces accurate/ adequate predictions by author's suggestion and deviations, and limitations are appropriately explained. In summation, external validation was performed in accordance with the guidelines.

4.1.8 NHS model

The NHS model is based on the UKDS Outcomes Model with an adjustment for nephropathy. Therefore, it can be generally concluded that accurate or adequate results can be obtained with the simulation. Regarding the assessment of the external validation of the results of nephropathy, one paper serves as data base since the model did not attend at any MtHCh. No definition of external validation and no explanation, how the three studies were selected, were presented. Furthermore, the model was externally validated against only two standard outcomes which were both laboratory parameters referring to renal disease (microabluminuria and gross proteinuria). The methods used were consistent with the respective studies, and methods were applied to assess sampling error, statistical significance, and sensitivity analysis. The description of the external validation is only partially detailed. With respect to the UKDPS Outcomes Model, the model produces accurate/ adequate predictions by author's suggestion and deviations, and limitations are reasonably explained. In conclusion, external validation was only partially performed in accordance with the guidelines.

4.1.9 Chicago model

The Chicago Model has been externally validated only once in a paper and did not participate at any MtHCh. No definition but a description of the process of external validation was given. Furthermore, no explanation of the one selected study was provided. Hence, the model has not been validated against many landmark studies. Additionally, the model was externally validated against only two standard outcomes (mortality (death)/survival and life expectancy). The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. The model produces accurate/ adequate predictions by author's suggestion, but the results are limited to the used population and some details of the

description of the process are missing. To conclude, external validation was only partially performed in accordance with the guidelines.

4.1.10 DiDACT model

The external validation of the DiDACT model was described in 2 papers, but in these results of external validation were only mentioned, and no detailed results were presented. Further, although the responsible research group participated in the 4th MtHCh, they declined to conduct an external validation because none of the proposed endpoints would fit their model. Therefore, it can be concluded that external validation was not performed in accordance with the guidelines based on the available information.

4.1.11 MICADO

MICADO has been externally validated sufficiently frequently in papers and participated on three MtHChs, but it has only been externally validated on the 9th MtHCh. The external validation has been described in adequate detail. No definition but a description of the process of external validation was given. Moreover, no explanation of the selected studies was provided. Nevertheless, MICADO has been externally validated against three landmark studies. A lot of standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error, statistical significance, and sensitivity analysis. The model produces mostly accurate/ adequate predictions by author's suggestion and deviations, and limitations are reasonably explained. To sum up, external validation was performed in accordance with the guidelines for the most part.

4.1.12 Markov model/Vijan

Markov model/Vijan has been externally validated once in a paper and did not participate at any MtHCh. Despite this, the external validation has been described in adequate detail. No definition but a description of the process of external validation was given. The authors stated that they validated the model against recent studies. In summary, the model was externally validated against two landmark studies. Moreover, the model was externally validated against only two standard outcomes (retinopathy and progression of HbA1c). The methods used were consistent with the respective studies, and methods were applied to perform sensitivity analysis. The model produces accurate/ adequate predictions by author's suggestion and deviations, and

limitations are sufficiently explained. Overall, external validation was performed in accordance with the guidelines for the most part.

4.1.13 Dutch diabetes model

The Dutch diabetes model has been externally validated only once in a paper and did not participate at any MtHCh. The results are not described in sufficient detail. On the one hand, the selection of studies is not justified in detail. On the other hand, although results are summarized and limitations are pointed out, the presentation of external validation processes is missing. Therefore, it can be concluded that external validation was not performed in accordance with the guidelines based on the available information.

4.1.14 SPHR Diabetes

SPHR Diabetes has been externally validated once in a paper and participated in two MtHChs but was only externally validated on the 9th MtHCh. No definition but a description of the process of external validation was given. In addition, no explanation of the selected studies was provided. However, the model has been validated against three landmark studies. Furthermore, many standard outcomes have been externally validated. The methods used were consistent with the respective studies, but no methods were applied to assess sampling error, statistical significance, or sensitivity analysis. External validation has been described in adequate detail, but the model does not produce accurate/ adequate predictions by author's suggestion what is explained reasonably. Altogether, external validation was performed in accordance with the guidelines for the most part.

4.1.15 Microsimulation/Javanbakht

External validation of microsimulation/Javanbakht was mentioned in one paper and the responsible research group did not participate on any MtHCh. No definition of external validation or description of study selection was provided. Furthermore, methods were described but no results were presented. External validation was not described in sufficient detail. In summary, it can be concluded that external validation was not performed in accordance with the guidelines based on the available information.

4.1.16 IHSD DPMM

IHSD DPMM has been externally validated sufficiently frequently in papers but did not participate on any MtHCh. A definition of external validation was given, but no justification for study selection was provided. The model was only externally validated against two landmark studies. The description of the external validation varies considerably between the papers. In the one paper exact descriptions of the results were given and in the other paper there was no presentation of the results. Where sufficient results were presented, they showed accurate/adequate predictions by author's suggestion and deviations were explained in a reasonable way. In conclusion, external validation was only partially performed in accordance with the guidelines.

4.1.17 Microsimulation:model form Caro JJ /Eastman

The Microsimulation:model form Caro JJ/Eastman has not been sufficiently externally validated according to available data. The responsible research group has not participated in any MtHCh and no definitions or descriptions of external validation have been published. It was only mentioned that adequate results were achieved in the external validation. Therefore, it can be concluded that external validation was not performed in accordance with the guidelines based on the available information.

4.1.18 Australian Diabetes Model

In the Australian Diabetes Model, external validation was only addressed in one paper and the responsible research group did not participate at any MtHCh. No definition of external validation or description of study selection was given. Furthermore, the model was not externally validated against landmark studies. Only two not common outcomes were externally validated (annual per patient cost estimates for diabetes and DALYs). While the same methods were used as in the corresponding studies and accurate/adequate results were obtained by author's suggestion, the description was not sufficiently detailed. To sum up, it can be concluded that external validation was not performed in accordance with the guidelines based on the available information.

4.1.19 Discrete event simulation model from Jiao

External validation was addressed in one paper for the Discrete event simulation model from Jiao and the responsible research group did not participate in any MtHCh. No definition of

external validation or description of study selection was provided. Furthermore, the model was not externally validated against landmark studies. Additionally, only one standard outcome was externally validated (life expectancy). The same methods were used as in the corresponding studies and methods to appraise the statistical significance were applied. Accurate/adequate results were obtained by author's suggestion and the process was described in reasonable detail. However, limitations of the results were not sufficiently discussed. In conclusion, external validation was only partially performed in accordance with the guidelines.

4.1.20 PREDICT-DM

PREDICT-DM has been externally validated once in a paper and did not participate at any MtHCh. No definition but a description of the process was presented. Studies were selected on the basis that a specific population, the baseline characteristics and/or outcomes matched the model characteristics. Thus, the model has been validated against two landmark studies. Additionally, eight standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. External validation has been described in adequate detail, and the model produces accurate/ adequate predictions by author's suggestion. The authors gave no explanation regarding the limitation of external validation. Altogether, external validation was only partially performed in accordance with the guidelines.

4.1.21 CDOM

CDOM has been externally validated once in a paper and did not participate at any MtHCh. A definition of external validation was provided, and studies were selected on the basis that baseline characteristics and/or outcomes matched the model characteristics. To sum up, it can be stated that CDOM has not been validated against many landmark studies. Moreover, six standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess statistical significance of results. External validation has been described in sufficient detail, and the model produces accurate/ adequate predictions by author's suggestion. Limitations of the model were indicated. To conclude, external validation was only partially performed in accordance with the guidelines.

4.1.22 Cornerstone Diabetes Simulation

One paper has covered external validation of Cornerstone Diabetes Simulation, and the responsible research group has not participated on any MtHCh. For this model, a mixture of definition and description of the process of external validation was utilized. Studies were selected that have also been used on MtHChs. Therefore, the model was externally validated against three landmark studies. Furthermore, a lot of standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error and statistical significance. The external validation was described in sufficient detail and accurate/ adequate predictions were obtained by author's suggestion. Deviations, and limitations are appropriately explained. In summation, external validation was performed in accordance with the guidelines for the most part.

4.1.23 Chinese Outcomes Model for T2DM

Chinese Outcomes Model for T2DM has been externally validated once in a paper and did not participate at any MtHCh. A definition of external validation was provided, but no explanation of the selected studies was indicated. Overall, the model was externally validated against many landmark trials. In addition, many standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error, and statistical significance. External validation was described in reasonable detail. The model produces accurate/ adequate predictions by author's suggestion and deviations, and limitations are sufficiently explained. In conclusion, external validation was performed in accordance with the guidelines for the most part.

4.1.24 BRAVO Model

BRAVO Model has been externally validated once in a paper and on the 9th MtHCh. A definition of external validation or a description of the process were not given. Either studies that have been used on other MtHChs were chosen for external validation, or there was no reason for the study selection. Nevertheless, the model has been validated against many landmark studies. Furthermore, many standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to assess sampling error, and statistical significance. External validation was described in reasonable detail. While accurate/adequate results are obtained by author's suggestion, a discussion of limitations of the

results is missing. In summary, external validation was performed in accordance with the guidelines for the most part.

4.1.25 Sheffield Diabetes Model

The Sheffield Diabetes Model has participated in two MtHChs and no paper for external validation has been published. On the 4th MtHCh, it was only externally validated. A definition of external validation was provided on the MtHCh, but there was no justification for the study selection. The model was therefore only externally validated against one landmark study. Furthermore, only three standard outcomes have been externally validated (CVD, MI, and stroke). The methods used were consistent with the respective study, and methods were applied to perform sensitivity analysis. The results were described in sufficient detail, but there is no discussion of the individual models, so it is not clear why the Sheffield Diabetes Model overestimated macrovascular events. Overall, external validation was only partially performed in accordance with the guidelines.

4.1.26 EAGLE

Like the Sheffield Diabetes Model, EAGLE has participated in two MtHChs and no paper for external validation has been published. On the 4th MtHCh, it was only externally validated. A definition of external validation was provided on the MtHCh, but there was no justification for the study selection. The model was therefore only externally validated against one landmark study. In addition, only three standard outcomes have been externally validated (CVD, MI, and stroke). The methods used were consistent with the respective study, and methods were applied to perform sensitivity analysis. The results were described in sufficient detail, but there is no discussion of the individual models, so it is not clear why EAGLE does not yield to accurate/ adequate results by author's suggestion. Therefore, external validation was only partially performed in accordance with the guidelines.

4.1.27 Evidence-Based Medicine Integrator

Evidence-Based Medicine Integrator has only participated on 5th MtHCh where it was externally validated. No paper for external validation has been published. Further, no definition but a description of the process was presented. Recent studies were selected for validation. Thus,

the model was externally validated against only two landmark studies. The methods used were consistent with the respective study. Additionally, eight standard outcomes have been externally validated. No methods were applied to appraise sampling error or statistical significance. Due to the only external validation on a MtHCh, the description of results lacks detail. Furthermore, the model does not provide accurate/adequate results by author's suggestion, which means that it cannot be considered generally applicable based on the performed external validation. Hence, external validation was only partially performed in accordance with the guidelines.

4.1.28 PROSIT Model

PROSIT Model has participated on 9th MtHCh where it was externally validated. No paper for external validation has been published. Moreover, no definition of external validation and no description of the study selection were provided. Consequently, the model was externally validated against only two landmark studies. Six standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to appraise sampling error, and statistical significance. Furthermore, the model simulations do not lead to accurate/adequate results by author's suggestion. Due to the only external validation on a MtHCh, this is not reasonably explained in detail. In conclusion, external validation was only partially performed in accordance with the guidelines.

4.1.29 TTM

Like PROSIT Model, TTM has only participated on 9th MtHCh where it was externally validated. No paper for external validation has been published. Moreover, no definition of external validation and no description of the study selection were provided. As a result, the model was externally validated against only two landmark studies. Seven standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to appraise sampling error, and statistical significance. Furthermore, the model simulations do not lead to accurate/adequate results by author's suggestion. Due to the only external validation on a MtHCh, this is not reasonably explained in detail. To sum up, external validation was only partially performed in accordance with the guidelines.

4.1.30 IQVIA CORE Diabetes Model

The IQVIA CORE Diabetes Model is a further development of the IMS CORE Diabetes Model [137]. In general, it can be considered as well externally validated as described above. The IQVIA CORE Diabetes Model has therefore only been externally validated at the 9th MtHCh. Although no definition of the external validation and no description of the study selection were given, the reference to the IMS Core Diabetes Model means that the model has been sufficiently externally validated against landmark studies. In addition, two landmark studies were added through participation on the 9th MtHCh. Overall, many standard outcomes have been externally validated. The methods used were consistent with the respective studies, and methods were applied to appraise sampling error, and statistical significance. The IQVIA CORE Diabetes Model simulations do not lead to accurate/adequate results by author's suggestion. Due to the participation on a MtHCh, this is not sufficiently explained in detail. Altogether, external validation was only partially performed in accordance with the guidelines.

4.2 Discussion on overall level

4.2.1 Discussion of search in data bases and screening

To generate a broad data base for answering the research questions, 15 databases were searched and additional relevant sources, which mainly consisted of reviews, were considered. After reduplication, 41% of the papers were removed. The higher percentage than usual (around 3-29%) [71, 73] of duplicates was due to the collection of hits from 15 different databases that are partly overlapping with each other and a broad search strategy to obtain the most complete database as possible. In total, there were 13,021 hits available for abstract screening, which is a large number and therefore prone to error.

Furthermore, during the abstract screening the exclusion criteria were not strictly applied. For example, conference abstracts and models simulating the disease process in T1DM were included in the abstract screening in order to find relevant papers by manual search. Against the background of the intended broad data basis, I have accepted to have false positive hits in our data for the full text screening. After the two screening steps 0,2% of all hits were included in the analysis which is less than in comparable studies (0,6-4,8%) [65, 70, 71, 73] because of the reasons already mentioned in this section.

4.2.2 Discussion of the relevance of external validation

The range of how external validation has been applied across models varies considerably. As ADA and ISPOR guidelines have already pointed out, one reason for this is certainly that not one model fits every population in the world and therefore adaptations to specific research questions and populations are always necessary. Therefore, the question arises to what extent external validation respectively the general applicability of a model is essential at all [11, 12].

The results of this review suggest that of 79 models available for full text screening, only 24 (30%) were finally included in the analysis of external validation. Although external validation as a quality criterion is desirable in itself, this aspect is sometimes evaluated slightly differently in the literature [11, 12]. For the understanding of the results discussed in the following, it should be kept in mind that the extent and accuracy of the description of the external validation depends on the question of generalizability. MtHChs have an important role in this context. These are diabetes modelling conferences where current topics of the research area are discussed. The advantage of these events is that models can raise their visibility in the research field and that a transparent process takes place with regard to external validation. Here, it can be observed that of the 30 models included in the analysis 13 (43%) did not participate in any Mt Hood Challenge.

4.2.3 Discussion of external validation results

If a model performs external validation and publishes results, it should be assessed to what extent these are consistent with the known guidelines in the research field (definition of external validation, the selection of studies for performing external validation, the conducting of the external validation with the methods used and the presentation of the results). For the most part, detailed definitions, or descriptions of the process of external validation were reported. However, 9 out of 30 models did not provide any definition of external validation (30%). A similar issue can be noticed in the selection of studies. Although a formal search took place only for CDC/ RTI model, especially the well-established models (IMS CORE Diabetes Model, Cardiff Model, UKPDS Outcomes Model, Archimedes Model, IHE/ ECHO-T2DM Model, CDC/ RTI Model, Michigan Model for Diabetes) were externally validated against many landmark studies either via paper or on MtHChs. This shows that no selection of ‘suitable’ studies has taken place in these models, which has been addressed as a possible problem in the guidelines [11–13]. Again, however, it is also evident that for 18 of 30 models (60%) no rationale for study selection was supplied at all. This, of course, makes the evaluation of the methodology difficult. The correspondence of the methods between simulation and corresponding studies was largely

achieved, but the extent was very variable. In some models, a very detailed attempt was made to reproduce the relevant study population with risk factors, and in others, either because of the study specification or because of the model, only a few risk factors were specified for the simulation, making it difficult to understand the results. The same applies to the number of standard outcomes which were externally validated. Only for 9 models (30%) were many standard outcomes used for external validation. It should be noted, as well, that in some cases only a part of a model and therefore only some standard outcomes have been externally validated, which is also in accordance with the guidelines. As described in the guidelines and on MtHChs, the reasons for missing or incomplete study selection and description of methods can be the mismatching characteristics of the study populations, the different study protocols, or the definitions of the outcomes of the studies to be simulated [11, 12, 75]. An essential point for the assessment of the external validation and the accordance with the guidelines is the description of the results and the methods employed. In line with the guidelines, the validation process should always be described as precisely as possible. While it is not possible to describe every detail in a publication, these should be available either online or on request [11–13]. Overall, the results were mostly described in detail, methods for assessing sampling error and statistical significance were applied, and limitations were adequately described. However, especially in the case of models that have only participated on MtHChs or have only published one paper for external validation, it is noticeable that the results are not particularly detailed in some cases or that a discussion of the limitations is missing in others. Hence, it is sometimes difficult to classify the results. There is no clear description or recommendation in the guidelines and on MtHCh of the methods to be applied [9–13, 75]. The only clear recommendation given is that the same methods should be used as in the corresponding study to be externally validated. This is to ensure comparability of results [12]. The reasons discussed in the literature are that models are so different that there is no clear recommendation for performing external validation. Furthermore, varying outcome definitions can hinder a standardized recommendation [9].

However, frequently used methods appear through this SLR. For example, the results (observed vs. simulated) are either compared in tabular form or scatterplots. As indicated in the guidelines, CIs are frequently reported for point estimates and when comparing observed vs. simulated outcomes, regression analysis with the coefficient (R^2) of determination is calculated to assess agreement of observations with the simulated results [11, 13].

4.3 Strengths and limitations

Strengths of the SLR are that it presents a comprehensive overview of external validation in the literature and the relevance of external validation for the simulation of T2DM. These include the following aspects in particular. First, six reviews were identified in the literature so far that deal with health economic aspects of diabetes modeling and model assessments in diabetes field [66]. Five of these name and/or give a brief overview of the results of external validation, but our study is the first to give a comprehensive overview of the external validation of T2DM models [65, 70–73]. Second, for a large part of the models, it does not seem to be the goal to be externally validated and to achieve the quality criterion of external validation. Third, the methodology is easily comprehensible and reproducible due to the use of the PRIMSA guidelines.

The limitations of the work are, firstly, that there is always the possibility that publications or models have not been included in the search [65]. This risk was attempted to be minimized by searching 15 literature databases, and not strictly applying the exclusion criteria in abstract screening to keep the data base broad. Secondly, it must be emphasized that this SLR only provides an overview of external validation in the literature and it is not a study that assesses whether statistical methods are appropriate to perform external validation or to assess its statistical significance. Referring to this, the SLR can serve as recommendation for future statistical practices that should reported at least in technical reports. Thirdly, there is no assessment of the quality of studies used for external validation. This was not the subject of this work, as the focus was to present an overview of the current practice of external validation and on the accordance with the guidelines. The assessment of the quality of studies is an interesting research question for the future. Fourth, the results are only up to date to the time of data collection. Fifthly, regarding the relevance of external validation, it has to be mentioned that studies used for external validation are not based on real treatment setting and including only respective inclusion criteria. Therefore, the question of the general applicability and transferability of external validation and its results to real treatment setting arises [65]. Sixthly, the results were presented qualitatively by us and there can always be misjudgments. I tried to minimize this risk by working in teams of at least two. Seventhly, for the categories of sufficiently frequent validation, landmark trials and standard outcomes, I had to set limits to enable a reproducibility of our results that can be considered arbitrary.

4.4 Recommendations respectively best practice examples for performing external validation in the future

As already mentioned, external validation serves as generalizability of a model. Not every model needs to fit every population, and this is often not the goal of individual research groups. However, external validation is a quality criterion. As a recommendation for the future practice of external validation, a research group should consider in advance whether the model should be generalizable and then communicate this clearly. If a research group decides to conduct external validation, the guidelines could serve as the basis for formal implementation. The goal of this review was to present the current state of science on the topic of external validation in T2DM simulation models. Furthermore, recommendations for the future practice of external validation should be provided.

Based on the research results, recommendations respectively best practice examples for performing external validation in the future are presented below.

In terms of quantity of external validation, the UKPDS Outcomes Model represents a good example as external validation has been reported in three papers and it has participated in all MtHChs since the 3rd and thus external validation has been additionally performed on the 4th, 5th, and 9th MtHCh.

Regarding the specification of a definition of external validation, there are some good examples in the literature. One example is the definition used in the Cardiff Model: "External validation compares output from the model with data not specifically used to construct the disease progression algorithms" [95]. Based on the research results, it is crucial that a definition of external validation is given at all.

For the generalizability of a model, it is important that it has been externally validated against as many landmark studies as possible. A more precise description of 'many' and 'landmark trials' by the guidelines would be desirable for the future. Nevertheless, there are of course different possibilities to find landmark trials for example to use studies that are known in the research field or studies that have been used in other external validations. A further essential element, as suggested in the guidelines, is however a formal search. In addition to selecting studies already used in other models, a formal search in PubMed was performed for the CDC/ RTI Model [109]. This model is the only one for which a formal search was performed. More emphasis should certainly be placed on this topic in the future.

In order to obtain accurate results from the simulation, it is also important to represent the study population as accurately as possible. The Cardiff Model serves as a good example in this context, as it incorporates various risk factors as well as different treatment options. Furthermore, consistency between the model and the study was considered: "For each validation exercise the model's demographics, baseline risk factor and prior event history cohort profile was initialized to each validation study's cohort profile. Clinical events, consistently defined between the publication and those predicted by the Cardiff Model were compared over the relevant time horizon. Where appropriate, each simulated cohort had treatment effect profile applied to consistent to that reported in each respective study" [95]. This accuracy is further reflected in the results of the review, as 14 risk factors and 12 outcomes were used for external validation. Again, it would be desirable in the future to know how many standard outcomes (e.g. macro- and microvascular events) are considered sufficient for external validation.

Regarding the methods used and their detailed description, the IMS CORE Diabetes Model is a good example. In two papers 10 years apart (2004 and 2014) external validation was performed with the most commonly used methods from the literature and described in detail. The modelers compared the cumulative incidence of the different treatment options in tables and then performed a regression analysis where observed vs. simulated outcomes were presented in scatterplots. The coefficient of determination (R^2) was calculated and CIs for point estimators were given [92, 93].

Finally, according to the guidelines, external validation should also include a discussion of limitations. Again, the IMS CORE Diabetes Model represents a good example for this topic. The following passages are good examples of this:

"A criticism of validation analyses such as this is that they only demonstrate that the model accurately predicts the outcomes in a clinical trial setting, and not in the real-life situation including factors such as non-compliance or pregnancy." [92]

"The largest discrepancy between published trial data and the CORE Diabetes Model simulation was in the third-order [external] validation of nephropathy in type 2 diabetes patients. […] The discrepancy in this third-order validation may be due to a risk factor in the population that was not measured or reported for the WESDR cohort as reported by Eastman *et al.*[14], or perhaps changes in treatment protocols, adherence or compliance that have not been described." [92]

Vemer et al. (2016) published a checklist "Assessment of the Validation Status of Health-Economic decision models (AdViSHE)" for validation in general. This can help to systematically present results of validation and to provide research groups an overview of their status of

validation [83]. As this checklist was published in 2016, only IHE/ ECHO-T2DM Model and PREDICT-DM have used it for their validation. For the future, it remains to be seen to what extent this tool will be adopted and contribute to improving the validation process.

4.5 Conclusion

The main objective of this thesis was to conduct a SLR to describe and appraise external validation approaches employed in the simulation models of diabetes. Papers were gathered by a formal systematic search in different literature sources and screened to their relevance. In total 21,951 hits were found through the search and 30 models finally included in the data extraction and analysis. This SLR gives an overview of the current practices of external validation in the literature: the majority of modelers did not assess the external validity of their diabetes simulation models at all, or did not provide sufficient details required in the above-mentioned guidelines. At the same time, “state of art” examples of external validation performed in accordance with the guidelines can be found in the literature. For the future, I can conclude the following from this work. It would be desirable for authors of models to clearly state whether they perform external validation and then describe the process as precisely as possible. If a research group of a model wants its model to be generalizable and thus externally validated, it might be desirable to participate on MtHChs. Referring to future guidelines, I often had to set limits to make the results reproducible. It would be desirable if the guidelines quantified the terms 'many' or 'frequent' and if there was a definition of landmark trial. I also only assessed the standard outcomes in terms of quantity and not whether they are qualitatively similar. Furthermore, there was no evaluation of the adequacy of the statistical methods. For the future, a more precise classification of the guidelines would be desirable.

5 References

1. International Diabetes Federation. IDF Diabetes Atlas. 9th ed. Brussels; 2019.

2. Yang W, Dall TM, Beronjia K, Lin J, Semilla AP, Chakrabarti R, et al. Economic costs of diabetes in the U.S. in 2017. Diabetes Care. 2018;41:917–28.

3. Peters ML, Huisman EL, Schoonen M, Wolffenbuttel BHR. The current total economic burden of diabetes mellitus in the Netherlands. Neth J Med. 2017;75:281–97.

4. Stollenwerk B, Becker C, Leidl R. Gesundheitsökonomische Modellierung bei Diabetes mellitus : Möglichkeiten und Grenzen. Diabetologe. 2011;7:99–104.

5. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4. ed. Oxford: Oxford Univ. Press; 2015.

6. Mt Hood Challenge Website. https://www.mthooddiabeteschallenge.com/. Accessed 16 Jan 2020.

7. Cohen N, Minshall ME, Sharon-Nash L, Zakrzewska K, Valentine WJ, Palmer AJ. Continuous Subcutaneous Insulin Infusion versus Multiple Daily Injections of Insulin. 2007;25:881–97. http://search.ebscohost.com/login.aspx?direct=true&db=eoh&AN=EP27223579&site=ehost-live.

8. Palmer Andrew J M, Si Lei P, Tew Michelle MpMPH, Hua Xinyang MS, Willis Michael S P, Asseburg Christian PPMP, et al. Computer Modeling of Diabetes and Its Transparency: A Report on the Eighth Mount Hood Challenge. Value Heal. 2018;21:724–31. doi:10.1016/j.jval.2018.02.002.

9. Si L, Willis MS, Asseburg C, Nilsson A, Tew M, Clarke PM, et al. Evaluating the Ability of Economic Models of Diabetes to Simulate New Cardiovascular Outcomes Trials: A Report on the Ninth Mount Hood Diabetes Challenge. Value Heal. 2020;23:1163–70. doi:10.1016/j.jval.2020.04.1832.

10. Palmer AJ, Roze S, Valentine WJ, McEwan P, Gillett M, Holmes M, et al. Computer Modeling of Diabetes and Its Complications: A report on the Fourth Mount Hood Challenge Meeting. Diabetes Care. 2007;30:1638–46. doi:10.2337/dc07-9919.

11. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. Diabetes Care. 2004;27:2262–5. http://www.ncbi.nlm.nih.gov/pubmed/15333499. Accessed 4 Sep 2017.

12. 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. Med Decis Mak. 2012;32:733–43.

13. Caro JJ, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, et al. Questionnaire to Assess Relevance and Credibility of Modeling Studies for Informing Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. Value Heal. 2014;17:174–82. doi:10.1016/j.jval.2014.01.003.

14. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: A high-risk state for diabetes development. Lancet. 2012;379:2279–90. doi:10.1016/S0140-6736(12)60283-9.

15. Bansal N. Prediabetes diagnosis and treatment: A review. World J Diabetes. 2015;6:296.

16. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. Diabetes Care. 2020;43 Supplement 1:14–31. doi:10.2337/dc20-S002.

17. World Health Organization, IDF. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Geneva; 2006. https://www.who.int/publications-detail/definition-and-diagnosis-of-diabetes-mellitus-and-intermediate-hyperglycaemia.

18. World Health Organization. Global Health Risks. Mortality and burden of disease attributable to selected major risks. Geneva; 2009. https://apps.who.int/iris/bitstream/handle/10665/44203/9789241563871_eng.pdf?sequence=1&isAllowed=y.

19. Magliano DJ, Islam RM, Barr ELM, Gregg EW, Pavkov ME, Harding JL, et al. Trends in incidence of total or type 2 diabetes: Systematic review. BMJ. 2019;366:1–12.

20. Basu S, Yoffe P, Hills N, Lustig RH. The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data. PLoS One. 2013;8.

21. World Health Organization. Definition of cardiovascular diseases. 2023. https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

22. World Health Organization. About diabetes. Complications of diabetes. 2019. https://www.who.int/diabetes/action_online/basics/en/index3.html.

23. Beckman JA, Creager MA. Vascular complications of diabetes. Circ Res. 2016;118:1771–85.

24. Orasanu G, Plutzky J. The Pathologic Continuum of Diabetic Vascular Disease. J Am Coll Cardiol. 2009;53:35–42. doi:10.1016/j.jacc.2008.09.055.

25. Fowler MJ. Microvascular and Macrovascular Complications of Diabetes. Clin Diabetes. 2008;26:77–82. doi:10.2337/diaclin.26.2.77.

26. Wiernsperger NF. In defense of microvascular constriction in diabetes. Clin Hemorheol Microcirc. 2001;25:55–62. http://www.ncbi.nlm.nih.gov/pubmed/11790871.

27. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: A systematic literature review of scientific evidence from across the world in 2007-2017. Cardiovasc Diabetol. 2018;17:1–19. doi:10.1186/s12933-018-0728-6.

28. World Health Organization. ICD-10 Version:2016. 2016. https://icd.who.int/browse10/2016/en.

29. Wolk MJ, Scheidt S, Killip T. Heart Failure Complicating Acute Myocardial Infarction. Circulation. 1972;45:1125–38. doi:10.1161/01.CIR.45.5.1125.

30. Leontidis G, Al-Diri B, Hunter A. A new unified framework for the early detection of the progression to diabetic retinopathy from fundus images. Comput Biol Med. 2017;90:98–115. doi:http://dx.doi.org/10.1016/j.compbiomed.2017.09.008.

31. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375:2215–22. doi:10.1016/S0140-6736(10)60484-9.

32. Huang Y, Cai X, Mai W, Li M, Hu Y. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. BMJ. 2016;355:i5953. doi:10.1136/bmj.i5953.

33. Yeboah J, Bertoni AG, Herrington DM, Post WS, Burke GL. Impaired fasting glucose and the risk of incident diabetes mellitus and cardiovascular events in an adult population: MESA (Multi-Ethnic Study of Atherosclerosis). J Am Coll Cardiol. 2011;58:140–6. doi:10.1016/j.jacc.2011.03.025.

34. The Emerging Risk Facotrs Collaboration. Diabetes Mellitus, Fasting Glucose, and Risk of Cause-Specific Death. N Engl J Med. 2011;364:829–41. doi:10.1056/NEJMoa1008862.

35. Yau JWY, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556–64.

36. Sabanayagam C, Banu R, Chee ML, Lee R, Wang YX, Tan G, et al. Incidence and progression of diabetic retinopathy: a systematic review. Lancet Diabetes Endocrinol. 2019;7:140–9. doi:10.1016/S2213-8587(18)30128-1.

37. World Health Organization. Universal eye health. A global action plan 2014-2019. Geneva; 2013.

38. Klein R, Klein BEK. Vision Disorders in Diabetes. In: Diabetes in America. 2nd ed. National Institutes of Health and National Institute of Diabetes and Digestive and Kidney Diseases; 1995. p. 293–338.

39. Fakhruddin S, Alanazi W, Jackson KE. Diabetes-Induced Reactive Oxygen Species: Mechanism of Their Generation and Role in Renal Injury. J Diabetes Res. 2017;2017:1–30. doi:10.1155/2017/8379327.

40. Thomas MC, Cooper ME, Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. Nat Rev Nephrol. 2016;12:73–81. doi:10.1038/nrneph.2015.173.

41. Narres M, Claessen H, Droste S, Kvitkina T, Koch M, Kuss O, et al. The incidence of end-stage renal disease in the diabetic (compared to the non-diabetic) population: A systematic review. PLoS One. 2016;11:1–28. doi:10.1371/journal.pone.0147329.

42. Narres M, Kvitkina T, Claessen H, Droste S, Schuster B, Morbach S, et al. Incidence of lower extremity amputations in the diabetic compared with the non-diabetic population: A systematic review. PLoS One. 2017;12.

43. Mulnier HE, Seaman HE, Raleigh VS, Soedamah-Muthu SS, Colhoun HM, Lawrenson RA. Mortality in people with Type 2 diabetes in the UK. Diabet Med. 2006;23:516–21. doi:10.1111/j.1464-5491.2006.01838.x.

44. American Diabetes Association. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2019. Diabetes Care. 2019;42 Supplement 1:90–102. doi:10.2337/dc19-S009.

45. American Diabetes Association. 3. Prevention or Delay of Type 2 Diabetes: Standards of Medical Care in Diabetes—2019. Diabetes Care. 2019;42 Supplement 1:29–33. doi:10.2337/dc19-S003.

46. American Diabetes Association. 5. Lifestyle Management: Standards of Medical Care in Diabetes—2019. Diabetes Care. 2019;42 Supplement 1:46–60. doi:10.2337/dc19-S005.

47. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes Self-Management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. J Acad Nutr Diet. 2015;115:1323–34. doi:10.1016/j.jand.2015.05.012.

48. Galaviz KI, Weber MB, Straus A, Haw JS, Venkat Narayan KM, Ali MK. Global diabetes prevention interventions: A systematic review and network meta-analysis of the real-world impact on incidence, weight, and glucose. Diabetes Care. 2018;41:1526–34.

49. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med. 2015;373:2117–28.

50. Breton MC, Guénette L, Amiche MA, Kayibanda JF, Grégoire JP, Moisan J. Burden of diabetes on the ability to work. A systematic review. Diabetes Care. 2013;36:740–9.

51. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Bärnighausen T, et al. The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study. Lancet Diabetes Endocrinol. 2017;5:423–30. doi:10.1016/S2213-8587(17)30097-9.

52. Committee for a Responsible Federal Budget. American Health Care: Health Spending and the Federal Budget. 2018. https://www.crfb.org/papers/american-health-care-health-spending-and-federal-budget. Accessed 21 Jul 2020.

53. Zhuo X, Zhang P, Hoerger TJ. Lifetime direct medical costs of treating type 2 diabetes and diabetic complications. Am J Prev Med. 2013;45:253–61. doi:10.1016/j.amepre.2013.04.017.

54. Kähm K, Laxy M, Schneider U, Rogowski WH, Lhachimi SK, Holle R. Health Care Costs Associated With Incident Complications in Patients With Type 2 Diabetes in Germany. Diabetes Care. 2018;41:971–8.

55. Rudmik L, Drummond M. Health economic evaluation: Important principles and methodology. Laryngoscope. 2013;123:1341–7.

56. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15:677–87.

57. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006;15:1295–310. doi:10.1002/hec.1148.

58. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: Selecting the appropriate approach. J Heal Serv Res Policy. 2004;9:110–8.

59. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). Lancet. 2002;360:711–5.

60. Diabetes Prevention Program Research Group. Within-Trial Cost-Effectiveness of Lifestyle Intervention or Metformin for the Primary Prevention of Type 2 Diabetes. Diabetes Care. 2003;26:2518–23. doi:10.2337/diacare.26.9.2518.

61. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002;346:393–403.

62. Herman WH, Hoerger TJ, Brandle M, Hicks K, Sorensen S, Zhang P, et al. The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. Ann Intern Med. 2005;142:323–32.

63. Valentine WJ, Palmer AJ, Erny-Albrecht KM, Ray JA, Cobden D, Foos V, et al. Cost-effectiveness of basal insulin from a US health system perspective: comparative analyses of detemir, glargine, and NPH. Adv Ther. 2006;23:191–207. doi:10.1007/bf02850126.

64. Lifetime Benefits and Costs of Intensive Therapy as Practiced in the Diabetes Control and Complications Trial. JAMA J Am Med Assoc. 1996;276:1409–15. doi:10.1001/jama.1996.03540170053032.

65. Tarride J-EE-E, Hopkins R, Blackhouse G, Bowen JM, Bischof M, Von Keyserlingk C, et al. A review of methods used in long-term cost-effectiveness models of diabetes mellitus treatment. Pharmacoeconomics. 2010;28:255–77. doi:10.2165/11531590-000000000-00000.

66. Ogurtsova K, Heise TL, Linnenkamp U, Dintsios C-M, Lhachimi SKSK, Icks A. 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;6:1–10.

67. Eastman RC, Javitt JC, Herman WH, Dasbach EJ, Copley-Merriman C, Maier W, et al. Model of Complications of NIDDM: II. Analysis of the health benefits and cost-effectiveness of treating NIDDM with the goal of normoglycemia. Diabetes Care. 1997;20:735–44. doi:10.2337/diacare.20.5.735.

68. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Med Decis Making. 2012;32:667–77. doi:10.1177/0272989X12454577.

69. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation.

Pharmacoeconomics. 1998;13:397–409.

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

71. Charokopou M, Sabater FJJ, Townsend R, Roudaut M, McEwan P, Verheggen BGG, 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;32:207–18. doi:10.1185/03007995.2015.1102722.

72. Becker C, Langer A, Leidl R. The quality of three decision-analytic diabetes models: A systematic health economic assessment. Expert Rev Pharmacoeconomics Outcomes Res. 2011;11:751–62. doi:10.1586/erp.11.68.

73. Henriksson M, Jindal R, Sternhufvud C, Sorstadius E, Bergenheim K, Willis M. A systematic review of cost-effectiveness models in type 1 diabetes mellitus. Value Heal. 2015;18 (7):A607. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed17&AN=72084354.

74. Kirsch F. A Systematic Review Of Markov Models Evaluating Multi-Component Disease Management Programs. Value Heal. 2015;18:A502. doi:10.1016/j.jval.2015.09.1422.

75. Palmer AJ, Group, BmediSCI, MBBS TMH 5 M. Computer Modeling of Diabetes and Its Complications: A Report on the Fifth Mount Hood Challenge Meeting. Value Heal. 2013;16:670–85. doi:10.1016/j.jval.2013.01.002.

76. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices - Modeling Studies. Value Heal. 2003;6:9–17.

77. McHaney R. Computer Simulation. A Practical Perspective. San Diego: Academic Press; 1991.

78. Altman DG, Royston P. What do we mean by validating a prognostic model ? Stat Med. 2000;19:453–73.

79. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009;6:e1000097. doi:10.1371/journal.pmed.1000097.

80. Free reference manager. https://www.mendeley.com/?interaction_required=true. Accessed 10 Jan 2020.

81. Abstract Screening Service. https://rayyan.qcri.org/welcome. Accessed 10 Jan 2020.

82. Cloud manager. https://www.google.com/sheets/about/. Accessed 10 Jan 2020.

83. Vemer P, Corro Ramos I, van Voorn GAK, Al MJ, Feenstra TL. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. Pharmacoeconomics. 2016;34:349–61.

84. Ogurtsova K, Heise TL, Linnenkamp U, Dintsios C-M, Lhachimi SKSK, Icks A. 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;267:1–10.

85. Farmer AJ, Stevens R, Hirst J, Lung T, Oke J, Clarke P, et al. Optimal strategies for identifying kidney disease in diabetes: Properties of screening tests, progression of renal dysfunction and impact of treatment - Systematic review and modelling of progression and cost-effectiveness. Health Technol Assess (Rockv). 2014;18:1–127. doi:10.3310/hta18140.

86. O'Hanlon CE, Cooper JM, Lee SM, John P, Churpek M, Chin MH, et al. Life Expectancy Predictions for Older Diabetic Patients as Estimated by Physicians and a Prognostic Model. MDM Policy Pract. 2017;2:1–9. doi:10.1177/2381468317713718.

87. Vijan S, Hofer TP, Hayward RA. Estimated benefits of glycemic control in microvascular complications in type 2 diabetes. Ann Intern Med. 1997;127:788–95. http://www.ncbi.nlm.nih.gov/pubmed/9382399.

88. Niessen LW, Dijkstra R, Hutubessy R, Rutten GE, Casparie AF. Lifetime health effects and costs of diabetes treatment. Neth J Med. 2003;61:355–64. http://www.ncbi.nlm.nih.gov/pubmed/14768718%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed6&NEWS=N&AN=14768718.

89. Walker A, Colagiuri S, Mclennan M. Cost – Benefit Model of Diabetes Prevention and Care. Model Construction, Assumptions and Validation. 2002.

90. Jiao F, Wan EYF, Fung CSC, Chan AKC, McGhee SM, Kwok RLP, et al. Cost-effectiveness of a primary care multidisciplinary Risk Assessment and Management Program for patients with diabetes mellitus (RAMP-DM) over lifetime. Endocrine. 2019;63:259–69. doi:10.1007/s12020-018-1727-9.

91. Palmer AJ, Roze S, Valentine WJ, Minshall ME, Foos V, Lurati FM, et al. The CORE

Diabetes Model: Projecting Long-term Clinical Outcomes, Costs and Costeffectiveness of Interventions in Diabetes Mellitus (Types 1 and 2) to Support Clinical and Reimbursement Decision-making. Curr Med Res Opin. 2004;20:S5–26. doi:10.1185/030079904X1980.

92. Palmer AJ, Rozea S, Valentinea WJ, Minshallb ME, Foosa V, Luratia FM, et al. Validation of the CORE Diabetes Model Against Epidemiological and Clinical Studies. Curr Med Res Opin. 2004;20:S27–40. doi:10.1185/030079904X2006.

93. McEwan P, Foos V, Palmer JL, Lamotte M, Lloyd A, Grant D. Validation of the IMS CORE Diabetes Model. Value Heal. 2014;17:714–24. doi:10.1016/j.jval.2014.07.007.

94. McEwan P, Peters JR, Bergenheim K, Currie CJ. Evaluation of the costs and outcomes from changes in risk factors in type 2 diabetes using the Cardiff stochastic simulation cost-utility model (DiabForecaster). Curr Med Res Opin. 2006;22:121–9. doi:10.1185/030079906X80350.

95. McEwan P, Ward T, Bennett H, Bergenheim K. Validation of the UKPDS 82 risk equations within the Cardiff Diabetes Model. Cost Eff Resour Alloc. 2015;13:12. doi:10.1186/s12962-015-0038-8.

96. 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;47:1747–59. doi:10.1007/s00125-004-1527-z.

97. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. 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;56:1925–33. doi:10.1007/s00125-013-2940-y.

98. Pagano E, Gray A, Rosato R, Gruden G, Perin PC, Merletti F, et al. Prediction of mortality and macrovascular complications in type 2 diabetes: validation of the UKPDS Outcomes Model in the Casale Monferrato Survey, Italy. Diabetologia. 2013;56:1726–34. doi:10.1007/s00125-013-2933-x.

99. Song M, Alexander CM, Mavros P, Lopez VA, Malik S, Phatak HM, et al. Use of the UKPDS Outcomes Model to predict all-cause mortality in U.S. adults with type 2 diabetes mellitus: comparison of predicted versus observed mortality. Diabetes Res Clin Pract. 2011;91:121–6.

100. Reynoso-Noverón N, Mehta R, Almeda-Valdes P, Rojas-Martinez R, Villalpando S,

Hernández-Ávila M, et al. Estimated incidence of cardiovascular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey. Cardiovasc Diabetol. 2011;10:1. doi:10.1186/1475-2840-10-1.

101. Eddy DM, Schkessinger L, Schlessinger L. Archimedes: A trial-validated model of diabetes. Diabetes Care. 2003;26:3093–101. doi:10.2337/diacare.26.11.3093.

102. Stern M, Williams K, Eddy D, Kahn R. Validation of Prediction of Diabetes by the Archimedes Model and Comparison With Other Predicting Models. Diabetes Care. 2008;31:1670–1. doi:10.2337/dc08-0521.

103. Eddy DM, Schlessinger L. Validation of the archimedes diabetes model. Diabetes Care. 2003;26:3102–10. http://care.diabetesjournals.org/content/26/11/3102.full-text.pdf.

104. Eddy D, Cohen M, Shum K DJ. Validation Methodology and Results. 2013. https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjMl9XZk8nZAhUHDOwKHSNgCCQQFggmMAA&url=https%3A%2F%2Fwww.evidera.com%2Fwp-content%2Fuploads%2F2017%2F10%2FPhysiology-Validation-Report1.pdf&usg=AOvVaw2t6iW479fwmpomT_z.

105. Eddy DM, Schlessinger L. Methods for Building and Validating Equations for Physiology-Based Mathematical Models: Glucose Metabolism and Type 2 Diabetes in the Archimedes Model. Med Decis Mak. 2016;36:410–21. doi:10.1177/0272989X15601864.

106. Willis M, Asseburg C, He J. Validation of economic and health outcomes simulation model of type 2 diabetes mellitus (ECHO-T2DM). J Med Econ. 2013;16:1007–21. doi:10.3111/13696998.2013.809352.

107. 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;35:375–96. doi:10.1007/s40273-016-0471-3.

108. Lundqvist A, Steen Carlsson K, Johansen P, Andersson E, Willis M. Validation of the IHE Cohort Model of Type 2 Diabetes and the Impact of Choice of Macrovascular Risk Equations. PLoS One. 2014;9:e110235. doi:10.1371/journal.pone.0110235.

109. Hoerger TJ, Segel JE, Zhang P, Sorensen SW. Validation of the CDC-RTI Diabetes Cost-Effectiveness Model. 2009. https://www.rti.org/sites/default/files/resources/mr-0013-0909-hoerger.pdf.

110. Zhou H, Isaman DJM, Messinger S, Brown MB, Klein R, Brandle M, et al. A Computer

Simulation Model of Diabetes Progression, Quality of Life, and Cost. Diabetes Care. 2005;28:2856–63. doi:10.2337/diacare.28.12.2856.

111. Ye W, Brandle M, Brown MB, Herman WH. The Michigan Model for Coronary Heart Disease in Type 2 Diabetes: Development and Validation. Diabetes Technol Ther. 2015;17:701–11. doi:10.1089/dia.2014.0304.

112. Herman WH, Ye W, Griffin SJ, Simmons RK, Davies MJ, Khunti K, et al. Early Detection and Treatment of Type 2 Diabetes Reduce Cardiovascular Morbidity and Mortality: A Simulation of the Results of the Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care (ADDITION-Europe). Diabetes Care. 2015;38:1449 LP – 1455. doi:10.2337/dc14-2459.

113. Kuo S, Ye W, Duong J, Herman WH. Are the favorable cardiovascular outcomes of empagliflozin treatment explained by its effects on multiple cardiometabolic risk factors? A simulation of the results of the EMPA-REG OUTCOME trial. Diabetes Res Clin Pract. 2018;141:181–9. doi:10.1016/j.diabres.2018.04.040.

114. Huang ES, Zhang Q, Gandra N, Chin MH, Meltzer DO. The Effect of Comorbid Illness and Functional Status on the Expected Benefits of Intensive Glucose Control in Older Patients with Type 2 Diabetes: A Decision Analysis. Ann Intern Med. 2008;149:11. doi:10.7326/0003-4819-149-1-200807010-00005.

115. Bagust A, Hopkinson PK, Maier W, Currie CJ. An economic model of the long-term health care burden of Type II diabetes. Diabetologia. 2001;44:2140–55. %22%3CGo%22,.

116. Shearer AT, Bagust A, Liebl A, Schoeffski O, Goertz A. Cost-effectiveness of rosiglitazone oral combination for the treatment of type 2 diabetes in Germany. Pharmacoeconomics. 2006;24 SUPPL. 1:35–48. https://www.scopus.com/inward/record.uri?eid=2-s2.0-33644994101&partnerID=40&md5=357795448ebc5abb92389cee8d5e4498.

117. Bagust A, Evans M, Beale S, Home PD, Perry AS, Stewart M. A model of long-term metabolic progression of type 2 diabetes mellitus for evaluating treatment strategies. Pharmacoeconomics. 2006;24 Suppl 1 SUPPL. 1:5–19. http://search.ebscohost.com/login.aspx?direct=true&db=eoh&AN=EP32039993&site=ehost-live.

118. Van Der Heijden AWAW, Feenstra TL, Hoogenveen RT, Niessen LW, de Bruijne MC, Dekker JM, et al. Policy evaluation in diabetes prevention and treatment using a population-

based macro simulation model: The MICADO model. Diabet Med. 2015;32:1580–7. doi:10.1111/dme.12811.

119. Ramos IC, van Voorn GAK, Vemer P, Feenstra TL, Al MJ. A New Statistical Method to Determine the Degree of Validity of Health Economic Model Outcomes against Empirical Data. Value Heal. 2017;20:1–7. doi:10.1016/j.jval.2017.04.016.

120. Breeze PR, Thomas C, Squires H, Brennan A, Greaves C, Diggle PJ, et al. School for Public Health Research (SPHR) Diabetes Prevention Model: Detailed Description of Model Background, Methods, Assumptions and Parameters. 2015. http://eprints.whiterose.ac.uk/97829/1/Breeze et al_SPHR Diabetes Prevention Model.pdf.

121. Javanbakht M, Mashayekhi A, Baradaran HR, Haghdoost AA, Afshin A. Projection of diabetes population size and associated economic burden through 2030 in Iran: Evidence from micro-simulation Markov model and Bayesian meta-analysis. PLoS One. 2015;10:e0132505–e0132505.

122. Dall T, Su W, Storm M. Disease Prevention Microsimulation Model. Technical Report, version 3.9.16. 2016.

123. Dall TM, Storm M V, Semilla AP, Wintfeld N, O'Grady M, Venkat Narayan KM, et al. Value of lifestyle intervention to prevent diabetes and sequelae. Am J Prev Med. 2015;48:271–80. doi:10.1016/j.amepre.2014.10.003.

124. Caro JJ, Ward AJ, O'Brien JA. Lifetime costs of complications resulting from type 2 diabetes in the U.S. Diabetes Care. 2002;25:476–81.

125. Kazemian P, Wexler DJ, Fields NF, Parker RA, Zheng A, Walensky RP. 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;21:344–55. doi:10.1089/dia.2018.0393.

126. Lau SH. A simulation model for diabetes care. ProQuest Information & Learning; 2017.

127. Zhuo TS, Bartelt-Hofer J, Brown S, Lew E, Sauriol L, Annemans L, 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. PHARMACOECONOMICS-OPEN. 2020;4:37–44. doi:10.1007/s41669-019-0156-x.

128. Wu B, Ma J, Zhang S, Zhou L, Wu H. Development and validation of a Health Policy Model of Type 2 diabetes in Chinese setting. J Comp Eff Res. 2018;7:749–63. doi:10.2217/cer-

2018-0001.

129. Shao H, Fonseca V, Stoecker C, Liu S, Shi L. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). Pharmacoeconomics. 2018;36:1125–34. doi:10.1007/s40273-018-0662-1.

130. Gillett M, Chilcott J, Goyder L, Payne N, Thokala P, Freeman C, et al. Prevention of type 2 diabetes: risk identification and interventions for individuals at high risk Economic Review and Modelling. 2012.

131. Mueller E, Maxion-Bergemann S, Gultyaev D, Walzer S, Freemantle N, Mathieu C, et al. Development and Validation of the Economic Assessment of Glycemic Control and Long-Term Effects of Diabetes (EAGLE) Model. Diabetes Technol Ther. 2006;8:219–36. doi:10.1001/jama.1965.03080010064030.

132. Blum S, Vardi M, Brown JB, Russell A, Milman U, Shapira C, et al. Vitamin E reduces cardiovascular disease in individuals with diabetes mellitus and the haptoglobin 2-2 genotype. Pharmacogenomics. 2010;11:675–84. doi:10.2217/pgs.10.17.

133. Schramm W, Sailer F, Pobiruchin M, Weiss C. PROSIT open source disease models for diabetes mellitus. Stud Health Technol Inform. 2016;226:115–8. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84978708834&doi=10.3233%2F978-1-61499-664-4-115&partnerID=40&md5=5c409bb629a6680d4492ea84e72ea16b.

134. Dadwani RS, Laiteerapong N. Economic Simulation Modeling in Type 2 Diabetes. Curr Diab Rep. 2020;20. doi:10.1007/s11892-020-01306-y.

135. Mt Hood Challenges Archieve. https://www.mthooddiabeteschallenge.com/archive. Accessed 16 Jan 2020.

136. Neal B, Perkovic V, Mahaffey KW, De Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med. 2017;377:644–57.

137. Mt Hood Challenge Website. IQVIA CORE Diabetes Model. 2021. https://www.mthooddiabeteschallenge.com/core-ims.

138. Kolbin A, Kurylev A, Balykina Y, Proskurin M. First Russian type 2 diabetes mellitus simulation model with discrete events modeling. Health-economic analysis. Value Heal. 2013;16 (7):A437–8. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed15

&AN=71232860.

139. Shestakova M V., Kolbin AS, Karpov OI, Galstyan GR, Mayorov AY, Arepeva MA, et al. An economic value of the glycated hemoglobin test in diabetes mellitus type 2 diagnosis. Diabetes Mellit. 2020;22:504–14. doi:10.14341/DM12205.

140. Luo L, Pang B, Chen J, Li Y, Xie X. Assessing the Impact of Lifestyle Interventions on Diabetes Prevention in China: A Modeling Approach. Int J Environ Res Public Health. 2019;16:1677. doi:10.3390/ijerph16101677.

141. Hauner H, Maxion-Bergemann S, Müller E, Schulz M, Huppertz E, Bergemann R, et al. Disease management program (DMP) diabetes mellitus: simultation of therapeutic results of different guidelines. A new diabetes mellitus model (DMM). Dtsch Medizinische Wochenschrift. 2003;128:1167–72. doi:10.1055/s-2003-39354.

142. Palmer AJ, Weiss C, Sendi PP, Neeser K, Brandt A, Singh G, et al. The cost-effectiveness of different management strategies for Type I diabetes: A Swiss perspective. Diabetologia. 2000;43:13–26. doi:10.1007/s001250050003.

143. Phan TP, Alkema L, Tai ES, Tan KHX, Yang Q, Lim W-Y, et al. Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore. BMJ Open Diabetes Res Care. 2014;2:e000012. doi:10.1136/bmjdrc-2013-000012.

144. Schaufler TM. Cost-effectiveness of preventing type-2-diabetes mellitus in Germany. Gesundheitsokonomie und Qual. 2009;14:71–5. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed12&AN=354406331.

145. de Vries FM, Denig P, Visser ST, Hak E, Postma MJ. Cost-Effectiveness of Statins for Primary Prevention in Patients Newly Diagnosed with Type 2 Diabetes in The Netherlands. Value Heal. 2014;17:223–30. doi:10.1016/j.jval.2013.12.010.

146. Chen J, Alemao E, Yin D, Cook J. Development of a diabetes treatment simulation model: with application to assessing alternative treatment intensification strategies on survival and diabetes-related complications. Diabetes Obes Metab. 2008;10 SUPPL.1:33–42. doi:10.1111/j.1463-1326.2008.00885.x.

147. Jacobs-Van Der Bruggen MAM, Van Baal PH, Hoogenveen RT, Feenstra TL, Briggs AH, Lawson K, et al. Cost-effectiveness of lifestyle modification in diabetic patients. Diabetes Care. 2009;32:1453–8. doi:10.2337/dc09-0363.

148. O'Reilly D, Hopkins R, Blackhouse G, Clarke P, Hux J, Tarride JE, et al. Long-term cost-utility analysis of a multidisciplinary primary care diabetes management program in Ontario. Can J Diabetes. 2007;31:205–14. doi:10.1016/S1499-2671(07)13007-0.

149. Smolen HJ, Murphy DR, Gahn JC, Yu X, Curtis BH. The Evaluation of Clinical and Cost Outcomes Associated with Earlier Initiation of Insulin in Patients with Type 2 Diabetes Mellitus. J Manag Care Pharm. 2014;20:968–84. doi:10.18553/jmcp.2014.20.9.968.

150. Brändle M, Azoulay M, Greiner R-A. Cost-effectiveness of insulin glargine versus NPH insulin for the treatment of Type 2 diabetes mellitus, modeling the interaction between hypoglycemia and glycemic control in Switzerland. Int J Clin Pharmacol Ther. 2011;49:217–30. doi:10.5414/CPP49217.

151. Pollock RFF, Valentine WJJ, Pilgaard T, Nishimura H, RF P, WJ V, et al. The cost effectiveness of rapid-acting insulin aspart compared with human insulin in type 2 diabetes patients: an analysis from the Japanese third-party payer perspective. J Med Econ. 2011;14:36–46. doi:10.3111/13696998.2010.541045.

152. Brown JB, Russell A, Chan W, Pedula K, Aickin M. The global diabetes model: user friendly version 3.0. Diabetes Res Clin Pract. 2000;50:S15–46.

153. Barhak J. The reference model estimates medical practice improvement in diabetic populations. Simul Ser. 2017;49:13–24. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85020656882&partnerID=40&md5=52a09754e2400803eb8a134751957ec6.

154. Flessa S, Zembok A. Costing of diabetes mellitus type II in Cambodia. Health Econ Rev. 2014;4:24. doi:10.1186/s13561-014-0024-4.

155. Grima DT, Thompson MF, Sauriol L. Modelling cost effectiveness of insulin glargine for the treatment of type 1 and 2 diabetes in Canada. Pharmacoeconomics. 2007;25:253–66. doi:10.2165/00019053-200725030-00007.

156. Huang ES, Basu A, O'Grady M, Capretta JC, Huang ES, Basu A, et al. Projecting the Future Diabetes Population Size and Related Costs for the US. Diabetes Care. 2009;32:2225–9. doi:10.2337/dc09-0459.

157. González JC, Walker JH, Einarson TR, González CJ, Walker JH, Einarson TR, et al. Cost-of-illness study of type 2 diabetes mellitus in Colombia. Rev Panam Salud Publica-Pan Am J Public Heal. 2009;26:55–63. %3CGo.

158. Hayes AJ, Clarke PM, Voysey M, Keech A. Simulation of Quality-Adjusted Survival in

Chronic Diseases: An Application in Type 2 Diabetes. Med Decis Mak. 2011;31:559–70. doi:10.1177/0272989X11409049.

159. Kato M, Goto A, Tanaka T, Sasaki S, Igata A, Noda M. Effects of walking on medical cost: A quantitative evaluation by simulation focusing on diabetes. J Diabetes Investig. 2013;4:667–72. doi:10.1111/jdi.12114.

160. Roberts S, Craig D, Adler A, McPherson K, Greenhalgh T. Economic evaluation of type 2 diabetes prevention programmes: Markov model of low- and high-intensity lifestyle programmes and metformin in participants with different categories of intermediate hyperglycaemia. BMC Med. 2018;16:16. doi:10.1186/s12916-017-0984-4.

161. Lamotte M, Annemans L, Lefever A, Nechelput M, Masure J. A Health Economic Model to Assess the Long-Term Effects and Cost-Effectiveness of Orlistat in Obese Type 2 Diabetic Patients. Diabetes Care. 2002;25:303–8. doi:10.2337/diacare.25.2.303.

162. Lasalvia P, Baquero L, Otálora-Esteban M, Castañeda-Cardona C, Rosselli D, Otalora-Esteban M, et al. Cost-Effectiveness of Dulaglutide Compared with Liraglutide and Glargine in Type 2 Diabetes Mellitus Patients in Colombia. Value Heal Reg Issues. 2017;14:35–40. doi:10.1016/j.vhri.2016.10.006.

163. Liu XQ, Li CP, Gong H, Cui Z, Fan LL, Yu WH, et al. An economic evaluation for prevention of diabetes mellitus in a developing country: a modelling study. BMC Public Health. 2013;13:11. doi:10.1186/1471-2458-13-729.

164. Novak S, Lawson KA. An analysis of the long-term cost-effectiveness of intensive lifestyle intervention for type 2 diabetes mellitus prevention. ProQuest Diss Theses. 2005;:305. https://search.proquest.com/docview/305393472?accountid=14532.

165. Palmer AJ, Tucker DMD. Cost and clinical implications of diabetes prevention in an Australian setting: A long-term modeling analysis. Prim Care Diabetes. 2012;6:109–21. doi:10.1016/j.pcd.2011.10.006.

166. Sullivan SD, Garrison LP, Rinde H, Kolberg J, Moler EJ. Cost-effectiveness of risk stratification for preventing type 2 diabetes using a multi-marker diabetes risk score. J Med Econ. 2011;14:609–16. doi:10.3111/13696998.2011.602160.

167. Tilden DP, Mariz S, O'Bryan-Tear G, Bottomley J, Diamantopoulos A. A lifetime modelled economic evaluation comparing pioglitazone and rosiglitazone for the treatment of type 2 diabetes mellitus in the UK. Pharmacoeconomics. 2007;25:39–54.

168. Zhang Y. Robust Optimal Control for Medical Treatment Decisions. ProQuest Diss Theses. 2014;:155. https://search.proquest.com/docview/1643246558?accountid=14532.

169. Maetzel A, Ruof J, Covington M, Wolf A, Rouf J, Covington M, et al. Economic evaluation of orlistat in overweight and obese patients with type 2 diabetes mellitus. Pharmacoeconomics. 2003;21:501–12. doi:10.2165/00019053-200321070-00005.

170. Li Y, Padron NA, Mangla AT, Russo PG, Schlenker T, Pagan JA, et al. Using Systems Science to Inform Population Health Strategies in Local Health Departments: A Case Study in San Antonio, Texas. Public Health Rep. 2017;132:549–55.

171. Ikeda S, Kobayashi M, Tajima N. Cost-effectiveness analysis of voglibose for prevention of type 2 diabetes mellitus in Japanese patients with impaired glucose tolerance. J Diabetes Investig. 2010;1:252–8. doi:10.1111/j.2040-1124.2010.00052.x.

172. Norkus A, Ostrauskas R, Sulcaite R. The economic estimates of well-timed diagnostics and early treatment of type 2 diabetes mellitus. Medicina (Kaunas). 2005;41:877–84. https://www.scopus.com/inward/record.uri?eid=2-s2.0-33644877845&partnerID=40&md5=4da73c32808f572b1bf8cd7b0b4606c7.

173. Prada MR, Guzmán NA, Bejarano DC, Benavides EK. Cost-effectiveness of using detemir for patients with type 2 diabetes in risk of cardiovascular events and death in colombia. Salud Uninorte. 2014;30:200–9. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84929937456&partnerID=40&md5=fdfb7f8c9feaf724185ce3299313a598.

174. Nagy B, Zsólyom A, Nagyjánosi L, Merész G o, Steiner T, Papp E, et al. Cost-effectiveness of a risk-based secondary screening programme of type 2 diabetes. Diabetes Metab Res Rev. 2016;32:710–29. doi:10.1002/dmrr.2791.

175. Bertram MY, Lim SS, Barendregt JJ, Vos T. Assessing the cost-effectiveness of drug and lifestyle intervention following opportunistic screening for pre-diabetes in primary care. Diabetologia. 2010;53:875–81. doi:10.1007/s00125-010-1661-8.

176. Gæde P, Valentine WJ, Palmer AJ, Tucker DMD, Lammert M, Parving H-HH-H, et al. Cost-effectiveness of intensified versus conventional multifactorial intervention in type 2 diabetes: results and projections from the Steno-2 study. Diabetes Care. 2008;31:1510–5. doi:10.2337/dc07-2452.

177. Sutherland CS, Gillet M, Brennan A, Duran A, Ansolabehere X. The canard project (projet canard) cost-effectiveness analysis of real-world outcomes in a diabetic population: The impact

of baseline characteristics in france. Value Heal. 2013;16 (3):A284. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed15&AN=71061310.

178. Wetering Van De G, Verheggen B. Dealing with competing risks in diabetes mellitus: Comparison of results using a markov versus a Micro-Markov simulation model. Value Heal. 2014;17 (3):A190. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed16&AN=71488553.

179. Pang B, Xie X, Nelson J, Li Y, Pagan JA. Using Simulation Modeling to Compare Diabetes Prevention Strategies in China. Diabetes. 2017;66:A458. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emex&AN=616961159.

180. XIE X, PANG B, LI Y. Comparing Diabetes Screening Strategies in China—An Agent-Based Modeling Approach. Diabetes. 2018;67 Supplement 1:1283-P. doi:10.2337/db18-1283-P.

181. Laroche S, Hopley C, Bourguignon S, Strauss K. Structured Education Program For Injection Technique - Budget Impact Model. Value Heal. 2017;20:A575–6. doi:10.1016/j.jval.2017.08.1005.

182. Alsultan M, Al-Omar H, Vandewalle B, Placido M, Afonso-Silva M, Mota M, et al. Metformin Extended Versus Immediate Release In Saudi Arabia: A Cost-Effectiveness Analysis. Value Heal. 2017;20. https://search.proquest.com/docview/2113726923?accountid=14532.

183. Nadkarni G, Poojary P, Ferrandino R, Saha A, Chauhan K, Mazumdar M, et al. Population Health Impact of Sodium Glucose Co-Transporter 2 Inhibitors and Glucagon-Like Peptide-1 Receptor Agonists in US Adults With Type 2 Diabetes. In: Circulation. 2018. doi:10.1161/circ.137.suppl_1.mp52.

184. Correa MF, Li Y, Kum H-C-C, Lawley MA. Assessing the Effect of Clinical Inertia on Diabetes Outcomes: a Modeling Approach. J Gen Intern Med. 2019;34:372–8. doi:10.1007/s11606-018-4773-3.

185. Al-Omar H, Alsultan M, Vandewalle B, Placido M, Afonso-Silva M, Mota M, et al. Delaying Second Line Diabetes Treatment Onset In Patients Uncontrolled With Metformin Xr In Saudi Arabia: Cost-Effectiveness Analysis. Value Heal. 2017;20:A480.

doi:10.1016/j.jval.2017.08.460.

186. Nguyen E, Coleman CI, Nair S, Weeda ER. Cost-utility of empagliflozin in patients with type 2 diabetes at high cardiovascular risk. J Diabetes Complications. 2018;32:210–5. doi:10.1016/j.jdiacomp.2017.10.006.

187. McQueen RB, Breton MD, Craig J, Holmes H, Whittington MD, Ott MA, et al. Economic Value of Improved Accuracy for Self-Monitoring of Blood Glucose Devices for Type 1 and Type 2 Diabetes in England. J Diabetes Sci Technol. 2018;12:992–1001. doi:10.1177/1932296818769098.

188. Lian J, McGhee SM, So C, Chau J, Wong CKH, Wong WCW, et al. Long-term cost-effectiveness of a Patient Empowerment Programme for type 2 diabetes mellitus in primary care. Diabetes, Obes Metab. 2019;21:73–83. doi:10.1111/dom.13485.

189. Basu A, Sohn M-W, Bartle B, Chan KCG, Cooper JM, Huang E. Development and Validation of the Real-World Progression in Diabetes (RAPIDS) Model. Med Decis Mak. 2019;39:137–51. doi:10.1177/0272989X18817521.

190. Moss SE, Klein R, Klein BEK, Wong TY. Retinal Vascular Changes and 20-Year Incidence of Lower Extremity Amputations in a Cohort With Diabetes. Arch Intern Med. 2003;163:2505–10.

191. Zhou H, Isaman DJM, Messinger S, Brown MB, Klein R, Brandle M, et al. A computer simulation model of diabetes progression, quality of life, and cost. Diabetes Care. 2005;28:2856–63. doi:10.2337/diacare.28.12.2856.

192. Eastman RC, Javitt JC, Herman WH, Dasbach EJ, Zbrozek AS, Dong F, et al. Model of Complications of NIDDM: I. Model construction and assumptions. Diabetes Care. 1997;20:725–34. doi:10.2337/diacare.20.5.725.

193. Moss SE, Klein R, Klein BE. The prevalence and incidence of lower extremity amputation in a diabetic population. Arch Intern Med. 1992;152:610–6. http://www.ncbi.nlm.nih.gov/pubmed/1546925. Accessed 4 Sep 2017.

194. Moss, Scot E.; Klein, Ronald; Klein BEK. Long-term incidence of lower-extremity amputations in a diabetic population. Arch Fam Med. 1996;5:391–8. doi:10.1001/archfami.5.7.391.

195. Klein, Ronald; Klein, Barbara E.K.; Moss SE. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. Arch Ophthalmol. 1984;102:527–32.

doi:10.1001/archopht.1984.01040030405011.

196. Moss SE, Klein R, Klein BEK. Ten-year Incidence of Visual Loss in a Diabetic Population. Ophthalmology. 1994;101:1061–70. doi:10.1016/S0161-6420(94)31217-6.

197. Moss SE, Klein R, Klein BEK. The Incidence of Vision Loss in a Diabetic Population. Ophthalmology. 1988;95:1340–8. doi:10.1016/S0161-6420(88)32991-X.

198. Klein, Ronald; Klein, Barbara E. K.; Moss SE et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. XIV. Ten-Year Incidence and Progression of Diabetic Retinopathy. Arch Ophthalmol. 1994;112:1217. doi:10.1001/archopht.1994.01090210105023.

199. Klein, Ronald; Klein, Barbara E. K.; Moss SE et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. X. Four-Year Incidence and Progression of Diabetic Retinopathy When Age at Diagnosis Is 30 Years or More. Arch Ophthalmol. 1989;107:244–9. doi:10.1001/archopht.1989.01070010250031.

200. Klein, Ronald; Klein, Barbara E.K.; Moss SE et al. Glycosylated Hemoglobin Predicts the Incidence and Progression of Diabetic Retinopathy. JAMA J Am Med Assoc. 1988;260:2864–71. doi:10.1001/jama.1988.03410190112033.

201. Klein R, Klein BEK, Moss SE. Prevalence of Microalbuminuria in Older-Onset Diabetes. Diabetes Care. 1993;16:1325–30. doi:10.2337/diacare.16.10.1325.

202. Bruno G, Biggeri A, Merletti F, Bargero G, Ferrero S, Pagano G, et al. Low Incidence of End-Stage Renal Disease and Chronic Renal Failure in Type 2 Diabetes: A 10-year prospective study. Diabetes Care. 2003;26:2353–8. doi:10.2337/diacare.26.8.2353.

203. Group UPDS. Overview of 6 years' therapy of type II diabetes: a progressive disease. UK Prospective Diabetes Study 16. Diabetes. 1995;44:1249–58.

204. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ. 2000;321:405–12. http://www.ncbi.nlm.nih.gov/pubmed/10938048. Accessed 4 Sep 2017.

205. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet. 1998;352:837–53.

206. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-Year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med. 2008;359:1577–89.

207. Sasaki A, Uehara M, Horiuchi N, Hasegawa K, Shimizu T. A 15 year follow-up study of patients with non-insulin dependent diabetes mellitus (NIDDM) in Osaka, Japan long-term prognosis and causes of death. Diabetes Res Clin Pract. 1996;34:47–55. doi:10.1016/S0168-8227(96)01329-0.

208. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC, Bigger JT, et al. Effects of Intensive Glucose Lowering in Type 2 Diabetes. N Engl J Med. 2008;358:2545–59. doi:10.1056/NEJMoa0802743.

209. ACCORD Study Group, Cushman WC, Evans GW, Byington RP, Goff DC, Grimm RH, et al. Effects of Intensive Blood-Pressure Control in Type 2 Diabetes Mellitus. N Engl J Med. 2010;362:1575–85. doi:10.1056/NEJMoa1001286.

210. ADVANCE Collaborative Group, Patel A, MacMahon S, Chalmers J, Neal B, Billot L, et al. Intensive Blood Glucose Control and Vascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med. 2008;358:2560–72. doi:10.1056/NEJMoa0802987.

211. Woodward M, Patel A, Zoungas S, Liu L, Pan C, Poulter N, et al. Does glycemic control offer similar benefits among patients with diabetes in different regions of the world? Results from the ADVANCE trial. Diabetes Care. 2011;34:2491–5.

212. Patel A, ADVANCE Collaborative Group. Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. Lancet. 2007;370:829–40.

213. Knopp RH, d'Emden M, Smilde JG, Pocock SJ. Efficacy and Safety of Atorvastatin in the Prevention of Cardiovascular End Points in Subjects With Type 2 Diabetes: The Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus (ASPEN). Diabetes Care. 2006;29:1478–85. doi:10.2337/dc05-2415.

214. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, et al. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med. 2009;360:129–39.

215. Griffin SJ, Borch-Johnsen K, Davies MJ, Khunti K, Rutten GE, Sandbæk A, et al. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): A cluster-randomised trial. Lancet. 2011;378:156–67.

216. Lauritzen T, Griffin S, Borch-Johnsen K, Wareham NJ, Wolffenbuttel BHR, Rutten G.

The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening. Int J Obes. 2000;24:6–11.

217. Dahlöf B, Sever PS, Poulter NR, Wedel H, Beevers DG, Caulfield M, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-B. Lancet. 2005;366:895–906.

218. Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HAW, Livingstone SJ, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet (London, England). 2004;364:685–96. doi:10.1016/S0140-6736(04)16895-5.

219. Thomason MJ, Colhoun HM, Livingstone SJ, Mackness MI, Betteridge DJ, Durrington PN, et al. Baseline characteristics in the Collaborative AtoRvastatin Diabetes Study (CARDS) in patients with Type 2 diabetes. Diabet Med. 2004;21:901–5.

220. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, et al. Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus. N Engl J Med. 2013;369:1317–26. doi:10.1056/NEJMoa1307684.

221. Alpert W, Hospital M. Cardiovascular Effects of Intensive Lifestyle Intervention in Type 2 Diabetes. N Engl J Med. 2013;369:145–54. doi:10.1056/NEJMoa1212914.

222. Espeland M et al. Reduction in Weight and Cardiovascular Disease Risk Factors in Individuals With Type 2 Diabetes: One-year results of the Look AHEAD trial. Diabetes Care. 2007;30:1374–83. doi:10.2337/dc07-0048.

223. National Center for Health Statistics. National Health and Nutrition Examination Survey, 1988-1994. 1994. https://www.cdc.gov/nchs/nhanes/index.htm.

224. Collins R, Armitage J, Parish S, Sleight P, Peto R. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: A randomised placebo-controlled trial. Lancet. 2002;360:7–22.

225. Gerstein HC, Yusuf S, Mann JFE, Hoogwerf B, Zinman B, Held C, et al. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: Results of the

HOPE study and MICRO-HOPE substudy. Lancet. 2000;355:253–9.

226. Gerstein HC, Bosch J, Pogue J, Taylor DW, Zinman B, Yusuf S. Rationale and Design of a Large Study to Evaluate the Renal and Cardiovascular Effects of an ACE Inhibitor and Vitamin E in High-Risk Patients With Diabetes: The MICRO-HOPE Study. Diabetes Care. 1996;19:1225–8. doi:10.2337/diacare.19.11.1225.

227. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. Prevention of Cardiovascular Events and Death with Pravastatin in Patients with Coronary Heart Disease and a Broad Range of Initial Cholesterol Levels. N Engl J Med. 1998;339:1349–57. doi:10.1056/NEJM199811053391902.

228. Frick MH, Elo O, Haapa K, Heinonen OP, Heinsalmi P, Helo P, et al. Helsinki Heart Study: Primary-Prevention Trial with Gemfibrozil in Middle-Aged Men with Dyslipidemia. N Engl J Med. 1987;317:1237–45. doi:10.1056/NEJM198711123172001.

229. SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). SHEP Cooperative Research Group. JAMA. 1991;265:3255—3264. http://europepmc.org/abstract/MED/2046107.

230. Rifkind BM. Lipid research clinics coronary primary prevention trial: Results and implications. Am J Cardiol. 1984;54:30–4. doi:10.1016/0002-9149(84)90854-3.

231. Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. Br Med J. 1985;291 July:97–104.

232. Rubins HB, Robins SJ, Collins D, Fye CL, Anderson JW, Elam MB, et al. Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. N Engl J Med. 1999;341:410–8.

233. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. N Engl J Med. 1995;333:91–7.

234. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs Diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA J Am Med Assoc. 2002;288:2981–97. doi:10.1001/jama.288.23.2981.

235. Rahman M et al. Renal Outcomes in High-Risk Hypertensive Patients Treated With an Angiotensin-Converting Enzyme Inhibitor or a Calcium Channel Blocker vs a Diuretic. Arch Intern Med. 2005;165:936. doi:10.1001/archinte.165.8.936.

236. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, Michelson EL, et al. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: The CHARM-overall programme. Lancet. 2003;362:759–66.

237. Deedwania P, Barter P, Carmena R, Fruchart JC, Grundy SM, Haffner S, et al. Reduction of low-density lipoprotein cholesterol in patients with coronary heart disease and metabolic syndrome: analysis of the Treating to New Targets study. Lancet. 2006;368:919–28.

238. LaRosa JC, Grundy SM, Waters DD, Shear C, Barter P, Fruchart J-C, et al. Intensive Lipid Lowering with Atorvastatin in Patients with Stable Coronary Disease. N Engl J Med. 2005;352:1425–35. doi:10.1056/NEJMoa050461.

239. Flechtner-Mors M, Ditschuneit HH, Johnson TD, Suchard MA, Adler G. Metabolic and weight loss effects of long-term dietary intervention in obese patients: Four-year results. Obes Res. 2000;8:399–402.

240. Manley SE, Stratton IM, Cull CA, Frighi V, Eeley EA, Matthews DR, et al. Effects of three months' diet after diagnosis of Type 2 diabetes on plasma lipids and lipoproteins (UKPDS 45). Diabet Med. 2000;17:518–23.

241. Tinker LF. Low-Fat Dietary Pattern and Risk of Treated Diabetes Mellitus in Postmenopausal Women. The Women's Health Initiative Randomized Controlled Dietary Modification Trial. Arch Intern Med. 2008;168:1500–11.

242. Howard B V., Van Horn L, Hsia J, Manson JE, Stefanick ML, Wassertheil-Smoller S, et al. Low-Fat Dietary Pattern and Risk of Cardiovascular Disease. JAMA. 2006;295:655. doi:10.1001/jama.295.6.655.

243. Chamnan P, Simmons RK, Forouhi NG, Luben RN, Khaw KT, Wareham NJ, et al. Incidence of type 2 diabetes using proposed HbA1c diagnostic criteria in the european prospective investigation of cancer-norfolk cohort: Implications for preventive strategies. Diabetes Care. 2011;34:950–6.

244. National Centre for Social Research, University College London Department of Epidemiology and Public Health. Health Survey for England. 2003. https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=5098#!#1.

245. Bethel MA, Sloan FA, Belsky D, Feinglos MN. Longitudinal incidence and prevalence of adverse outcomes of diabetes mellitus in elderly patients. Arch Intern Med. 2007;167:921–7.

246. Hill C, Gerardo D, James F, Tyroler HA, Chambless LE, Romm J, et al. THE ATHEROSCLEROSIS RISK IN COMMUNITIES (ARIC) STUDY: DESIGN AND OBJECTIVES. Am J Epidemiol. 1989;129:687–702. doi:10.1093/oxfordjournals.aje.a115184.

247. Bash LD, Astor BC, Coresh J. Risk of Incident ESRD: A Comprehensive Look at Cardiovascular Risk Factors and 17 Years of Follow-up in the Atherosclerosis Risk in Communities (ARIC) Study. Am J Kidney Dis. 2010;55:31–41. doi:10.1053/j.ajkd.2009.09.006.

248. Dalrymple LS, Katz R, Kestenbaum B, Shlipak MG, Sarnak MJ, Stehman-Breen C, et al. Chronic kidney disease and the risk of end-stage renal disease versus death. J Gen Intern Med. 2011;26:379–85.

249. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health. 1951;41:279–81.

250. Adler AI, Boyko EJ, Ahroni JH, Stensel V, Forsberg RC, Smith DG. Risk Factors for Diabetic Peripheral Sensory Neuropathy: Results of the Seattle Prospective Diabetic Foot Study. Diabetes Care. 1997;20:1162–7. doi:10.2337/diacare.20.7.1162.

251. Zethelius B, Eliasson B, Eeg-Olofsson K, Svensson AM, Gudbjörnsdottir S, Cederholm J. 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;93:276–84.

252. Eeg-Olofsson K, Cederholm J, Nilsson PM, Zethelius B, Svensson AM, Gudbjörnsdóttir S, et al. New aspects of HbA1c as a risk factor for cardiovascular diseases in type 2 diabetes: An observational study from the Swedish National Diabetes Register (NDR). J Intern Med. 2010;268:471–82.

253. Kelly PJ, Clarke PM, Hayes AJ, Gerdtham UG, Cederholm J, Nilsson P, et al. Predicting mortality in people with Type 2 diabetes mellitus after major complications: A study using Swedish National Diabetes Register data. Diabet Med. 2014;31:954–62.

254. Ravid M, Brosh D, Ravid-Safran D, Levy Z, Rachmani R. Main risk factors for nephropathy in type 2 diabetes mellitus are plasma cholesterol levels, mean blood pressure, and hyperglycemia. Arch Intern Med. 1998;158:998–1004.

255. Ravid M, Brosh D, Levi Z, Bar-Dayan Y, Ravid D, Rachmani R. Use of enalapril to

attenuate decline in renal function in normotensive, normoalbuminuric patients with type 2 diabetes mellitus: A randomized, controlled trial. Ann Intern Med. 1998;128 12 PART 1:982–8.

256. Ansquer JC, Foucher C, Rattier S, Taskinen MR, Steiner G. Fenofibrate reduces progression to microalbuminuria over 3 years in a placebo-controlled study in type 2 diabetes: Results from the Diabetes Atherosclerosis Intervention Study (DAIS). Am J Kidney Dis. 2005;45:485–93.

257. Partanen J, Niskanen L, Lehtinen J, Mervaala E, Siitonen O, Uusitupa M. Natural history of peripheral neuropathy in patients with non-insulin-dependent diabetes mellitus. N Engl J Med. 1995;333:89–94.

258. Humphrey, Linda L.; Palumbo. Pasquale J.; Butters MA et al. The Contribution of Non—Insulin-Dependent Diabetes to Lower-Extremity Amputation in the Community. Arch Intern Med. 1994;154:885–92. doi:10.1001/archinte.1994.00420080085009.

259. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from Coronary Heart Disease in Subjects with Type 2 Diabetes and in Nondiabetic Subjects with and without Prior Myocardial Infarction. N Engl J Med. 1998;339:229–34. doi:10.1056/NEJM199807233390404.

260. Lee ET, Keen H, Bennett PH, Fuller JH, Lu M. Follow-up of the WHO multinational study of vascular disease in diabetes: general description and morbidity. Diabetologia. 2001;44:S3–13. doi:10.1007/PL00002936.

261. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med. 2006;355:2427–43.

262. Gu K, Cowie CC, Harris MI. Diabetes and decline in heart disease mortality in US adults. J Am Med Assoc. 1999;281:1291–7.

263. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet. 2008;371:1783–9. doi:10.1016/S0140-6736(08)60766-7.

264. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. N Engl J Med. 2001;344:1343–50. doi:10.1056/NEJM200105033441801.

265. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V. The Indian

Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). Diabetologia. 2006;49:289–97.

266. Kosaka K, Noda M, Kuzuya T. Prevention of type 2 diabetes by lifestyle intervention: A Japanese trial in IGT males. Diabetes Res Clin Pract. 2005;67:152–62.

267. Chiasson JL, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M. Acarbose for prevention of type 2 diabetes mellitus: The STOP-NIDDM randomised trial. Lancet. 2002;359:2072–7.

268. Digestive and Kidney Disease National Diabetes Date Group, National Institute of Diabetes. Diabetes in America. National Institutes of Health and National Institute of Diabetes and Digestive and Kidney Diseases; 1995.

269. Paas GRA, Veenhuizen KCW. Research on the reliability of the National Ambulent Register. Onderz naar betrouwbaarheid van LMR Utr Prism. 2002.

270. Nichols GA, Joshua-Gotlib S, Parasuraman S. Independent Contribution of A1C , Systolic Blood Pressure , and LDL Cholesterol Control to Risk of Cardiovascular Disease Hospitalizations in Type 2 Diabetes : An Observational Cohort Study. J Gen Intern Med. 2013;28:691–7.

271. Gall M-A, Rossing P, Skøtt P, Damsbo P, Vaag A, Bech K, et al. Prevalence of micro- and macroalbuminuria, arterial hypertension, retinopathy and large vessel disease in European Type 2 (non-insulin-dependent) diabetic patients. Diabetologia. 1991;34:655–61. doi:10.1007/BF00400995.

272. Escobedo J, Rana JS, Lombardero MS, Albert SG, Davis AM, Kennedy FP, et al. Association between albuminuria and duration of diabetes and myocardial dysfunction and peripheral arterial disease among patients with stable coronary artery disease in the BARI 2D study. Mayo Clin Proc. 2010;85:41–6.

273. Park J-Y, Kim H-K, Chung YE, Kim SW, Hong SK, Lee K-U. Incidence and Determinants of Microalbuminuria in Koreans With Type 2 Diabetes. Diabetes Care. 1998;21:530–4. doi:10.2337/diacare.21.4.530.

274. Chin MH, Drum ML, Jin L, Shook ME, Huang ES, Meltzer DO. Variation in Treatment Preferences and Care Goals Among Older Patients With Diabetes and Their Physicians. Med Care. 2008;46:275–86. doi:10.1097/MLR.0b013e318158af40.

275. Hoogenveen RT, Gijsen R, van Genugten MLL, Kommer GJ, Schouten JSAG, de Hollander, A E M et al. Dutch DisMod: constructing a set of consistent data for chronic disease modelling. 2000; March:1–117.

276. Ruwaard D. Diabetes Mellitus: from epidemiology to health policy. 1996. http://hdl.handle.net/1765/22585.

277. van Os N, Niessen LW, Koopmanschap MA, van der Lei J. Detailed analysis of the societal costs of diabetes mellitus. Ned Tijdschr Geneeskd. 2000;144:842–6. http://www.ncbi.nlm.nih.gov/pubmed/10816774.

278. van Os N, Niessen LW, Bilo HJG, Casparie AF, van Hout BA. Diabetes nephropathy in the Netherlands: A cost effectiveness analysis of national clinical guidelines. Health Policy (New York). 2000;51:135–47. doi:10.1016/s0168-8510(00)00063-4.

279. Esteghamati A, Etemad K, Koohpayehzadeh J, Abbasi M, Meysamie A, Noshad S, et al. Trends in the prevalence of diabetes and impaired fasting glucose in association with obesity in Iran: 2005-2011. Diabetes Res Clin Pract. 2014;103:319–27.

280. Sheehan TJ, DuBrava S, DeChello LM, Fang Z. Rates of weight change for black and white Americans over a twenty year period. Int J Obes. 2003;27:498–504.

281. Gerstein HC, Santaguida P, Raina P, Morrison KM, Balion C, Hunt D, et al. Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: A systematic overview and meta-analysis of prospective studies. Diabetes Res Clin Pract. 2007;78:305–12.

282. McCarthy DJ, Zimmet P, Dalton A, Segal L, Welborn TA. The rise and rise of diabetes in Australia, 1996, A Review of Statistics, Trends and Costs. 1996;:1–60.

283. Department of Health and Aged Care. Insights into the utilisation of health services in Australia based on linked administrative data. Canberra; 2000.

284. Mathers CD, Vos ET, Stevenson CE. The burden of disease and injury in Australia. Canberra; 1999.

285. The Emerging Risk Factors Collaboration. Diabetes Mellitus, Fasting Glucose, and Risk of Cause-Specific Death. N Engl J Med. 2011;364:829–41. doi:10.1056/NEJMoa1008862.

286. Chan J, So W, Ko G, Tong P, Yang X, Ma R, et al. The Joint Asia Diabetes Evaluation (JADE) Program: a web-based program to translate evidence to clinical practice in Type 2 diabetes. Diabet Med. 2009;26:693–9. doi:10.1111/j.1464-5491.2009.02751.x.

287. National Cancer Institute. NCI Dictionary of Cancer Terms. coronary heart disease.

288. Black WC. The CE Plane: A Graphic Representation of Cost-Effectiveness. Med Decis Mak. 1990;10:212–4.

6 Appendix

Search	Hits	Hits
((((((((("Diabetes Mellitus"[MeSH Terms]) OR diabet*[Title/Abstract]) OR T2DM[Title/Abstract]))	572987	578565
((((((((((markov[Title/Abstract]) OR "monte carlo"[Title/Abstract]) OR "montecarlo"[Title/Abstract]) OR stochastic[Title/Abstract]) OR deterministic[Title/Abstract]) OR "discrete event"[Title/Abstract]) OR "agent based"[Title/Abstract]) OR "agentbased"[Title/Abstract])) OR "Computer Simulation"[MeSH Terms]) OR ((((computer[Title/Abstract]) OR simulat*[Title/Abstract])) AND ((((model[Title/Abstract]) OR models[Title/Abstract]) OR modelling[Title/Abstract]) OR modeling[Title/Abstract]))	397849	403130
"1995/01/01"[Date - Publication] : "2017/08/18"[Date - Publication]	16261812	16377310
(Animals[MeSH Terms]) NOT Humans[MeSH Terms]	16441621	4359835
((((((Comment[Publication Type]) OR Letter[Publication Type]) OR "Newspaper Article"[Publication Type]) OR News[Publication Type]) OR Addresses[Publication Type]) OR Editorial[Publication Type]) OR "Published Erratum"[Publication Type]	1772258	1785856

((((((((((("Diabetes Mellitus"[MeSH Terms]) OR diabet*[Title/Abstract]) OR T2DM[Title/Abstract])) AND ((((((((((markov[Title/Abstract]) OR "monte carlo"[Title/Abstract]) OR "montecarlo"[Title/Abstract]) OR stochastic[Title/Abstract]) OR deterministic[Title/Abstract]) OR "discrete event"[Title/Abstract]) OR "agent based"[Title/Abstract]) OR "agentbased"[Title/Abstract])) OR "Computer Simulation"[MeSH Terms]) OR ((((computer[Title/Abstract]) OR simulat*[Title/Abstract])) AND ((((model[Title/Abstract]) OR models[Title/Abstract]) OR modelling[Title/Abstract]) OR modeling[Title/Abstract])))))) AND "1995/01/01"[Date - Publication] : "2017/08/18"[Date - Publication])) NOT ((Animals[MeSH Terms]) NOT ((Animals[MeSH Terms]) AND Humans[MeSH Terms]))))) NOT (((((((Comment[Publication Type]) OR Letter[Publication Type]) OR "Newspaper Article"[Publication Type]) OR News[Publication Type]) OR Addresses[Publication Type]) OR Editorial[Publication Type]) OR "Published Erratum"[Publication Type])	3313	3378

Table 11: Search Terms for MEDLINE

Number	Name of model	Author	Year	Source	Journal of publication	Development of former model
1	IMS CORE Diabetes Model	Palmer AJ. et al.	2004	[91]	Current Medical Research and Opinion	Yes, Accusim and IMIB [70]
2	Cardiff Model	McEwan P et al.	2006, 2015	[94, 95]	Current Medical Research and Opinion Cost Effectiveness and Resource Allocation	Yes, Microsimulation:model from Caro JJ/ Eastman
3	UKPDS Outcomes Model	Clarke PM et al. Hayes AJ et al.	2004 2013	[96, 97]	Diabetologia Diabetologia	No
4	Archimedes Model	Eddy DM et al.	2003	[101]	Diabetes Care	No
5	IHE/ ECHO T2DM-Model	Willis M et al. Willis M et al. Lundqvist A.	2013 2017 2014	[106–108]	Journal of Medical Economics PharmacoEconomics PLoS ONE	No
6	CDC/ RTI Model	Hoerger TJ et al.	2009	[109]	RTI Press	No
7	Michigan Model for Diabetes	Zhou et al.	2005	[110]	Diabetes Care	No
8	NHS model	Farmer AJ et al.	2014	[85]	Health Technology Assessment	Yes, UKPDS Outcomes Model
9	Chicago model	Huang ES et al.	2008	[114]	Annals of Internal Medicine	Yes, Microsimulation:model from Caro JJ/ Eastman
10	DiDACT model	Bagust A et al.	2001	[115]	Diabetologia	No
11	MICADO	Van Der Heijden A-WAW et al.	2015	[118]	DiabeticMedicine	No
12	Markov model/Vijan	Vijan S et al.	1997	[87]	Annals of Internal Medicine	No
13	Dutch diabetes model	Niessen LW et al.	2003	[88]	The Netherlands Journal of Medicine	Yes, Microsimulation:model from Caro JJ/ Eastman
14	SPHR Diabetes	Breeze PR et al.	2015	[120]	none	No
15	Microsimulation/Javanbakht	Javanbakht M et al.	2015	[121]	PLoS ONE	No
16	IHSD DPMM	Dall T et al.	2016	[122]	none	No
17	Microsimulation:model from Caro JJ/ Eastman	Caro JJ et al.	2002	[124]	Diabetes Care	No
18	Australian Diabetes Model	Walker A et al.	2003	[89]	none	No
19	Discrete event simulation model from Jiao	Jiao F et al.	2019	[90]	Endocrine	No
20	PREDICT-DM	Kazemian P et al.	2019	[125]	Diabetes Technology & Therapeutics	No
21	CDOM	Lau SH	2017	[126]	ProQuest	No
22	Cornerstone Diabetes Simulation	Zhuo TS et al.	2020	[127]	PharmacoEconomics	Yes, UKPDS Outcomes Model

23	BRAVO Model	Shao H et al.	2018	[129]	PharmacoEconomics	No
24	Sheffield Diabetes Model	Gillet M et al.[130]	2012	[130]	None	Yes, Microsimulation:model from Caro JJ/ Eastman
25	EAGLE	Mueller E et al.	2006	[131]	Diabetes Technology & Therapeutics	No
26	Evidence-Based Medicine Integrator	Blum S et al.	2010	[132]	Pharmacogenomics	No
27	PROSIT Model	Schramm W et al.	2016	[133]	Studies in Health Technology and Informatics	No
28	Chinese Outcomes Model for T2DM	Wu B et al.	2018	[128]	Journal of Comparative Effectiveness Research	No
29	Russian T2DM model	Shestakova MV et al.	2020	[138, 139]	Diabetes mellitus	Yes, UKPDS Outcomes Model and EAGLE
30	An agent-based model/Luo	Luo L. et al.	2019	[140]	International Journal of Environmental Research and Public Health	No
31	Diabetes Mellitus Model (DMM)	Hauner H. et al.	2003	[141]	Deutsche Medizinische Wochenschrift	No
32	Swiss model	Palmer A et al.	2000	[142]	Diabetologia	No
33	Markov models/Singapore	Phan TP et al.	2014	[143]	BMJ Open Diabetes Research & Care	No
34	Markov models/German	Schaufler	2009	[144]	Gesundheitsökonomie und Qualitätsmanagement	No
35	Markov models/de Vries	de Vries FM et al.	2014	[145]	Value in Health	No
36	JADE Model	Chen J et al.	2008	[146]	Diabetes Obesity & Metabolism	Yes, UKPDS Outcomes Model
37	Chronic Diseases Model (CDM) at the Dutch National Institute for Public Health and the Environment	Jacobs-van der Bruggen MAM et al.	2009	[147]	Diabetes Care	No
38	Ontario Diabetes Economic Model (ODEM)	O'Reilly D et al.	2007	[148]	Canadian Journal of Diabetes	Yes, UKPDS Outcomes Model
39	Medical Decision Modeling Inc.	Smolen H et al.	2014	[149]	Journal of Managed Care Pharmacy	No
40	DES model	Brändle M et al.	2011	[150]	International Journal of Clinical Pharmacology and Therapeutics	No
41	Discret event simulation/Pollock	Pollock R et al.	2011	[151]	Journal of Medical Economics	No
42	Global diabetes model (GDM)	Brown JB et al.	2000	[152]	Diabetes Research and Clinical Practice	Yes, Microsimulation:model from Caro JJ/ Eastman
43	Reference Model	Barhak J	2017	[153]	Simulation Series	No

44	Markov model/Cambogia	Flessa S and Zembok A	2014	[154]	Health Economics Review	No
45	Markov models/Grima	Grima DT et al.	2007	[155]	PharmacoEconomics	No
46	Markov models/Huang	Huang ES et al.	2009	[156]	Diabetes Care	No
47	Markov models/Gonzalez	González JC et al.	2009	[157]	Revista Panamericana De Salud Publica-Pan American Journal of Public Health	Yes, Michigan Model for Diabetes
48	Markov models/Hayes	Hayes AJ et al.	2011	[158]	Medical Decision Making	No
49	Markov models/Kato	Kato M et al.	2013	[159]	Journal of Diabetes Investigation	No
50	Markov models/Roberts	Roberts S et al.	2018	[160]	BMC Medicine	No
51	Markov models/Lamotte	Lamotte M et al.	2002	[161]	Diabetes Care	No
52	Markov models/Lasalvia	Lasalvia P et al.	2017	[162]	None	No
53	Markov models/Liu	Liu XQ et al.	2013	[163]	BMC Public Health	No
54	Markov models/Novak	Novak S	2005	[164]	ProQuest Dissertations and Theses	No
55	Markov models/Palmer	Palmer AJ and Tucker DMD	2012	[165]	Primary Care Diabetes	No
56	Markov models/Sullivan	Sullivan SD et al.	2011	[166]	Journal of Medical Economics	No
57	Markov models/Tilden	Tilden DP et al.	2007	[167]	PharmacoEconomics	No
58	Markov models/Zhang	Zhang Y	2014	[168]	ProQuest Dissertations and Theses	No
59	Markov models/Maetzel	Maetzel A et al.	2003	[169]	PharmacoEconomics	No
60	Agent-based/Li	Li Y et al.	2017	[170]	Public Health Reports	No
61	Markov models/ikeda	Ikeda S et al.	2010	[171]	Journal of Diabetes Investigation	No
62	Markov models/Norkus	Norkus A. et al.	2005	[172]	Medicina (Kaunas)	No
63	Markov models/Prada	Prada MR et al.	2014	[173]	Salud Uninorte	No
64	Syreon	Nagy B et al.	2016	[174]	Diabetes/Metabolism Research and Reviews	No
65	Microsimulation:Prevention Project	Bertram M et al.	2010	[175]	Diabetologia	No
66	Steno	Gæde, P et al.	2008	[176]	Diabetes Care	No
67	CANARD model	Sutherland C et al.	2013	[177]	Value in Health	No
68	DELTA	Itering Van De G and Verheggen B	2014	[178]	Value in Health	No

69	New York Academy of Medicine Diabetes Simulation Model-China	Pang B et al.	2017	[179]	Diabetes	No
70	An agent-based model/Xie	Xie X et al.	2018	[180]	Diabetes	No
71	A budget impact model/Laroch	Laroche S et al.	2017	[181]	Value in Health	No
72	A cost-effectiveness model/Alsultan	Asultan M et al.	2017	[182]	Value in Health	No
73	A microsimulation model/Nadkarni	Nadkarni G et al.	2018	[183]	Circulation	No
74	An agent-based model/Correa	Correa MF et al.	2019	[184]	Journal of General Internal Medicine	No
75	Discrete Event Simulation/Al-omar	Al-Omar H et al.	2017	[185]	Value in Health	No
76	EMPA-REG	Nguyen E et al.	2018	[186]	Journal of Diabetes and its Complications	No
77	Markov model/McQueen	McQueen RB et al.	2018	[187]	Journal of Diabetes Science and Technology	No
78	Patient level simulation model/Lian	Lian J et al.	2019	[188]	Diabetes, Obesity and Metabolism	Yes, UKPDS Outcomes Model
79	RAPIDS	Basu A et al.	2019	[189]	Medical Decision Making	No

Table 12: List of models from abstract screening

Model	Only definition of external validation	Only description of external validation process (no exact definition given)	Definition and description of external validation	No definition given	Definition of external validation	Type of validation
IMS CORE Diabetes Model			x		"Third-order [external] validation was made against published epidemiological or clinical studies which had not been used to provide input data for transition probabilities in the CORE Diabetes Model" [92] "External validity quantifies how well the model predicts outcomes observed in the real world" and "External validation compares clinical events predicted by the CDM with observed clinical outcomes using studies not directly used to inform disease progression within the model." [93]	Validation of clinical outcomes (cumulative incidence) by using number of events and relative risks. [92, 93]
Cardiff Model	x				"External validation compares output from the model with data not specifically used to construct the disease progression algorithms." [95]	"Clinical events, consistently defined between the publication and those predicted by the Cardiff Model were compared over the relevant time horizon." [95]
UKPDS Outcomes Model	x				"to gain acceptance, it is important that such models can demonstrate reliable predictive performance not just with the patient sample used to develop the model, but with external patient populations who have not been involved in the model's development and estimation"[98]	In a specific group/population: the applicability of the UKPDS Outcomes Model among T2DM patients in populations outside the UK and with a wider range of patient characteristics [98–100] External validation was performed by comparing predicted

				cumulative incidences of diabetes-related health outcomes/ mortality with the observed cumulative incidences/ mortality in a different clinical setting. [98–100]
Archimedes Model	x		Studies which were not used to build the model. [103] In an external validation excercise "a model is set up to simulate a real study such as a clinical trial, and the model's results are compared to the actual outcomes." and "Three main types of external validations are conducted: simulation of clinical trials, simulation of cohort follow-up studies, and comparisons of age-specific incidence rates in cohort studies and registries" [104]	Validating the cumulative incidence of outcomes over time. [103] Validating the model's prediction of the incidence of diabetes and compare with alternative prediction models (logistic models). [102] Validation of health outcomes, and for three trials biomarker values. [104]
IHE/ ECHO-T2DM Model		x	"external (and predictive) validation (i.e., testing whether the model accurately predicts actual outcomes observed in patients in clinical trials or observational registries)"[108] Data source was not used to build the model. [106] "External validation involves simulating interventional or observational clinical studies and comparing model predictions with actual observed outcomes (i.e., assessing the degree of concordance) and is the primary focus of this study." [107]	Comparison of cumulative incidence of outcomes between model and study [106–108]

CDC/ RTI Model	x		"For the validation, we input baseline model parameters from studies that were used to create the model (internal validation) and those that were found from a review of the literature (external validation)." [109]	Valiation of clinical outcomes (cumulative incidence) [109]
Michigan Model for Diabetes	x		"studies […] were not used to develop our model or calibrate its parameters." [111]	Comparison of incidence of model and studies [111] Validation of incidence of cardiovascular events [112]
NHS model		x	not given	Comparing proportion of patients with a complication. [85]
Chicago Model	x		"the present study aimed to externally validate the prognostic accuracy of the Chicago Model compared with the status quo of prognostication: the physician's judgment. " [86]	Comparing life expectancy of model and physician's prediction. [86]
DiDACT model	x		Estimated prevalence was validated by comparison with international rates. [116]	Estimated prevalences were compared with external data. [116] mentioned, not described [117]
MICADO	x		"When the empirical data are not used to estimate the input parameters of the model, this is often called independent or external validation." [119] "To externally validate the MICADO model, we compared model-based estimates with empirical estimates of the incidence of end-stage microvascular complications in the Netherlands." [118]	Comparison of absolute number of patients with a complication [119] Comparison of estimated incidences [118]

Markov model/Vijan		x		"One way to test the validity of this model is to determine whether it can accurately predict the rates of microvascular disease reported in actual patient populations. " [87]	Comparison of predicted incidence of model and study [87]
Dutch diabetes model		x		"we validated model outputs, comparing model output data with empirical data from other sources" [88]	Comparison of different outcomes (life expectancy, complication occurrence) [88]
SPHR Diabetes		x		Four tests were developed "to compare model outcomes with reported data from external data sources." [120]	Comparing incidence, prevalence, mortality and distributions (mean, SD, median) of metabolic risk factors with observations. [120]
Microsimulation/Javanbakht			x	not given	Comparison of prediction of number of people with diabetes [121]
IHSD DPMM	x			External validation is defined as replication of "findings of studies (e.g., clinical trials) not used in model development". [122]	Validation of clinical outcomes (cumulative incidence, prevalence and mortality of diabetes, comorbidities, mortality, average values of biometrics (BMI, BP)): % of patients experienced events or having condition. [122] Validation of annual transition rates to diabetes absence intervention and incidence of cardiovascular events [123]
Microsimulation:model form Caro JJ /Eastman			x	not given	not given
Australian Diabetes Model			x	not given	Comparing annual per patient costs for diabetes and DALYs. [89]

Discrete event simulation model from Jiao				x	not given	Validation of life expectancy. [90]
PREDICT-DM		x			"We then externally validated PREDICT-DM against the VADT and the Look AHEAD trials to ensure that PREDICT-DM adequately predicts the outcomes of a study not used to inform disease progression within the simulation model."[125]	Validation of clinical outcomes. [125]
CDOM	x				"Here, external validity refers to the application of the study results to other populations or settings."[126]	Validation of of survival. [126]
Cornerstone Diabetes Simulation			x		"external validity (i.e., model results are consistent with actual real-world outcomes)" [127]	Validation of clinical outcomes. [127]
Chinese Outcomes Model for T2DM	x				"externally consistent with the observed outcomes not specifically employed to develop the disease risk equations." [128]	Validation of clinical outcomes (cumulative incidence). [128]
BRAVO Model				x	not given	Validation of mortality, incidence of marcovascular, and microvascular complications. [9]
Sheffield Diabetes Model	x				External validation: "tests against data independent from the model" [10]	Validation of cumulative incidence of cardiovascular events, all-cause mortality, and quality-adjusted life expectancy taking into account macrovascular complications and the costs of complications by therapy allocation over time horizons of 4 years. [10]
EAGLE	x				External validation: "tests against data independent from the model" [10]	Validation of cumulative incidence of cardiovascular events, all-cause mortality, and quality-adjusted life expectancy taking into account macrovascular

						complications and the costs of complications by therapy allocation over time horizons of 4 years. [10]
Evidence-Based Medicine Integrator		x			"Performance of "four external validation analyses against three recent clinical trials" [75]	Validation of mortality, incidence of marcovascular, and microvascular complications. [75]
PROSIT Model				x	not given	Validation of mortality, incidence of marcovascular, and microvascular complications. [9]
TTM				x	not given	Validation of mortality, incidence of marcovascular, and microvascular complications. [9]
IQVIA CORE Diabetes Model				x	not given	Validation of mortality, incidence of marcovascular, and microvascular complications. [9]
Sum	8	10	3	9		

Table 13: Definition, description and use of external validation

Model	Methods used in all published validation studies	Reference	Results	Graphs/ Figures	Tables
IMS CORE Diabetes Model	1. Regression analysis [92, 93] 2. Two-sided paired Student t-test (5%) for difference of event rates [93] 3. CI [93] 4. Mean absolute percentage error (MAPE) and root mean square percentage error (RMSPE). [93]	[92]	1. Accurate/adequate predictions 2. Only clinical settings and not real-life setting are considered with validation. 3. Discrepancy in the validation of nephropathy	Scatterplots of observed vs. predicted outcomes	Overview of validation results
		[93]	1. Accurate/adequate predictions 2. Discussion of limitations	1. Scatterplots of observed vs. predicted outcomes 2. point estimates and confidence intervals	1. Overview of validation results 2. Summary measures of goodness of fit
Cardiff Model	1. Regression analysis 2. Mean absolute percentage error (MAPE) 3. Coefficient of determination (R^2)	[95]	1. accurate/adequate predictions 2. "missing risk factor information [should be] imputed with care."	Scatterplots of observed vs. predicted outcomes	1. Overview of regression analysis 2. Summary measures of goodness of fit
UKPDS Outcomes Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) and estimating discrimination (c-statistic) and calibration (Hosmer-Lemeshow 2) [100] 2. CI [98, 99] 3. Sensitivity analysis, in which baseline risk factors are changed by ± 15% [98]	[98]	1. accurate/adequate predictions 2. "long-term (>15-year) predictions should be used with caution."	Prediction of incidences over time	Overview of validation results
		[99]	1. Accurate/adequate predictions 2. Model overpredicts mortality in patients with a duration of diabetes of 6 years or longer	point estimates and confidence intervals	Comparison of baseline characteristics
		[100]	1. Accurate/adequate predictions 2. unprecise if baseline characteristics of persons are different to those used to construct the model	none	none

Archimedes Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) [103, 105] 2. Regression analysis [103] 3. Kaplan-Meier curves and log-rank test. [103] 4. Receiver operating characteristic (ROC) curve, and Spearman correlation tests [102] 5. Hazard rates, hazard ratios, and combined hazard ratio for meta-analysis and CI [104, 105] 6. “validation hazard ratio” (vHR) calculated by fitting a Cox proportional hazard model. Hypothesis testing with a log-rang test for Cox model. [104]	[103]	Accurate/adequate predictions	1. Prediction of incidences over time compared with observed data 2. Kaplan Meier-curve 3. Scatterplots of observed vs. predicted outcomes	Overview of validation results
		[102]	Accurate/adequate predictions	receiver operating characteristics (ROC) curve	none
		[104]	1. Accurate/adequate predictions 2. "However, in the DPP metformin did not reduce OGTT by the same magnitude as metformin reduced FPG, and therefore the trial showed a smaller reduction in progression to diabetes. Based on this validation and other findings, the role of OGTT and its calculation in the Model is being reconsidered. Until the calculation of OGTT is improved, diabetes projects will focus on FPG and HbA1c."	1. point estimates and confidence intervals 2. Prediction of incidences over time compared with observed data	Overview of validation results all
		[105]	Accurate/adequate predictions	Prediction of incidences over time compared with observed data	none

IHE/ ECHO-T2DM Model	1.Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) [106] 2. Regression analysis [106–108] 3.Coefficient of determination (R^2) [106–108] 4. F-test to assess coefficients of regression analysis [107] 5. Mean average error (MAE), root mean squared error (RMSE), mean squared log of the accuracy ratio (MSLAR), and mean squared logit error (MSLE) [107]	[108]	1. Accurate/adequate predictions 2. Swedish NDR and UKPDS-OM2 slightly underestimate macrovascular outcome	Scatterplots of observed vs. predicted outcomes	none
		[106]	1. Accurate/adequate predictions 2. "The F-test fails to reject the null hypothesis of agreement between the model" 3. "While most of the points lie relatively close to the identity line, there is a tendency for the predicted values to exceed the actual trial values and several outliers are noticeable."	Scatterplots of observed vs. predicted outcomes	Overview of validation results
		[107]	1. Accurate/adequate predictions 2. failure to reject F test 3. under-prediction of macrovascular events	Scatterplots of observed vs. predicted outcomes	1. Overview of validation results and 95% CI 2. Overview of regression analysis
CDC/ RTI Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Regression analysis 3. Coefficient of determination (R^2)	[109]	1. Accurate/adequate predictions 2. Limitation to populations that were used in the construction (United States and United Kingdom 3. Studies, which do not fit the model structure were excluded from the results 4. MI and stroke were underestimated after 9 years of simulation horizon	1. Scatterplots of observed vs. predicted outcomes 2. Prediciton of incidences over time compared with observed data	Overview of validation results
Michigan Model for Diabetes	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Regression analysis 3. Coefficient of determination (R^2)	[112]	Accurate/adequate predictions	none	Overview of validation results

		[111]	1. Accurate/adequate predictions for 5 studies 2. "However, eight of the observed incidence rates were outside the simulated 95% confidence intervals provided by the simulation model." 3. "Because the sample size used in a study affects the Monte Carlo error, we performed each simulation with the number of patients reported in the trial with 500 repetitions. The resulting 95% confidence intervals are likely to be too narrow because they did not take into account the uncertainty in model parameters and unmeasured or unreported characteristics of the study population."	Scatterplots of observed vs. predicted outcomes	Overview of validation results
NHS model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Sensitivity analysis, and CI	[85]	1. Accurate/adequate predictions 2. lack of testing under one year	Prediction of incidences over time compared with observed data	none
Chicago Model	1. sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) 2. receiver-operating characteristic (ROC) curve, and c-statistic, and integrated Brier score (IBS)	[86]	1. Accurate/adequate predictions 2. subjective data basis (assessment of the physicians) 3. results are limited to used population	1. Boxplots of life expectancy (LE) predictions 2. Kaplan-Meier-Curve 3. ROC curve	Overview of validation results
DiDACT model	none described	[117]	long-term effects are difficult to interpret due to interactions of therapies	none	none
		[116]	none describe	none	none
MICADO	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Sensitivity analysis, and CI	[118]	Accurate/adequate predictions	Scatterplots of observed vs. predicted outcomes	Overview of validation results

		[119]	1. "the model may be considered valid only given relatively low requirements on accuracy for this specific outcome." 2. "Results indicate a high probability of a valid outcome is associated with relatively wide accuracy intervals. In particular, 25% deviation from the observed outcome implied approximately 60% expected validity."	1. Histogram for simulated number of patients with ESRD (end-stage renal disease) 2. Validity curve (deviation from observed outcomes vs. Probability that the model outcome is valid)	Overview of validation results
Markov model/Vijan	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Sensitivity analysis.	[87]	1. Accurate/adequate predictions 2. "The only exceptions are 1) the rate of early retinopathy in the UKPDS (28% to 32% according to our model compared with 37% seen in the trial [55]) and 2) the 10-year rate of blindness from retinopathy in the Wisconsin Epidemiologic Study (0.1% to 0.8% according to our model compared with approximately 1.1% seen in the study [30]; this rate is calculated on the basis of the estimate that one quarter of cases of blindness in the study were due to diabetic retinopathy) (15). "	none	none
Dutch diabetes model	none described	[88]	1. Accurate/adequate predictions 2. In the outcomes, we "found only minor differences, which we explain by the lack of an, increasing, incidence trend, underestimation in the registries and varying diagnostic criteria."	none	none
SPHR Diabetes	Comparing outcomes (incidences, prevalence, differences, absolute values, proportions)	[120]	1. No accurate/adequate predictions 2. "We suggest that the model may be more accurate at predicting diabetes incidence over a longer time period due to the nature of the quadratic equations used to predict HbA1c."	1. Comparison of different observed biomarkers and prevelances with predicted values 2. Development of biomarkers over time	Overview of validation results

Microsimulation/Javanbakht	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Check if there is less than 5% difference between estimations.	[121]	none describe	none	none
IHSD DPMM	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. CI	[123]	1. Accurate/adequate predictions 2. "For some populations the validation activities suggested that simulated growth rate in biometrics and disease onset appeared to be high (or low) when compared to published sources."	none	Overview of validation results
		[122]	none describe	none	Overview of validation results
Microsimulation:model form Caro JJ /Eastman	none described	[124]	Accurate/adequate predictions	none	none
Australian Diabetes Model	Comparing outcomes (incidences, prevalence, differences, absolute values, proportions)	[89]	Accurate/adequate predictions	none	none
Discrete event simulation model from Jiao	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. CI	[90]	Accurate/adequate predictions	none	Overview of validation results
PREDICT-DM	1. Mean absolute percentage error (MAPE), median absolute percentage error (MEDAPE), and root mean square percentage error (RMSPE) 2. Bland–Altman graphs 3. Intraclass coefficient (ICC) and CI 4. Kaplan-Meier curves	[125]	Accurate/adequate predictions	1. Bland–Altman graphs 2. Kaplan-Meier-Curve	Overview of validation results

CDOM	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Log-rank test	[126]	1. Accurate/adequate predictions 2. signifcant differences in CHD, heart failure, and cancer 3. "depending on the setting in which these patients are managed, the model will need to be adjusted"	Comparing survival of observed vs. predicted outcomes	Overview of validation results
Cornerstone Diabetes Simulation	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Regression analysis 3. Coefficient of determination (R^2)	[127]	1. Accurate/adequate predictions 2. For ADVANCE overestimation of all-cause mortality, blood pressure und blood glucose intervention	Scatterplots of observed vs. predicted outcomes	Overview of validation results
Chinese Outcomes Model for T2DM	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Mean absolute percentage error (MAPE) 3. Regression analysis 4. Coefficient of determination (R^2)	[128]	1. Accurate/adequate predictions 2. Slightly underestimation in eastern popluations and overestimation in western popluation	Scatterplots of observed vs. predicted outcomes	none
BRAVO Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Regression analysis 3. Coefficient of determination (R^2) 4. F-test 5. hazard ratio 6. Mean absolute percentage error (MAPE) 7. CI	[9, 129]	Accurate/adequate predictions	Scatterplots of observed vs. predicted outcomes with confidence intervals	Overview of validation results
Sheffield Diabetes Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Sensitivity analysis	[10]	The model overestimates macrovascular events.	none	Overview of validation results
EAGLE	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. Sensitivity analysis	[10]	Underestimation of MI and stroke. Overestimation of any acute CVD event.	none	Overview of validation results

Evidence-Based Medicine Integrator	Comparing outcomes (incidences, prevalence, differences, absolute values, proportions)	[75]	Overestimation or underestimation of incidences. No robust	Scatterplots of observed vs. predicted outcomes	Overview of validation results
PROSIT Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. hazard ratio 3. Mean absolute percentage error (MAPE) 4. CI	[9]	Models underestimated treatment effects	none	Overview of validation results
TTM	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. hazard ratio 3. Mean absolute percentage error (MAPE) 4. CI	[9]	Models underestimated treatment effects	none	Overview of validation results
IQVIA CORE Diabetes Model	1. Comparing outcomes (incidences, prevalence, differences, absolute values, proportions) 2. hazard ratio 3. Mean absolute percentage error (MAPE) 4. CI	[9]	Models underestimated treatment effects	none	Overview of validation results

Table 14: Results of external validation

Studies	Reference to used studies	Reference to validation paper	Year of MtHood Challenge	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Moss et al. (2003)	[190]	[92]		X																								1
ISDR	[191, 192, 201, 193–200]	[92, 104, 107]		X			X								X													3
Eastman et al. (1997)	[192]	[92, 109]		X					X																			2
Bruno et al. (2003)	[202]	[92, 109]		X					X																			2
UKPDS 16	[203]	[87]													X													1
UKPDS 35	[204]	[87, 94]			X																							1
UKPDS 33	[205]	[92, 93, 107, 115, 124]		X				X					X							X						X		5
UKPDS 80	[206]	[92, 93, 107]		X				X																		X		3
Sasaki et al. (1996)	[207]	[92, 106, 107, 109]		X				X	X																			3
ACCORD	[208, 209]	[75, 85, 93, 109, 111, 127]	2010	X	X	X		X	X	X															X	X		8
ADVANCE	[210–212]	[75, 93, 95, 109, 111, 118, 127]	2010	X	X	X		X	X	X				X											X	X	X	10
ASPEN	[213]	[75, 93, 95, 107, 127]	2010	X	X	X		X	X	X															X	X	X	9
VADT	[214]	[93, 95, 111, 125]		X	X					X													X			X		5
ADDITION	[215, 216]	[93, 95, 111, 120]		X	X					X							X											4
ASCOT	[217]	[93, 95]		X	X																							2
CARDS	[218, 219]	[10, 93, 95, 104, 106, 107, 109]	2004	X	X	X	X	X	X											X						X	X	9
SAVOR-TIMI	[220]	[95]			X																							1
AHEAD	[221]	[95]			X																							1
Look AHEAD	[221, 222]	[104, 107, 125]					X	X															X			X		4
National Health and Nutrition Examination	[223]	[99]				X																						1

Survey 1988–1994																												
EMPA-REG	[49]	[9]	2018		X	X		X	X	X				X			X										X	8
CANVAS	[136]	[9]	2018		X	X		X	X	X				X			X										X	8
DPP	[61]	[103, 104]					X																					1
HPS	[224]	[103, 104]					X																					1
Micro-HOPE	[225, 226]	[103, 107]					X	X																				2
LIPID	[227]	[103]					X																					1
HHS	[228]	[103]					X																					1
SHEP	[229]	[103, 104]					X																					1
LRC-CPPT	[230]	[103]					X																					1
MRC	[231]	[103]					X																					1
VA-HIT	[232]	[103]					X																					1
WOSCOPS	[233]	[103]					X																					1
ALLHAT	[234, 235]	[104]					X													X								2
CHARM	[236]	[104]					X																					1
TNT	[237, 238]	[104]					X													X								2
Flechtner-Mors	[239]	[104]					X																					1
UKPDS45	[240]	[104]					X																					1
WHI DMT	[241, 242]	[104]					X																					1
EPIC-Norfolk	[243]	[120]															X											1
HSE 2011	[244]	[120]															X											1
Medicare cohort	[245]	[104]					X																					1
ARIC cohort	[246, 247]	[104]					X																					1
CHS	[248]	[104]					X																					1
FHS original cohort	[249]	[104]					X																					1
SEATTLE	[250]	[104]					X																					1
SIdish NDR	[251–253]	[107, 118]						X						X														2
Ravid et al. (1998)	[254]	[109]							X																			1

Ravid et al. (1998)	[255]	[109]							X																				1
Ansquer et al. (DAIS) (2005)	[256]	[109]							X																				1
Partanen et al. (1995)	[257]	[92, 109]		X					X																				2
Humphrey et al. (1994)	[258]	[109]							X																				1
Haffner et al. (1998)	[259]	[109]							X																				1
Lee et al. (2001)	[260]	[109]							X																				1
ADOPT	[261]	[109]							X	X																	X		3
Gu et al. (1999)	[262]	[109]							X																				1
Li et al. (2008)	[263]	[109]							X																				1
Tuomilehto et al. (2001)	[264]	[109]							X																				1
Ramachandran et al. (2006)	[265]	[109]							X																				1
Kosaka et al. (2005)	[266]	[109]							X																				1
Chiasson et al. (2002)	[267]	[109]							X																				1
National Diabetes Data Group (1995)	[268]	[116]											X																1
Dutch Medical Register (2003)	[269]	[118]												X															1
Kaiser Permanente	[270]	[118]												X															1
Gall et al. (1991)	[271]	[85]									X																		1
Bari 2D	[272]	[85]									X																		1
Park et al. (1998)	[273]	[85]									X																		1
Survey (physicians surveyed for Chicago Model)	[274]	[86]										X																	1
Hoogenveen er al. (2000)	[275]	[88]														X													1
Ruwaard (1996)	[276]	[88]														X													1
van Os et al. (2000)	[277]	[88]														X													1

van Os et al. (2000)	[278]	[88]														X												1
SuRFNCD	[279]	[86]																X											1
Sheehan et al. (2003)	[280]	[98]																	X										1
Gerstein et al. (2007)	[281]	[98]																	X										1
Mc Carty er al. (1996)	[282]	[89]																			X								1
Department of Health and Aged Care (2000)	[283]	[89]																			X								1
Mathers, Vos and Stevenson (1999)	[284]	[89]																			X								1
The Emerging Risk Factors Collaboration (2011)	[285]	[90]																				X							1
JADE Registry	[286]	[126]																						X					1
ADVANCE Asian Study		[128]																								X			1
Hong Kong Diabetes Regis-try (HKDR)		[128]																								X			1
RAMP-DM		[128]																								X			1
Osaka		[128]																								X			1
EMPATHY		[128]																								X			1
JDCS		[128]																								X			1
J-EDIT		[128]																								X			1
				15	12	7	24	12	23	8	3	1	2	6	2	4	5	1	2	4	3	1	2	1	3	16	5	162

1. IMS CORE Diabetes Model, 2. Cardiff Model, 3. UKPDS Outcomes Model, 4. Archimedes Model, 5. IHE/ ECHO-T2DM Model, 6. CDC/ RTI Model, 7. Michigan Model for Diabetes, 8. NHS model, 9. Chicago Model, 10. DiDACT model, 11. MICADO, 12. Markov model/Vijan, 13. Dutch diabetes model, 14. SPHR Diabetes, 15. Microsimulation/Javanbakht, 16. IHSD DPMM, 17. Microsimulation:model form Caro JJ/Eastman, 18. Australian Diabetes Model, 19. Discrete event simulation model from Jiao, 20. PREDICT-DM, 21. CDOM, 22. Cornerstone Diabetes Simulation, 23. Chinese Outcomes Model for T2DM, 24. BRAVO Model

X Validation was performed

Table 15: Studies used for external validation in paper-based models grouped by model and study

Studies	25	26	27	28	29	30	
Moss et al. (2003)							1
WESDR							3
Eastman et al.							2
Bruno et al. (2003)							2
UKPDS 16							1
UKPDS 35							1
UKPDS 33							5
UKPDS 80							3
Sasaki et al. (1996)							3
ACCORD			X				9
ADVANCE			X				11
ASPEN							9
VADT							5
ADDITION							4
ASCOT							2
CARDS	X	X					11
SAVOR-TIMI							1
AHEAD							1
Look AHEAD							4
Casale Monferrato Study							2
National Health and Nutrition Examination Survey 1988–1994							1
Diabetes Care System West Friesland, Netherlands (A van der Heijden, G Nijpels)							2
EMPA-REG				X	X	X	11

CANVAS				X	X	X	11
DPP							1
HPS							1
Micro-HOPE							2
LIPID							1
HHS							1
SHEP							1
LRC-CPPT							1
MRC							1
VA-HIT							1
WOSCOPS							1
ALLHAT							2
CHARM							1
TNT							2
Flechtner-Mors							1
UKPDS45							1
WHI DMT							1
EPIC-Norfolk							1
HSE 2011							1
Medicare cohort							1
ARIC cohort							1
CHS							1
FHS original co-hort							1
SEATTLE							1
Swedish NDR							2
Ravid et al. (1998)							1
Ravid et al. (1998)							1
Ansquer et al. (DAIS) (2005)							1

Partanen et al. (1995)							2
Humphrey et al. (1994)							1
Haffner et al. (1998)							1
Lee et al. (2001)							1
ADOPT							3
Gu et al. (1999)							1
Li et al. (2008)							1
Tuomilehto et al. (2001)							1
Ramachandran et al. (2006)							1
Kosaka et al. (2005)							1
Chiasson et al. (2002)							1
National Diabetes Data Group (1995)							1
Dutch Medical Register (2003)							1
Kaiser Permanente							1
Gall et al. (1991)							1
Bari 2D							1
Park et al. (1998)							1
Survey (physicians surveyed for Chicago Model)							1
Hoogenveen er al. (2000)							1
Ruwaard (1996)							1
van Os et al. (2000)							1
van Os et al. (2000)							1
SuRFNCD							1
Sheehan et al. (2003)							1

Gerstein et al. (2007)							1
Mc Carty er al. (1996)							1
Department of Health and Aged Care (2000)							1
Mathers, Vos and Stevenson (1999)							1
The Emerging Risk Factors Collaboration (2011)							1
JADE Registry							1
ADVANCE Asian Study							1
Hong Kong Diabetes Registry (HKDR)							1
RAMP-DM							1
Osaka							1
EMPATHY							1
JDCS							1
J-EDIT							1
	1	1	2	2	2	2	172

25. Sheffield Diabetes Model, 26. EAGLE, 27. Evidence-Based Medicine Integrator, 28. PROSIT Model, 29. TTM, 30. IQVIA CORE Diabetes Model

X Validation was performed

Table 16: Studies used for external validation in MtHChs based models grouped by model and study

Risk factors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		26	27	28	29	30	
Socioeconomic																									49							54
Age	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X		21	X	X					23
Sex	X	X	X	X		X	X	X	X	X			X	X		X	X	X	X	X	X	X	X		19	X	X					21
Education									X																1							1
Income									X																1							1
Race (skin colour)						X			X								X			X		X	X		6	X						7
marital status									X																1							1
Laboratory result				X																					57							63
Lipids	X							X																	23						X	26
Triglycerides			X																						1							1
Cholesterol		X	X		X	X	X				X			X			X			X			X		10							10
HDL		X	X		X									X								X	X		6	X	X					8
LDL		X							X													X		X	4							4
other laboratory results																									33							36
HbA1c	X	X	X		X		X	X	X		X	X	X	X			X			X	X	X	X	X	17	X	X				X	20
serum creatinine																				X			X		2							2
eGFR		X			X			X														X			4							4
Albuminuria		X	X																			X			3							3
urine albumin:creatinine ratio																							X		1							1

white blood cell amount		X			X																	X			3							3
Micro- or macroalbuminuria					X			X																	2							2
Hemoglobin level																						X			1							1
T2DM characteristics																									10							12
Duration of disease			X					X	X									X							4	X	X					6
T2DM					X				X			X									X				4							4
diabetes incidence															X										1							1
diabetes medicines									X																1							1
Cardiological																									22							25
Blood pressure	X	X			X	X		X			X			X		X									8	X					X	10
Systolic blood pressure		X	X				X	X	X					X						X		X	X	X	10		X					11
Heart rate		X			X																	X			3							3
hypertension medicins									X																1							1
Weight																									11							12
BMI			X				X	X	X		X			X		X	X					X		X	10							10
weight																									0						X	1
weight gain		X																							1							1

Comorbidities																									21							21
Retinopathy					X													X				X			3							3
nephropathy									X									X				X			3							3
neuropathy									X									X				X			3							3
peripheral vascular disease																		X			X	X			3							3
Macrovascular Disease																		X							1							1
CHF																					X	X			2							2
CVD history																				X			X		2							2
Atrial fibrillation																						X			1							1
MI																						X			1							1
Stroke																					X	X			2							2
Behavior																									12							14
Smoker		X	X		X		X	X	X		X			X			X			X		X		X	12	X	X					14
Other																									15							15
Severe hypoglycemia																								X	1							1
EQ-5D														X											1							1
lipid-lowering medicines									X											X			X		3							3
blood-pressure lowering medicines																				X			X		2							2
oral antiglycemices																				X			X		2							2

anticoagulants																			X			X		2							2
Mortality															X			X						2							2
Prevalence																		X						1							1
Ratio diagnosed undiagnosed																		X						1							1
	5	14	11	3	12	5	7	11	18	2	5	3	3	10	2	4	7	11	2	13	7	22	14	6	197	8	7	0	0	0	4	216

1. IMS CORE Diabetes Model, 2. Cardiff Model, 3. UKPDS Outcomes Model, 4. Archimedes Model, 5. IHE/ ECHO-T2DM Model, 6. CDC/ RTI Model, 7. Michigan Model for Diabetes, 8. NHS model, 9. Chicago Model, 10. DiDACT model, 11. MICADO, 12. Markov model/Vijan, 13. Dutch diabetes model, 14. SPHR Diabetes, 15. Microsimulation/Javanbakht, 16. IHSD DPMM, 17. Microsimulation:model form Caro JJ/Eastman, 18. Australian Diabetes Model, 19. Discrete event simulation model from Jiao, 20. PREDICT-DM, 21. CDOM, 22. Cornerstone Diabetes Simulation, 23. Chinese Outcomes Model for T2DM, 24. BRAVO Model, 25. Sheffield Diabetes Model, 26. EAGLE, 27. Evidence-Based Medicine Integrator, 28. PROSIT Model, 29. TTM, 30. IQVIA CORE Diabetes Model

X Validation was performed

Table 17: Risk factors used for validation exercises

Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		25	26	27	28	29	30	
Incidence				X												X									2							2
Cumulative incidence						X																			1							1
Diabetes prevelance															X										1							1
Mortality (death)/ survival	X	X	X	X	X	X	X		X		X			X			X			X	X	X	X	X	16			X	X	X	X	20
Life expactancy									X				X						X						3							3
Prevalence of complications										X															1							1
Macrovascular diseases											X											X			2			X				3
CVD (cardiovascular disease)		X	X	X	X	X	X				X			X									X	X	10	X	X	X	X			14
CHD (coronary heart disease)		X			X																X	X			4			X				5
Angina														X								X		X	3							3
MI (fatal and non fatal)	X	X	X	X	X	X	X				X			X						X		X	X	X	13	X	X		X	X	X	18
Cardiovascular death	X	X	X	X	X		X				X			X						X		X	X		11				X	X	X	14
CHF (congestive heart failure)	X	X	X		X		X				X									X	X		X	X	10			X				11
Hospitalisation for HF		X	X		X		X							X										X	6					X	X	8
Hospitalisation for Angina		X	X											X										X	4					X	X	6
TIA														X											1							1
Stroke	X	X	X	X	X	X	X				X			X						X	X	X	X	X	14	X	X		X	X	X	19

PVD (peripheral vascular disease)																					X	X			2							2
Revascularisation (due to diabetic foot ulcers (DFU))							X																	X	2							2
(Lower extremity) amputation	X	X	X		X	X					X			X									X		8				X	X	X	11
Sum Macrovascular diseases	5	9	8	4	8	4	7			0	7		0	9	0	0								8	90	3	3	4	5	6	6	117
Microvascular diseases		X															X					X			3							3
Retinopathy	X				X						X	X					X					X			6			X				7
Cataract	X																								1							1
Blindness					X	X																	X		3							3
Microalbuminuria	X				X	X		X												X		X			6			X				7
Gross proteinuria	X				X	X		X												X					5							5
Nephropathy												X	X									X			3							3
End-stage renal disease/ renal failure	X	X	X		X	X					X			X						X	X		X		10							10
Death from end-stage renal disease	X																								1							1
Neuropathy					X	X							X									X	X		5			X				6
Sum Microvascular diseases	6	2	1	0	6	5	0	2	0	0	2	2	2	1	0	0	2	0						0	43	0	0	3	0	0	0	46
Other Complications																																
Hypoglycaemia events requiring assistance	X					X																			2							2
Annual per patient cost estimates for diabetes																		X							1							1

DALYs																		X							1							1
Cancer																					X				1							1
Progression of HbA1c										X															1							1
	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	0	1	0	0	0	6	0	0	0	0	0	0	6
Total sum	13	12	10	6	15	12	8	2	2	2	10	2	3	11	1	1	3	2	1	8	7	13	10	9	163	3	3	8	6	7	7	197

1. IMS CORE Diabetes Model, 2. Cardiff Model, 3. UKPDS Outcomes Model, 4. Archimedes Model, 5. IHE/ ECHO-T2DM Model, 6. CDC/ RTI Model, 7. Michigan Model for Diabetes, 8. NHS model, 9. Chicago Model, 10. DiDACT model, 11. MICADO, 12. Markov model/Vijan, 13. Dutch diabetes model, 14. SPHR Diabetes, 15. Microsimulation/Javanbakht, 16. IHSD DPMM, 17. Microsimulation:model form Caro JJ/Eastman, 18. Australian Diabetes Model, 19. Discrete event simulation model from Jiao, 20. PREDICT-DM, 21. CDOM, 22. Cornerstone Diabetes Simulation, 23. Chinese Outcomes Model for T2DM, 24. BRAVO Model, 25. Sheffield Diabetes Model, 26. EAGLE, 27. Evidence-Based Medicine Integrator, 28. PROSIT Model, 29. TTM, 30. IQVIA CORE Diabetes Model

Macrovascular: Cardiovascular disease (CVD): 1. coronary heart disease (CHD)/ coronary artery disease (I25.1) def.: disease of the blood vessels supplying the heart muscle (· angina pectoris (I20), · MI (I21-22), · Cardiovascular death, explanation: For this event there is no special code. It refers to the underlying disease, · CHF (congestive heart failure) I50.0), 2. peripheral artery disease (I73.9) def.: disease of blood vessels supplying the arms and legs (Revascularisation (due to diabetic foot ulcers (DFU), (Lower extremity) amputation), 3. cerebrovascular disease I60-I69 def.: disease of the blood vessels supplying the brain (stroke (I61-63)) [21, 22, 25, 28, 287]

Microvascular: 1. Retinopathy (E11.3†), 2. Nephropathy (E11.2†), 3. Neuropathy (E11.4†) [22]

+ Validated Outcome

Table 18: Validated outcomes grouped per model and outcome

7 Acknowledgment

Ein langer Weg mit viel harter Arbeit geht nun zu Ende. Ich möchte mich dabei bei allen bedanken die mich auf diesem Weg unterstützt haben. Im privaten Bereich sind dies allen voran meine Frau Julia ohne die ich sicher diese lange Zeit nicht so erfolgreich überstanden hätte. Dann möchte ich meinen Eltern besonders danken, die mir diese Ausbildung über die ganzen Jahre ermöglicht haben. Weiterhin meiner Schwester Melanie und ihrem Mann Christian für hilfreiche Tipps. Schließlich zählt mein Kommilitone Bahne Bahners dazu, mit dem ich das Studium gemeistert habe und der mir mit fachlichem Rat zur Seite stand.

Im universitären Bereich möchte ich meiner Betreuerin Katherine Ogurtsova besonders für die Mühe und Ausdauer während der Promotion danken, die mir immer mit Rat und Tat zur Seite stand. Ferner möchte ich mich noch bei Ute Linnenkamp and Thomas Heise bedanken, die beim Abstractscreening viel geholfen haben. Außerdem möchte ich noch entscheidend Frau Icks danken, die immer zeitnah bei Fragen zur Verfügung stand und die Arbeit durch ihre Anmerkungen entscheidend vorangebracht hat. Schließlich möchte ich noch Herrn Dintsios und Herrn Holle für das Korrekturlesen und abschließende hilfreiche Hinweise danken, welche die Arbeit abgerundet haben.