ORIGINAL RESEARCH ARTICLE

Comparing the Cohort and Micro-Simulation Modeling Approaches in Cost-Effectiveness Modeling of Type 2 Diabetes Mellitus: A Case Study of the IHE Diabetes Cohort Model and the Economics and Health Outcomes Model of T2DM

Michael Willis[1] · Adam Fridhammar[1] · Jens Gundgaard[2] · Andreas Nilsson[1] · Pierre Johansen[2]

Abstract

Background Economic modeling is widely used in estimating cost-effectiveness in type 2 diabetes mellitus. Because type 2 diabetes is complex and patients are heterogenous, the cohort modeling approach may generate biased estimates of cost-effectiveness. The IHE Diabetes Cohort Model (IHE-DCM) was constructed using the cohort approach as an alternative for stakeholders with limited resources, some of whom have voiced reasonable concerns about a lack of transparency with type 2 diabetes micro-simulation models and long run times.

Objectives The objective of this study was to inform decision makers by investigating the direction and magnitude of bias of IHE-DCM cost-effectiveness estimates that can be attributed to the cohort modeling approach.

Methods Simulation scenarios inspired by the 9th Mount Hood Diabetes Challenge were simulated with IHE-DCM and with a micro-simulation model, the Economic and Health Outcomes Model of T2DM (ECHO-T2DM), and key metrics (absolute and incremental costs and quality-adjusted life-years, event rates, and cost-effectiveness) were compared for evidence of systematic differences. The models were harmonized to the extent possible to ensure that differences were driven primarily by the unit of observation and not by other model differences.

Results IHE-DCM run times were faster and IHE-DCM produced uniformly larger estimates of absolute life-years, quality-adjusted life-years, and costs than ECHO-T2DM but smaller between-arm (incremental) differences. Estimated incremental cost-effectiveness ratios and net monetary benefits varied similarly and predictably across the scenarios. On average, IHE-DCM estimates of incremental cost-effectiveness ratios and net monetary benefits were CAN\$269 (3%) and CAN\$2935 (10%) smaller, respectively, than ECHO-T2DM.

Conclusions There was little evidence that estimated cost-effectiveness metrics, the outcomes that matter most to stakeholders, differed systematically.

1 Introduction

Type 2 diabetes mellitus (T2DM) is a chronic and progressive disease hallmarked by hyperglycemia. Chronic hyperglycemia, together with common co-morbidities

✉ Michael Willis
mw@ihe.se

1   The Swedish Institute for Health Economics, Box 2017, 220 02 Lund, Sweden

2   Novo Nordisk A/S, Søborg, Denmark

such as obesity, hypertension, and dyslipidemia, is associated with high risks for serious micro- and macro-vascular complications and premature mortality [1, 2]. Currently, T2DM cannot be cured and treatment consists primarily of managing blood glucose and cardiovascular risk factors (e.g., blood pressure and serum lipids) to postpone or prevent the development of disease complications [3].

The economic burden of T2DM is substantial [4–6], cost-effectively allocating scarce resources among competing resources is challenged not only by the limited time and resources available to economic stakeholders in general but also by an unusual degree of decision-making uncertainty (e.g., clinical and economic implications that

Key Points

Efficiently allocating scarce resources for chronic and progressive diseases such as type 2 diabetes mellitus (T2DM) is challenged by limited time and resources and an unusual degree of decision-making uncertainty (e.g., clinical and economic implications that extend far beyond trial durations, patient heterogeneity, evolving practice patterns, and practice patterns that differ between trials and ordinary use).

To extrapolate trial data to longer decision-making time horizons, economic modeling is routinely used. While economic models of T2DM would ideally be user friendly, transparent, fast, and accurate (i.e., good external validity), the complexity of T2DM generally requires comprehensive (including parallel sets of complications and sophisticated treatment-switching algorithms) to ensure good predictive accuracy. Established T2DM models are generally slow and relatively opaque, which imposes an additional demand on economic stakeholders for case-specific expertise to evaluate the suitability of manufacturer-submitted models and in some cases to run the models with tight deadlines.

To address a need that some economic stakeholders have for greater user friendliness and faster run times, the IHE Diabetes Cohort Model was constructed using the cohort rather than the micro-simulation approach. A well-known limitation of cohort modeling, however, is an inability to adequately model patient heterogeneity (at least not without a health state explosion) and a potential for biased cost-effectiveness estimates.

In exercises designed to evaluate the potential magnitude of bias of the IHE Diabetes Cohort Model, we compared results generated for a set of simulation scenarios with those of a micro-simulation model (Economic and Health Outcomes Model of T2DM), chosen because the structures are otherwise generally similar and because it was possible to harmonize the models even more to minimize between-model simulation differences. We found systematic differences in simulated costs and quality-adjusted life-years, but little evidence of systematic differences in the incremental costs and quality-adjusted life-years that underlie cost-effectiveness metrics or in incremental cost-effectiveness ratios and net monetary benefits themselves.

extend far beyond trial durations, large number of interdependent micro- and macrovascular complications with multiple treatment targets, patient heterogeneity, evolving

practice patterns, and practice patterns that can differ widely between trials and ordinary use) [7].

The evidence used by economic stakeholders to make decisions is routinely generated using economic models that support extrapolation of trial data to time horizons sufficient to capture the full costs and benefits of intervention (often lifetime). A large number of economic models of T2DM are available [8]. Ideally, these models would be user friendly, transparent, fast, and accurate (i.e., good external validity). To obtain good predictive accuracy given the complexity of T2DM, however, these models must include a large set of interdependent micro- and macrovascular complications and sophisticated long-term treatment managers that challenge these goals.

Economic simulation models can generally be divided according to whether they represent the hypothetical patients as unique individuals (micro-simulation) or as average members of a representative cohort (cohort modeling) [9, 10]. Both approaches have well-known advantages and disadvantages. Micro-simulation models can accommodate patient heterogeneity and interdependent health states while maintaining a compact form because individual hypothetical patients can be assigned and carry with them a large amount of personal information, which enables simulation of personalized treatment pathways and event risks and realistic patient histories [9, 10]. The primary disadvantages of the micro-simulation approach in T2DM are a lengthy model code (often in high-level programming language rather than the more accessible Microsoft Excel® [Microsoft, Seattle, WA, USA]), computational intensiveness [9, 11], and an additional demand on the economic stakeholder for case-specific disease and programming expertise to evaluate the suitability of manufacturer-submitted models. Indeed, the code underlying most current models of T2DM is generally impenetrable to most non-programmers and run times (numbering in hours and sometimes days) can be limiting. The Canadian Agency for Drugs and Technologies in Health (CADTH), for example, has announced pending updates to its Category 1 Requirements that include model run times for the base-case analysis and key scenario analyses of less than 1 business day and programming in Microsoft Excel® [12].

Cohort models can approximate a micro-simulation model if the disease is discretized into enough health states, but "state explosion" and the paradoxical possibility the model is less manageable and transparent than a corresponding micro-simulation model [10], thus the micro-simulation approach has been widely used for T2DM [8, 13, 14]. Pragmatic cohort models can be constructed without a "state explosion", however, even for complicated diseases without a complete sacrifice of predictive accuracy. The IHE Diabetes Cohort Model (IHE-DCM) [15] was designed and

constructed in Microsoft Excel® (Microsoft) with this in mind to address reasonable concern about a lack of transparency in micro-simulation models and has demonstrated external validity on par with other micro-simulation models of T2DM [15]. The benefits include fewer parameters, faster run times, and convenient use of Microsoft Excel®, all of which can be appealing to stakeholders tasked with understanding (and potentially running) the models under time pressure [16]. The primary disadvantage of the cohort modeling approach is the potential for biased estimates of the incremental cost-effectiveness ratio (ICER), which arises when there is "uncaptured" patient heterogeneity that forces the cohort approach to simulate non-linear relationships with average patient characteristics [16]. To manage the large number of parallel health states, Visual Basic for Applications was used to program key model functions as macros, thus sacrificing some of the potential gains in transparency.

The IHE-DCM has been used to estimate long-term cost consequences of T2DM in Sweden [17], to estimate the cost-effectiveness of anti-hyperglycemic treatments [18–22], and to support HTA submissions in Sweden, Norway, and Canada [23–25]. Given the possibility that the cohort modeling approach produces biased estimates for complex diseases like T2DM, stakeholders can benefit from an empirical investigation of the likely magnitude and direction (i.e., the potential penalty to be traded against the other benefits). Indeed, CADTH conjectured that "there may be a significant degree of bias …" involved in a recent application using IHE-DCM, owing in part to the model design (including absence of patient variability and the non-linear relationship between biomarkers and outcomes) [25]. The Norwegian Medicines Agency had similar reservations about the cohort approach, though they concluded that IHE-DCM was appropriate given shorter run times and greater transparency [23].

2 Objective

The objective of this study was to inform decision makers by investigating the direction and magnitude of bias of IHE-DCM cost-effectiveness estimates attributable to using the cohort modeling approach.

3 Methods

We borrowed well-established cross-validation tools [11, 13, 26] to examine whether cost-effectiveness estimates generated by IHE-DCM are tangibly biased by comparing IHE-DCM results from a set of scenarios inspired by the 9th Mount Hood Diabetes Challenge with corresponding results produced by an otherwise similar micro-simulation model—the Economic and Health Outcomes Model of T2DM

(ECHO-T2DM). Similar analyses have been performed previously for other diseases, including chronic obstructive pulmonary disease [27], human immunodeficiency virus [28], and hepatitis C [29]. While such an exercise cannot provide a definitive (and universal) answer to concerns about possible bias, and it does not address the academic discussion of how much accuracy is reasonable to swap for increased transparency [30], it can provide a careful examination of how two otherwise similar models respond to the same stimuli (both absolutely and incrementally) and thus inform stakeholders charged with interpreting evidence generated by IHE-DCM.

3.1 The Models

IHE-DCM uses the cohort approach to model the cost-effectiveness of competing treatment alternatives for representative hypothetical patients with T2DM [18–22]. It is constructed with Markov health states representing important microvascular complications (retinopathy, neuropathy, and nephropathy) and macrovascular complications (myocardial infarction, ischemic heart disease, heart failure, and stroke) and dead, updated in annual cycles. Microvascular event risks are sourced primarily from the National Institutes of Health model [31] and Bagust et al. [32]. Multiple sets of macrovascular and mortality event risks are supported in the model [33–36], of which the UK Prospective Diabetes Study Outcomes Model 2 equations [36] were used in this exercise. Treatment effects are applied as changes in biomarkers (applied during the first year of treatment) and biomarker evolution is simulated until the predefined time horizon is reached. Treatment algorithms allow for treatment intensification when glycemic goals are not met. Unit costs and quality-adjusted life-year (QALY) disutility weights are applied based on health outcomes. The simulation time horizon is user defined and the probabilistic sensitivity analysis (PSA) is supported for treatment effects, risk coefficients, biomarker drifts, adverse event rates, unit costs, and QALYs. A more complete description can be found in the Electronic Supplementary Material (ESM). IHE-DCM performed in line with other micro-simulation models in internal and external validation exercises covering 12 long-term clinical studies, though there was a tendency to overestimate the macrovascular outcomes [15]. Model validity has been described formally using the Assessment of the Validation Status of Health-Economic decision modeling tool [37] (see the ESM).

ECHO-T2DM was chosen as the micro-simulation model because it has a similar (albeit not identical) structure (e.g., health states, biomarkers, risk predictions, as well as outcomes) and model features (e.g., treatment intensification following poor glycemic control), an ability to simulate common risk equations (both models support multiple sets), and flexibility. Furthermore, as both models were available to

the study authors, the models could be modified to further improve standardization and reduce noise attributable to factors other than the modeling approach (something not possible when cross-validating against previously published results in the literature). ECHO-T2DM is validated [38, 39] and has participated in the 5th through 9th Mount Hood Diabetes Network Challenges [8, 13, 26]. A full description can be found in the ESM and tests of its validity are described using the Assessment of the Validation Status of Health-Economic decision modeling tool [37] (see the ESM).

The main differences in the models and the steps taken to harmonize them are presented in Table 1.[1] Briefly, we harmonized the model structures used in this exercise by: (1) selecting the same sets of macrovascular and mortality risk prediction equations (UKPDS 82) [36], (2) simplifying the ECHO-T2DM insulin treatment algorithm to duplicate the simpler regimen supported by IHE-DCM, and (3) aligning diverse inputs such as microvascular risk elasticities with glycosylated hemoglobin (HbA$_{1c}$) and systolic blood pressure and drifts of clinical biomarkers. However, the models simulate end-stage renal disease risk and estimated glomerular filtration rate (eGFR) progression differently, which could not be resolved directly, thus eGFR progression in IHE-DCM was loaded as closely as possible to ECHO-T2DM. Health states for kidney disease and foot ulcer also differed, which was handled by disabling the cost and QALY consequences for micro-and macroalbuminuria in IHE-DCM and for chronic kidney disease (CKD) stages as well as foot ulcer in ECHO-T2DM. Because these standardizations entail that the simulated versions of the models are somewhat artificial, a sensitivity analysis was performed using the models "as intended" (i.e., not harmonized).

3.2 Reference Case

A set of simulation scenarios was designed with inspiration from the "Reference Case" simulation developed for the 9th Mount Hood Diabetes Challenge Network (convened in Dusseldorf, Germany in 2018) [40] and based loosely on the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) trial [41]. The Mount Hood Diabetes Challenge Network Reference Case was chosen as it is well known in diabetes modeling circles and permits comparison with publicly available results for 11 other models of diabetes [8]. In a first step, the Reference Case was simulated exactly as per the Challenge instructions [42], which

importantly extends the reach of this analysis by supporting comparison with 11 different diabetes models that have uploaded results to the online Mount Hood Diabetes Network Registry [8] (because of the harmonization, the results reported here for ECHO-T2DM differ slightly from those online).

Baseline patient characteristics were sourced from the Challenge instructions and, as necessary, from ADVANCE trial publications (see Table 2). Quality-adjusted life-year disutility weights were sourced entirely from the Challenge instructions (see Table 1 of the ESM). The Mount Hood Challenge simulation consisted of a control arm compared with five hypothetical treatment profiles, the first four of which considered changes in individual biomarkers one at a time and the last of which included the combined set of biomarker changes. For this application, we simulated the combined set of biomarker changes (see Table 3). As per the Challenge instructions, male and female individuals are simulated separately (though baseline characteristics were otherwise identical), biomarkers were kept constant over time, and the simulation time horizon was 40 years. We supplemented the Reference Case by including a vector of unit costs reflecting the Canadian treatment setting (see Table 1 in the ESM), which enabled consideration of cost-effectiveness metrics. Fictional, but not unreasonable, annual costs were applied for the control and intervention arms (CAN$1000 vs CAN$2500). A porobabilistic sensitivity analysis was used in the base case for both models, which is consistent both with micro-simulation modeling and with ordinary use of IHE-DCM (though it may differ from common practice with cohort modeling in general). Preliminary simulations found that cost-effectiveness metrics stabilized at or well before 500 cohorts (with 1000 individuals per cohort for ECHO-T2DM), ICER for IHE-DCM, and net monetary benefits (NMB) for ECHO-T2DM based on model functionalities. Conservatively, 1000 cohorts (and 2000 individuals per cohort for ECHO-T2DM) were chosen (see Fig. 7 in the ESM).

3.3 Expanded Reference Case

The restriction of homogeneous patients at baseline (and the absence of biomarker evolution and rescue medication) in the Reference Case artificially limits a key difference between cohort and micro-simulation modeling and limits generalizability of the exercise. Inspired by the Mount Hood Reference Case, we created a more realistic simulation scenario that captures patient heterogeneity, natural evolution of biomarkers, and treatment intensification. We also added biomarker treatment effects for HbA$_{1c}$ and eGFR to the control arm (see Table 3). Because cost-effectiveness is rarely estimated separately for male and female individuals in T2DM, the sexes were pooled. Treatment

---

[1] This differs from a conventional cross-validation analysis, in which the models are applied without adjustment to the same decision-making problem. In this exercise, the models were adjusted to remove structural differences to better isolate differences attributable to the unit of representation (i.e., cohort vs patient).

Table 1 Key model differences (IHE Diabetes Cohort Model [IHE-DCM] vs Economic and Health Outcomes Model of T2DM [ECHO-T2DM] and method of standardization)

Type	IHE-DCM	ECHO-T2DM	Method of standardization
Unit of observation	Cohort model	Microsimulation	N/A
Cycle length	1 year	1 year, but foot submodule operates in monthly cycles	N/A
Health states			
Macrovascular	IHD, MI, stroke, CHF	IHD, MI, stroke, CHF	N/A
Neuropathy	Symptomatic neuropathy, PVD, LEA	symptomatic neuropathy, PVD, foot ulcer, LEA	Exclude costs and QALYs for foot ulcer
Retinopathy	BDR, PDR, ME, blindness	BDR, PDR, ME, blindness	
Kidney disease	Micro- and macroalbuminuria, ESRD	CKD stages 1–5, ESRD	Exclude costs and QALYs for micro-and macroalbuminuria as well as for CKD stages 1–5
Risk predictions			
Macrovascular	UKPDS 68 [35] and 82 [36], NDR (Kiadaliri et al. [34]), FDS [33]	UKPDS 68 [35] and 82 [36], NDR (Kiadaliri et al. [34] and Zethelius et al. [44]), ADVANCE [45] Second event for MI and stroke in UKPDS 82 [36] evaluated within simulation only	UKPDS 82 [36] is used in this application
Neuropathy	Transition probabilities from Eastman et al. [31] and Bagust et al. [32] with HbA_{1c} power function from Eastman et al. [31]	Transition probabilities from Eastman et al. [31] and Bagust et al. [32] with HbA_{1c} power function from Eastman et al. [31]	N/A
Retinopathy	Transition probabilities from Eastman et al. [31] with a power function for HbA_{1c} [31]	Transition probabilities from Eastman et al. [31] with a power function for HbA_{1c} and SBP	Deactivate the link between SBP and PDR in ECHO-T2DM
Kidney disease	Transition probabilities from Eastman et al. [31] with a power function for HbA_{1c}	Transition probabilities for micro-and macroalbuminuria from UKPDS 64 [46] with a power function for HbA_{1c}. Risk of microalbuminuria adjusted with NHANES data for eGFR level. ESRD directly linked with eGFR decline modeled in line with the CDC CKD model [47]	Apply linear decline of eGFR in IHE-DCM estimated from the CDC CKD model [47] to approximate eGFR decline in ECHO-T2DM
Mortality	UKPDS 68 [35] and UKPDS 82 [36]	UKPDS 68 [35] and UKPDS 82 [36], BRAVO	UKPDS 82 [36] used for both models
Baseline patient characteristics	Includes mean values for a wide range of demographics, biomarkers, and clinical history inputs	Includes mean (SD) values for a wide range of demographics, biomarkers, and clinical history inputs with an option to include correlation between biomarkers	SD included in ECHO-T2DM but baseline biomarkers were assumed to be uncorrelated
Efficacy data and AEs for intervention and comparator	First year change applied at initiation of the treatment for wide range of biomarkers AE risk applied as an annual event rate	First year change applied at initiation of the treatment for wide range of biomarkers AE risk applied as an annual event rate	N/A
Insulin rescue medication	Applied in same way as intervention and comparator treatment	Supports insulin rescue with a fixed dose working in a similar way as the treatment profile for intervention and comparator. Supports flexible dose with adjusted effects	Use fixed dose insulin rescue in ECHO-T2DM

Table 1 (continued)

Type	IHE-DCM	ECHO-T2DM	Method of standardization
Biomarker value progression	Annual linear change	Annual linear change supported for all biomarkers, additionally supports the non-linear evolution equations for HbA$_{1c}$ change sourced from UKPDS 68. eGFR decline dependent on current eGFR level and albuminuria status based on the CDC CKD model	Apply linear change for HbA$_{1c}$ and estimate a linear decline for IHE-DCM based on the CDC CKD model [47]
Other disease management (ACE inhibitors, statins)	Supports anti-hypertensive, anti-dyslipidemia, and weight management drugs	Supports anti-hypertensive, anti-dyslipidemia, and weight management drugs with associated RRs for events (e.g., stroke) associated with use of ACE inhibitors and statins	No treatment for other disease management is included
Costs and QALYs			
Treatment costs	Annual cost applied for all treatment arms	Annual cost applied for all treatment arms	N/A
Unit costs for complications	Apply event cost in year of the event and an annual cost for each event. Possible to assign separately event costs for first and second MI, stroke, and LEA	Apply first-year cost and subsequent annual follow-up cost. Support possibility to assign non-fatal and fatal event costs separately	Assign same unit costs for fatal, non-fatal, first and second events
QoL	Disutility applied for current health state(s) and demographics deducted from base utility level. Support disutility for first and subsequent events for MI and stroke	Disutility approach applied with disutility for current health state(s) and demographics deducted from base utility level	Assign same disutility for first and second events in IHE-DCM

ACE angiotensin-converting enzyme, AEs adverse events, BDR background diabetic retinopathy, BRAVO building relating assessing and validating outcomes, CDC Center of Disease Control, CHF congestive heart failure, CKD chronic kidney disease, eGFR estimated glomerular filtration rate, ESRD end-stage renal disease, FDS Freemantle Diabetes Study, IHD ischemic heart disease, LEA lower extremity amputation, ME macular edema, MI myocardial infarction, N/A not applicable, NDR National Diabetes Registry, PDR proliferative diabetic retinopathy, PVD peripheral vascular disease, QALY quality-adjusted life-years, QoL quality of life, RR relative risk, SBP systolic blood pressure, SD standard deviation, UKPDS UK Diabetes Protective Diabetes Study

Table 2 Baseline patient characteristics

Parameter	Reference case (male and female individuals separately)	Expanded Reference case		Early disease (hypothetical)		Late disease (hypothetical)	
	Mean/%	Mean/%	SD	Mean/%	SD	Mean/%	SD
Demographics							
Age (years)	66.00	66.00	6.00	**50.00**	**6.00**	**75.00**	**6.00**
Male (%)	–	50.0	–	50.0	–	50.0	–
Disease duration (years)[a]	8.00	8.00	–	**2.50**	–	**15.00**	–
Ethnicity/race (%)							
African Americans	1.9	1.9	–	1.9	–	1.9	–
Caucasian	98.1	0.0	–	0.0	–	0.0	–
Clinical indicators							
Atrial fibrillation (%)	0.0	0.0	–	0.0	–	0.0	–
Smokers (%)	0.0	0.0	–	0.0	–	0.0	–
HbA$_{1c}$ (%)	7.50	8.00	1.50	8.00	1.50	8.00	1.50
SBP (mmHg)	145.00	145.00	22.00	**135.00**	**22.00**	**155.00**	**22.00**
BMI (kg/m^2)	28.00	28.00	5.00	**27.00**	**5.00**	**30.00**	**5.00**
WBC (*10^6)	7.00	7.00	1.90	7.00	1.90	7.00	1.90
HR (beat/minute)	79.00	79.00	12.00	79.00	12.00	79.00	12.00
Total cholesterol (mg/dL)[b]	200.77	200.77	1.20	**190.00**	**46.33**	**210.00**	**46.33**
LDL cholesterol (mg/dL)[b]	115.83	115.83	39.77	**110.00**	**39.77**	**130.00**	**39.77**
HDL cholesterol (mg/dL)[b]	50.19	50.19	15.44	50.19	15.44	50.19	15.44
Triglycerides (mg/dL)[c]	176.99	176.99	1.50	176.99	132.74	176.99	132.74
eGFR (mL/min/1.73 m^2)	70.00	70.00	15.00	**80.00**	**15.00**	**60.00**	**15.00**
Co-morbidities (%)							
Microalbuminuria	–	25.6	–	**10.3**	–	**30.0**	–
Macroalbuminuria	–	3.6	–	**1.9**	–	**20.0**	–
IHD (not including MI)	–	6.1	–	**0.0**	–	**12.0**	–
MI	–	12.0	–	**2.0**	–	**25.0**	–
CHF	–	6.1	–	**0.0**	–	**10.0**	–
Stroke	–	9.2	–	**1.4**	–	**20.0**	–

Bolded values for early disease and late disease cohorts indicate changes from expanded reference case

BMI body mass index, *CHF* congestive heart failure, *eGFR* estimated glomerular filtration rate, *HbA$_{1c}$* glycosylated hemoglobin, *HDL* high-density lipoprotein, *HR* heart rate, *IHD* ischemic heart disease, *LDL* low-density lipoprotein, *MI* myocardial infarction, *SBP* systolic blood pressure, *SD* standard deviation, *WBC* white blood cell

[a]Entered as minimum/maximum in the Economic and Health Outcomes Model of T2DM

[b]Entered as mmol/L in the IHE Diabetes Cohort Model

intensification starting with basal insulin and followed by a basal and bolus insulin regimen was applied when HbA$_{1c}$ was ≥ 8% (see Table 2 in the ESM). Note: these results are not comparable to those stored in the Mount Hood Diabetes Network Registry [8].

In addition to the base case, 18 additional scenarios were created and simulated to evaluate whether systematic differences between the models (and modeling approaches) could be identified and, if so, which model features drive them. The scenarios are presented in Table 4 and can broadly be sub-divided into tests of the treatment algorithm, the importance of PSA, economic parameters (i.e., costs of treatment and QALY disutility weights), different patient sub-groups

(male vs. female individuals, early disease, and late disease), and differences in the CKD sub-model. Baseline patient characteristics for early and late disease are presented in Table 2. As these scenarios are each based on model harmonization to minimize between-model differences unrelated to the cohort vs micro-simulation approaches, we also simulated a less artificial scenario in which the models were simulated as intended.

3.4 Analysis

We compared estimated model outcomes (including costs, QALYs, and ICERs and NMBs defined based on QALYs

Table 3 Treatment profiles

	Reference Case		Expanded Reference Case			
	Intervention	Control	Intervention		Control	
Treatment	Mean	Mean	Mean	SE	Mean	SE
Mean (SE) change from baseline						
HbA$_{1c}$, %	− 1	0	− 1	0.05	− 0.5	0.05
SBP, mmHg	− 10	0	− 10	1.5	0	0
LDL cholesterol, mg/dL	− 19.305	0	− 19.305	1.5	0	0
BMI, kg/m^2	− 1	0	− 1	0.05	0	0
eGFR	0	0	0	0	− 5	1.000
Rates of adverse events (per P-Y of exposure):						
Non-severe hypoglycemia	0	0	0.5	0	1	0
Severe hypoglycemia	0	0	0	0	0	0
Corresponding[a] HbA$_{1c}$, %	7.5	7.5	7.5	0	7.5	0
Drifts						
HbA$_{1c}$, %	0	0	0.14	0	0.14	0
SBP, mmHg	0	0	0.3	0	0.3	0
Total cholesterol, mg/dL	0	0	0	0	0	0
LDL cholesterol, mg/dL	0	0	0	0	0	0
HDL cholesterol, mg/dL	0	0	0	0	0	0
Triglycerides, mg/dL	0	0	0	0	0	0
BMI, kg/m^2	0	0	0	0	0	0
eGFR (ECHO-T2DM)	0	0	Estimated from CDC-CKD model [47, 48][b]		Estimated from CDC model [47, 48][b]	
eGFR (IHE-DCM)	0	0	− 2.43	0	− 2.43	0
Treatment costs (CAN$)						
Drug	2500	1000	2500	0	1000	0

BMI body mass index, *CDC* Center of Disease Control, *CHF* congestive heart failure, *CKD* chronic kidney disease, *ECHO-T2DM* Economic and Health Outcomes Model of T2DM, *eGFR* estimated glomerular filtration rate, *HbA$_{1c}$* glycosylated hemoglobin, *HDL* high-density lipoprotein, *IHD* ischemic heart disease, *IHE-DCM* IHE Diabetes Cohort Model, *LDL* low-density lipoprotein, *MI* myocardial infarction, *P-Y* patient-year, *SBP* systolic blood pressure, *SE* standard error

[a]Hypoglycemic event rates increase or decrease inversely with HbA$_{1c}$, thus hypoglycemic event rates are adjusted from mean HbA$_{1c}$ in the trial from which event rates are sourced to match the HbA$_{1c}$ of the patients being simulated in both models. See also the ESM

[b]Annual decline for patients with eGFR > 60 mL/min/1.73 m^2 − 1.1, patients with eGFR > 60 and macroalbuminuria − 4.1, patients with eGFR < 60 − 2.8, patients with eGFR < 60 and macroalbuminuria − 5.2

gained) under the maintained assumption that systematic differences can largely be attributed to the modeling approach (cohort vs micro-simulation) given our attempts to otherwise harmonize the models and input parameters. Numerical differences between models were calculated and assessed, for costs and QALYs at both the absolute and incremental levels. Mean differences were calculated across the base case and all scenarios in the Expanded Reference Case. Because harmonization was incomplete, however, some noise will inevitably enter, thus we assessed concordance statistically using three different methods (for the Reference Case, only visual assessment was performed):

1. We plotted the mean and 95% confidence intervals for incremental costs, incremental QALYs, and NMB estimated for both models for the base case and for the 18 scenario analyses. The proportion of point estimates for each model falling within the 95% confidence interval produced by the other model was generated for each outcome (ICERs were excluded because more than 2.5% of replications produced negative values). Ninety percent was considered a threshold for concordance.

2. At an anonymous reviewer's suggestion, we conducted a formal hypothesis test for costs, QALYs, and NMB using the paired *t* test with a null hypothesis of concordance (significance level of 5%). We performed the test for ICERs as well because they are of considerable interest to decision makers, but one of the scenarios had to be omitted because it produced cost savings and QALY gains (i.e., a negative ICER). To ensure that violation of normality does not invalidate the results, the non-parametric Wilcoxon Signed Rank Test was also

Table 4 Simulation scenarios: expanded reference case

#	Description	Change in parameter(s)	Base case
Base case	Expanded MH reference Challenge	–	–
1	Higher HbA$_{1c}$ target	HbA$_{1c}$ target of 8.5%	HbA$_{1c}$ target 8.0%
2	Higher HbA$_{1c}$ drift for comparator	HbA$_{1c}$ drift higher for comparator (0.24%)	0.14% for both arms
3	Treatment intensification	Intensification after year 3 (both arms)	HbA$_{1c}$ target 8.0%
4	Model applied "as intended"	Flexible insulin dose (ECHO-T2DM)	Fixed insulin dose (both models)
		All unit costs/QALYs (both models)	Zero costs/QALYs for foot ulcer, CKD stages, and micro- and macroalbuminuria
5	No PSA (second-order uncertainty)	PSA inactive for both models	PSA active
6	Early disease patients	More favorable age, disease duration, biomarkers (SBP, TC, LDL, BMI, eGFR), and micro- and macrovascular disease histories at baseline	ADVANCE baseline patient characteristics [50]
7	Late disease patients	Less favorable age; disease duration; SBP, TC, LDL, BMI, and eGFR biomarkers; and more patients with micro- and macrovascular disease at baseline	ADVANCE baseline patient characteristics [50]
8	Alternative QALY decrements for key outcomes	QALY weights sourced from CODE-2 [51]: MI (-0.028), stroke (-0.115), CHF (-0.028), IHD (-0.028), ESRD (-0.175), and LEA (-0.272)	CADTH recommended disutility weights
9	Treatment effects	HbA$_{1c}$ lowering only: intervention (-1%) and comparator (-0.5%)	Treatment effects for SBP, LDL, BMI, and eGFR
10	Rebound of treatment effects following discontinuation	Full rebound of TE following discontinuation	No rebound in HbA$_{1c}$
11	Discontinuation of treatments	No discontinuation of interventions following intensification to insulin	Discontinuation of interventions following intensification to insulin
12	Lower insulin cost	50% lower (basal CAN$500/year and bolus CAN$800/year)	Cost of fixed-dose basal: CAN$1000/year; Cost of fixed-dose bolus: CAN$1600/year
13	Higher treatment costs (intervention)	50% higher (CAN$3750)	CAN$2500 for intervention in BC
14	Lower treatment costs (intervention)	50% lower (CAN$1250)	CAN$2500 for intervention in BC
15	No modeling of CKD	eGFR kept constant over time	eGFR decline in line with the CDC-CKD model [47]
16	Deterministic cohort model	PSA inactive for IHE-DCM	PSA active for IHE-DCM
17	Female subgroup	100% female	50% female/50% male
18	Male subgroup	100% male	50% female/50% male

BC, BMI body mass index, *CDC* Center for Disease Control, *CHF* congestive heart failure, *CKD* chronic kidney disease, *ECHO-T2DM* Economic and Health Outcomes Model of T2DM, *eGFR* estimated glomerular filtration rate, *ESRD* end-stage renal disease, *HbA$_{1c}$* glycosylated hemoglobin, *IHE-DCM* IHE Diabetes Cohort Model, *IHD* ischemic heart disease, *LDL* low-density lipoprotein, *LEA* lower-extremity amputation, *MH* Mt Hood, *PSA* probabilistic sensitivity analysis, *MI* myocardial infarction, *QALY* quality-adjusted life-year, *SBP* systolic blood pressure, *TC* total cholesterol, *TE* treatment effects

performed. Because the results of modeling different simulation scenarios are not akin to independent draws from a population (i.e., there is considerable dependence), this test is over-powered and thus too likely to reject the null hypothesis of concordance.

3. At the same reviewer's suggestion, we also performed a test loosely based on methods proposed by Corro Ramos and colleagues [43] in which we calculated the number of PSA iterations for each model for which the estimated NMB falls within the 95% confidence interval produced by the other model for the base case scenario (ICERs were excluded because 95% confidence intervals could not be generated). Note, the Corro Ramos et al. approach is designed to assess the validity of model estimates by comparing with clinical data rather than predictions from a different model.

Because important differences can be masked when looking only at the aggregate level, we also compared cumulative event incidences in the Expanded Reference Case for IHE-DCM and ECHO-T2DM (95% confidence intervals are not generated by IHE-DCM). Specifically, the proportion of the 14 IHE-DCM-predicted cumulative event incidence rates in the base case falling within the 95% confidence intervals for the corresponding ECHO-T2DM micro-simulation estimates was calculated. Ninety percent was considered a threshold for concordance. Biomarker evolution curves were examined to ensure that the simulations were properly implemented.

4 Results

4.1 Comparison of Model Implementation

Run times differed substantially by model. On a personal computer with a 16-GB random access memory and an I7-processor, run times for the base case analysis were approximately 45 min for IHE-DCM and 30 h for ECHO-T2DM. For the scenario analysis without PSA (i.e., running only one cohort), run times were less than 1 min for IHE-DCM and between 2 and 3 min for ECHO-T2DM. In part because there are more parameters in micro-simulation though also because ECHO-T2DM has more model features, the analysts (authors AN and AL) noted that loading and double checking ECHO-T2DM took longer than IHE-DCM.

4.2 Reference Case

Key results for the Reference Case are presented in Tables 3 and 4 of the ESM for male and female individuals, respectively. Estimated life-years predicted by IHE-DCM were approximately 1 year longer for male individuals and

0.6 years longer for female individuals for both treatment arms than for ECHO-T2DM, which is consistent with the larger predicted QALYs and total costs. The between-model differences were smaller at the incremental level. Incremental predicted life-years were 0.61 and 0.47 years for IHE-DCM vs 0.71 and 0.55 years for ECHO-T2DM, for male and female individuals, respectively. The between-model differences in incremental predicted QALYs were smaller by about half. Incremental predicted total costs differed by CAN$294 for male and CAN$462 for female individuals, which yielded ICERs (per QALY gained) of CAN$29,309 for IHE-DCM vs CAN$27,654 for ECHO-T2DM for male individuals and CAN$38,680 for IHE-DCM vs CAN$37,109 for ECHO-T2DM for female individuals. At a willingness-to-pay threshold of CAN$50,000, NMBs (based on QALYs gained) were $13,293 for IHE-DCM vs CAN$15,452 for ECHO-T2DM for male individuals and CAN$6,199 for IHE-DCM vs CAN$7518 for ECHO-T2DM for female individuals. The cumulative incidences for micro- and macrovascular complications are presented in Figs. 3 and 4 of the ESM. With the exception of kidney complications, IHE-DCM predictions fell well within the 95% confidence intervals.

4.3 Expanded Reference Case

Key results for the Expanded Reference Case are presented in Table 5. Predicted absolute life-years, QALYs, and total costs were (as with the Reference Case) larger for IHE-DCM for both treatment arms. Incremental (between-arm) differences were again smaller, though the between-model gap differences were larger than in the Reference Case (0.46 vs 0.60 life-years gained, 0.67 vs 0.72 QALYs gained, and net cost increases of CAN$3719 vs CAN$5098 for IHE-DCM and ECHO-T2DM, respectively). Uncertainty as indicated by 95% confidence intervals was similar for the two models for costs, but about twice as high for IHE-DCM for QALYs (with the difference largely attributable to hypoglycemia event rates). Estimated ICERs were CAN$5542 for IHE-DCM and CAN$7059 for ECHO-T2DM and NMBs were CAN$28,834 and CAN$31,009, respectively. While 95% confidence intervals could not be calculated for ICERs, the 95% confidence intervals for NMBs were also about twice as wide for IHE-DCM and the lower bound was below 0 (CAN$-5833).

Estimated survival curves were visually similar, though slightly higher for IHE-DCM (see Fig. 1). Estimated 40-year cumulative incidence rates for IHE-DCM fell within the 95% confidence intervals for ECHO-T2DM predictions for each outcome, though IHE-DCM generated generally lower estimates than ECHO-T2DM (see Fig. 2). These cumulative incidences are also presented in a scatterplot in Fig. 6 of the ESM, with the values for IHE-DCM on the horizontal axis

Table 5 Detailed cost-effectiveness estimates for the Expanded Reference Case, by model

	IHE-DCM			ECHO-T2DM		
	Intervention	Control	Difference	Intervention	Control	Difference
Cost drivers						
Treatment	31,473	24,186	7287	29,372	21,674	7698
Non-insulin AHA	–	–	–	15,040	4666	10,375
Insulin AHA	–	–	–	14,332	17,009	− 2677
MI	12,384	13,152	− 768	12,828	13,540	− 712
IHD	7391	7830	− 439	5524	5685	− 161
CHF	9228	9659	− 431	7343	7282	62
Stroke	14,349	16,454	− 2105	9391	10,025	− 634
PVD	–	–	–	183	165	18
Retinopathy	247	286	− 40	250	264	− 14
CKD	4051	4007	44	6890	8199	− 1309
Neuropathy	–	–	–	310	319	− 8
Amputation				3716	3449	267
Lower extremity disease[a]	5142	4972	171	4209	3932	277
Hypoglycemia	483	655	− 171	602	711	− 109
Total costs (95% CI)	84,266	80,547	3719 (− 2807; 10,245)	76,410	71,312	5098 (− 3162; 9694)
Disutility drivers						
MI	0.097	0.099	− 0.002	0.089	0.089	0.000
IHD	0.081	0.086	− 0.004	0.059	0.060	− 0.001
CHF	0.094	0.097	− 0.003	0.071	0.069	0.002
Stroke	0.086	0.095	− 0.009	0.077	0.078	− 0.001
PVD	–	–	–	0.090	0.081	0.009
Retinopathy	0.033	0.038	− 0.005	0.034	0.037	− 0.002
CKD	0.040	0.040	0.000	0.083	0.098	− 0.015
Neuropathy	–	–	–	0.025	0.025	− 0.001
Amputation event	–	–	–	0.050	0.046	0.004
Lower extremity disease	0.201	0.192	0.009	–	–	–
Hypoglycemia	0.642	0.807	− 0.165	0.598	0.740	− 0.142
Excess weight	0.202	0.268	− 0.066	0.266	0.283	− 0.017
Survival[b]			0.471			− 0.558
Total			0.671			0.722
Health outcomes (discounted)						
LYs (95% CI)	13.169	12.710	0.459 (0.096; 0.822)	11.150	10.546	0.604 (0.180; 1.033)
QALYs (95% CI)	10.966	10.295	0.671 (− 0.029; 1.371)	8.978	8.256	0.722 (0.359; 1.205)
Survival at end of year 40	0.8%	0.7%	0.2%	0.3%	0.2%	0.1%
Net monetary benefits			29,834 (− 5833; 65,501)			31,009 (14,562; 55,026)
Incremental cost per QALY gained			5542			7059

AE adverse event, *AHA* anti-hyperglycemic agent, *CHF* congestive heart failure, *CI* confidence interval, *CKD* chronic kidney disease, *ECHO-T2DM* Economic and Health Outcomes Model of T2DM, *IHD* ischemic heart disease, *IHE-DCM* IHE Diabetes Cohort Model, *LY* life-year, *MI* myocardial infarction, *PVD* peripheral vascular disease, *QALY* quality-adjusted life-year

[a]Note: this is the outcome measure for IHE-DCM. Individual components are calculated separately for ECHO-T2DM, but we calculated this *ex post* as PVD + amputation + neuropathy for comparison purposes

[b]For consistency with the other disutility measures, only the disutility associated with survival differences is reported (with the value of one intervention normalized to 0)

and for ECHO-T2DM on the vertical axis. Points along the 45-degree line indicate equality and the dotted lines plot the best-fitting linear regression lines.

A cost-effectiveness scatterplot plane is presented in Fig. 3, with each point representing incremental QALYs and costs for one of the 1000 cohort replicates for the two
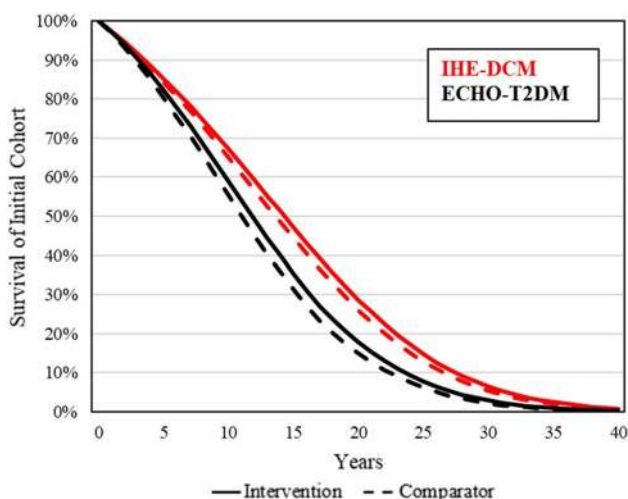
Fig. 1 Forty-year intervention and comparator survival for Expanded Reference Case, by model. *ECHO-T2DM* Economic and Health Outcomes Model of T2DM, *IHE-DCM* IHE Diabetes Cohort Model

models (IHE-DCM in black and ECHO-T2DM in red). Though uncertainty is larger for IHE-DCM, the scatterplots largely coincide. Cost-effectiveness acceptability curves

are largely similar as well (see Fig. 5 in the ESM). Both models predict a low probability of cost savings, but the predicted probabilities that the intervention is cost-effective are about 70% at a willingness-to-pay of CAN$10,000 per QALY gained rising to 96% for IHE-DCM and 100% for ECHO-T2DM at a willingness-to-pay of CAN$50,000 per QALY gained. The modified Corro Ramos et al. test found that estimated NMB for IHE-DCM fell within the 95% confidence interval generated by ECHO-T2DM for 72% of the PSA iterations. For ECHO-T2DM, estimated NMB fell within the 95% confidence interval generated by IHE-DCM for 98% of the replications.

Similarities at the aggregate level may mask some differences at the granular level. For example, IHE-DCM simulated greater cost offsets for avoided stroke and ischemic heart disease events than ECHO-T2DM, but ECHO-T2DM predicted cost offsets for CKD while IHE-DCM predicted a modest cost increase. Simulated biomarker evolution curves diverged over time, especially for HbA$_{1c}$ and body mass index where the start of rescue insulin medication occurred at the same time and induced stair step patterns in IHE-DCM, largely because of differential survival in the
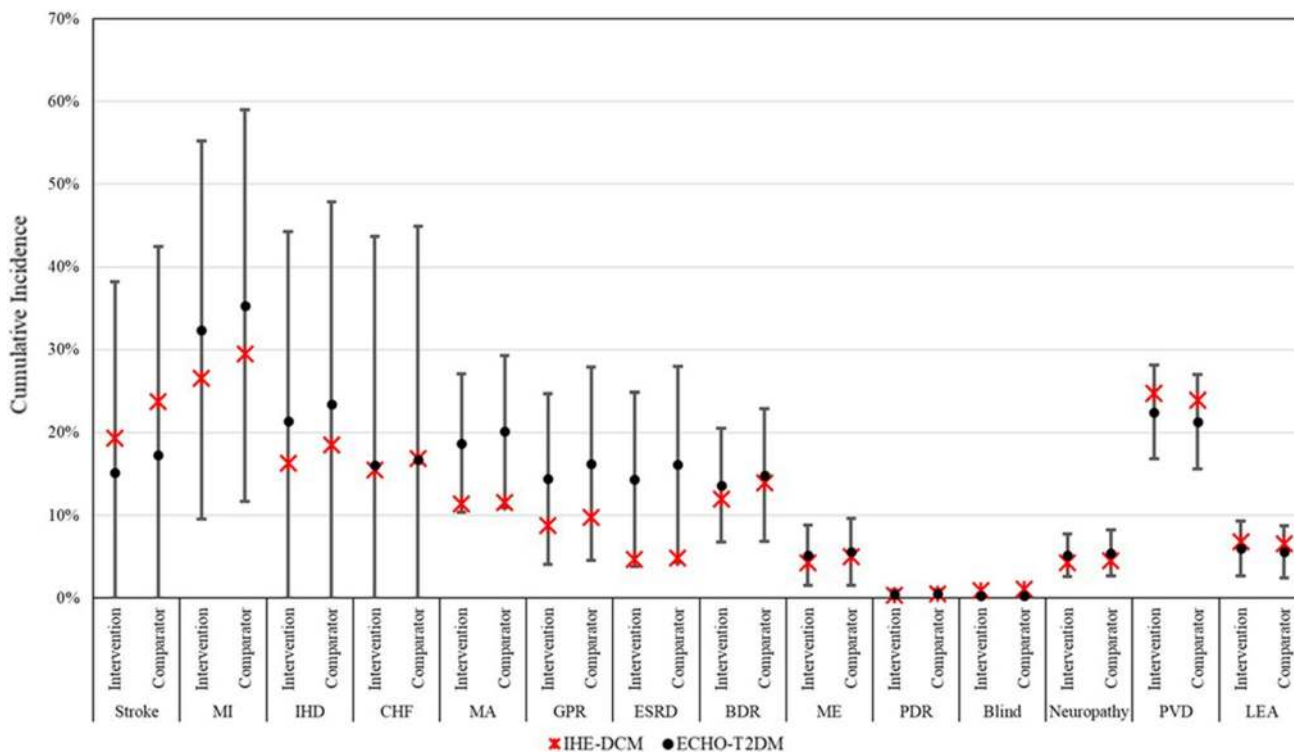
Fig. 2 Forty-year predicted cumulative incidence rates for Expanded Reference Case, by model. The *whiskers* represent the estimated 95% confidence interval of the cumulative incidence in the Economic and Health Outcomes Model of T2DM (ECHO-T2DM). *BDR* background diabetic retinopathy, *CHF* congestive heart failure, *ESRD* end-stage renal disease, *GPR* gross proteinuria (macroalbuminuria), *IHD* ischemic heart disease, *IHE-DCM* IHE Diabetes Cohort Model, *LEA* lower extremity amputation, *MA* microalbuminuria, *ME* macular edema, *MI* myocardial infarction, *PDR* proliferative diabetic retinopathy, *PVD* peripheral vascular disease
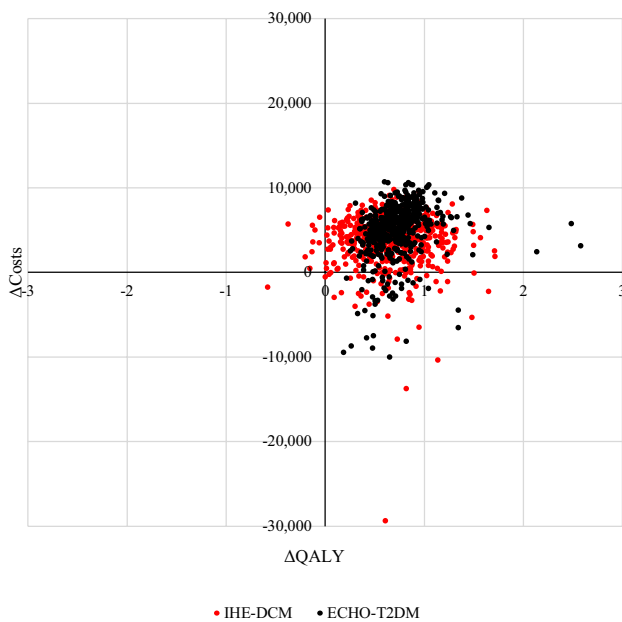
Fig. 3 Cost-effectiveness plane for Expanded Reference Case, by model (one sample point per cohort replication). *ECHO-T2DM* Economic and Health Outcomes Model of T2DM, *IHE-DCM* IHE Diabetes Cohort Model, *QALY* quality-adjusted life-year

heterogeneous ECHO-T2DM simulated population (see Fig. 8 of the ESM).

Results of the scenario analysis demonstrated that the two models changed in predictable (and mostly similar) ways to the parameter changes. IHE-DCM produced consistently greater life-years, QALYs, and absolute costs for both treatment arms than ECHO-T2DM (summary results are presented in Table 5 in the ESM) and IHE-DCM also generated consistently lower mean incremental costs, QALYs, and NMBs. Mean ICER for the base case and the 18 scenarios (excluding one scenario for which intervention was dominant for both models) were CAN$10,299 for IHE-DCM and CAN$10,417 for ECHO-T2DM. IHE-DCM generated a lower ICER in ten of the 18 cases (with well-behaved ICERs).

Individually, the results of the scenarios were generally predictable and robust. Sub-group analysis was notable, for example, only because the early disease cohort was associated with a noticeable change in incremental costs (especially for IHE-DCM). This affected predicted ICERs in relative terms, though the effect was less for the NMB (CAN$41,300 for IHE-DCM vs CAN$43,411 for ECHO-T2DM). The results were most affected by assumptions about CKD, where structural differences could be least standardized. Keeping eGFR constant over time increased the ICERs for both models compared with the base case, with between-model differences driven largely by changes

in incremental costs. Using the model "as intended" (rather than standardized) had limited impact on the results.

Mean and 95% confidence intervals (note, only for scenarios with PSA activated) for incremental costs, incremental QALYs, and the NMB are plotted in Fig. 4. Neither model had a mean value that fell outside of the 95% confidence interval for the other model in the base case or any of the 18 scenarios. Paired t tests uniformly rejected the null hypothesis of between-model equality of the absolute costs ($p < 0.001$) and QALYs ($p < 0.001$), incremental costs ($p < 0.001$) and QALYs ($p < 0.001$), and the NMB ($p < 0.009$). For the scenarios with well-behaved ICER estimates, however, the t test failed to reject between-model equality ($p < 0.68$).

5 Discussion

Using well-established cross-validation tools [11] modified to allow structural standardization of the models, we examined whether IHE-DCM produces systematically biased estimates of cost-effectiveness related to the cohort approach. In a simple Reference Case performed to enable comparison with the results of 11 other models that participated in the 9th Mount Hood Diabetes Challenge, IHE-DCM produced consistently greater absolute survival, QALYs, and costs than ECHO-T2DM, which is consistent with the difference between modeling homogenous patients and heterogeneous patients when event risks are non-linear (specifically convex) in key parameters [16]. Between-model differences were generally small at the incremental level (i.e., different between the two comparator arms) used to construct cost-effectiveness metrics, however, and the ICER and NMB, which were also similar between models. As expected, IHE-DCM was considerably faster compared with ECHO-T2DM, with a run time of approximately 45 min compared with 30 h using ECHO-T2DM, an important aspect for many stakeholders under time constraints.

This same pattern was observed for the more realistic Expanded Reference Case and 18 scenario analyses, and both models responded to changes in model parameters similarly and predictably. This was supported statistically; incremental costs, incremental QALYs, and NMBs for each model fell uniformly within the 95% confidence interval generated by the other model. There was more uncertainty in the results of IHE-DCM, which was driven in large part by uncertainty in the parameter estimate for the hypoglycemia event rate (eliminating it roughly halved the confidence interval). The estimates of ECHO-T2DM falls within even half of the 95% confidence intervals generated by IHE-DCM. Estimates in the base case by IHE-DCM of the 40-year cumulative incidence of study outcomes, moreover, fell within the 95% confidence intervals generated by ECHO-T2DM. While the

(A) Incremental Costs
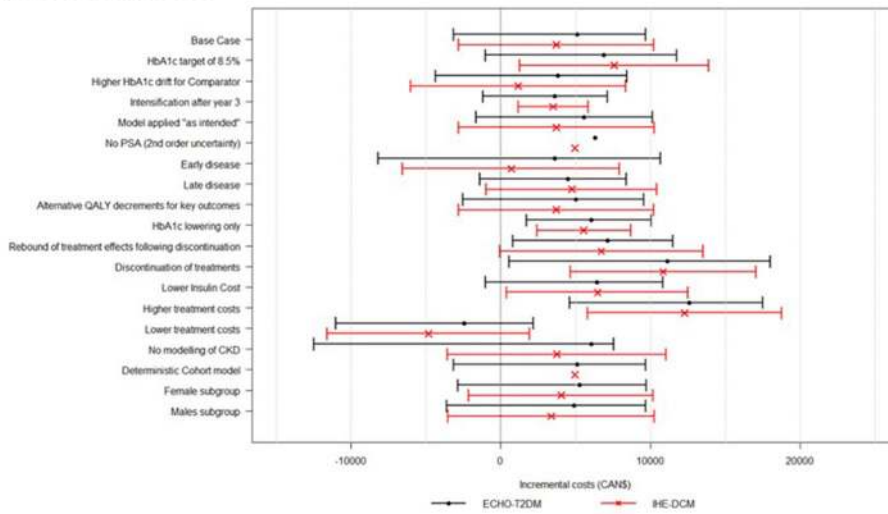
(B) Incremental QALYs
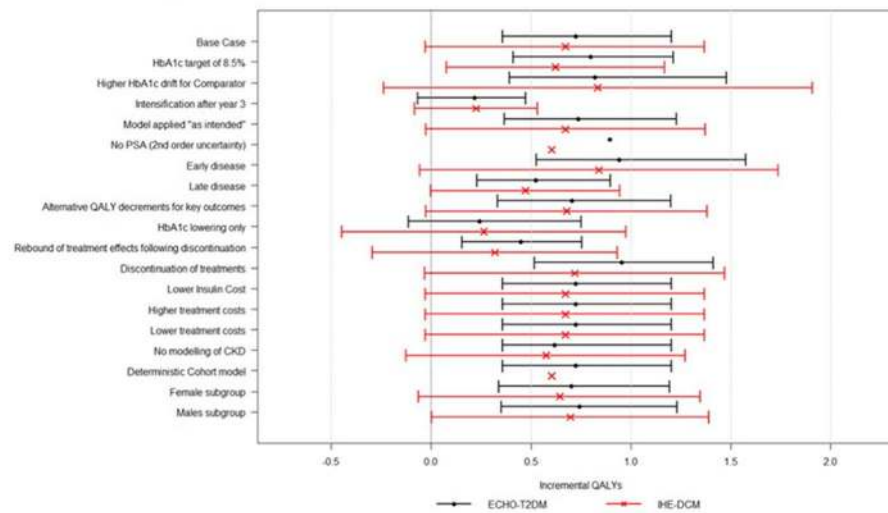
(C) Net Monetary Benefits Associated with Intervention (Willingness-to-Pay of CAN$50,000 per QALY)
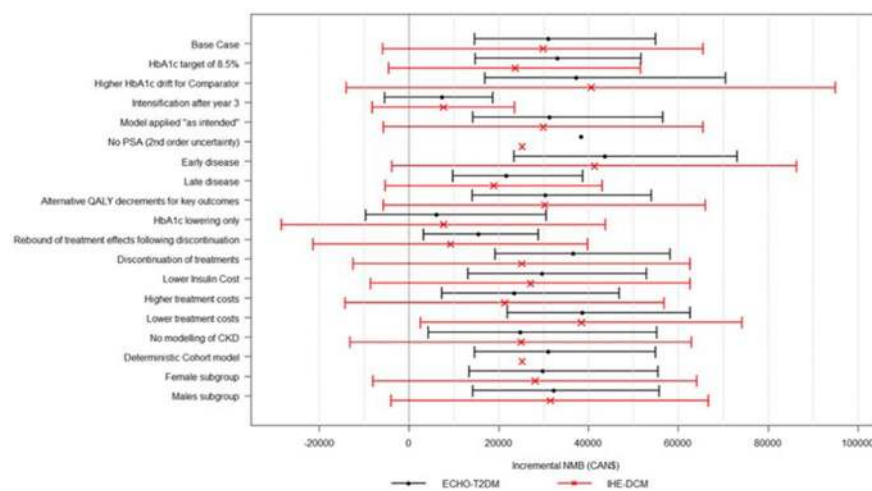
◄Fig. 4 Mean estimates using the IHE Diabetes Cohort Model (IHE-DCM) and Economic and Health Outcomes Model of T2DM (ECHO-T2DM) for Expanded Reference Case with 95% confidence interval for ECHO-T2DM, by scenario. **a** Incremental costs. **b** Incremental quality-adjusted life-years (QALYs). **c** Net monetary benefits associated with intervention (willingness to pay of CAN$50,000 per QALY). *CKD* chronic kidney disease, *HbA₁c* glycosylated hemoglobin, *PSA* probabilistic sensitivity analysis

paired *t* tests did find statistically significant between-model differences in incremental costs, incremental QALYs, and the NMB for these 19 scenarios, the paired *t* test is grossly overpowered to reject the null hypothesis in this setting as the simulation scenarios (i.e., the sample draws) are not independent of each other. Interestingly, however, the paired *t* test failed to reject between-model differences for the ICER ($p < 0.68$) for the 18 scenarios for which both incremental costs and incremental QALYs were positive (producing a meaningful ICER). Further underscoring this absence of clear bias in cost-effectiveness estimates, there was no discernible pattern as to which model produced more favorable cost-effectiveness estimates, with each more favorable in roughly half of the scenarios.

The trade-off between cohort modeling and micro-simulation is sometimes (perhaps mistakenly) cast as a choice between time and transparency vs accuracy. Both models satisfy International Society for Pharmacoeconomics and Outcomes Research recommendations for model transparency, which accept complexity and call instead for a technical report that describes the structure, components, equations, and computer code that would enable experts to reproduce the model (full technical transparency) and non-technical documentation that, at a minimum, describes the type of model and intended applications, funding sources, model structure, inputs and outputs, data sources, model validation, and model limitations [11]. While transparency in a general sense is hard to quantify, and no fit-for-purpose diabetes models are likely to achieve "transparency" in a general sense, analysts (authors AN and AL) generally considered that IHE-DCM was easier to grasp and work with (and is constructed with approximately 50% fewer lines of code).

This analysis has several strengths, including the use of two models that were relatively similar and required limited standardization. Many of the remaining differences could be standardized to minimize the extent that differences would be driven by model differences other than units of observations. The scenarios were inspired by the Mount Hood Reference Case, which permits comparison (at least of the Reference Case results) with 11 health economic models of diabetes that participated in the 9th Mount Hood Diabetes Challenge Network. Finally, a wide range of scenarios was considered that explored different aspects of the model to enhance generalizability.

The models could not be entirely standardized, however, and remaining differences must be considered when interpreting the results of this analysis (i.e., between-model differences may reflect more than just the potential bias related to cohort vs micro-simulation modeling). In particular, the main structural difference is the modeling of CKD, for which there are different methods of simulating disease progression (transition probability vs biomarker driven) and which clearly impact the results. Indeed, for the cumulative incidence, the CKD outcomes (micro-and macroalbuminuria and end-stage renal disease) were clear outliers and the mean estimates for the IHE-DCM model were just within the 95% confidence interval of ECHO-T2DM. Foot ulcer is included only in ECHO-T2DM. To mitigate the impact on the analysis, costs and QALY weights were set to 0. The indirect impact on overall results was limited because foot ulcer affected only the risk of congestive heart failure (though patients simulated to develop congestive heart failure had in turn increased risks for ischemic heart disease, myocardial infarction, and mortality) and the simulated incidence of foot ulcer was low. Second, while the scenarios were constructed to mimic a cost-effectiveness analysis, the simulated scenarios are purely hypothetical.

While this study cannot provide a definitive (and universal) answer to concerns about possible bias, and it does not address the academic discussion of how much accuracy is reasonable to swap for increased transparency [30], this exercise provides a careful examination of how two otherwise similar models respond to the same set of stimuli (both absolutely and incrementally), which can be valuable for stakeholders charged with interpreting evidence produced by IHE-DCM.

6 Conclusions

The IHE-DCM was faster to load and to run than the micro-simulation model used in this study (ECHO-T2DM) and the modeling details are likely to be more easily understood by external reviewers, which can be an advantage for economic stakeholders with limited time and resources. Despite systematic differences in absolute predicted survival, QALYs, and costs, estimated cost-effectiveness metrics were similar suggesting that any bias related to the cohort approach is small in the outcomes that matter most. We believe that both models are suitable for use in cost-effectiveness evaluations for interventions in T2DM; the selection of one over the other should be made on the basis of stakeholder needs, resources, and preferences.

Compliance with ethical standards

References

1. Fowler MJ. Microvascular and macrovascular complications of diabetes. Clin Diabetes. 2008;26(2):77–82. https://doi.org/10.2337/diaclin.26.2.77.

2. 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 (Clin Res Ed). 2000;321(7258):405–12.

3. ADA. Standards of medical care in diabetes: 2020. Diabetes Care. 2020;43(S1).

4. ADA. Economic costs of diabetes in the U.S. in 2017. Diabetes Care. 2018;41(5):917–28.

5. Kanavos P, van den Aardweg S, Schurer. Diabetes expenditure, burden of disease and management in 5 EU countries. LSE Health, London School of Economics, 2012. Available from: https://eprints.lse.ac.uk/54896/. [Accessed Apr 2020].

6. Bommer C, Sagalova V, Heesemann E, Manne-Goehler J, Atun R, Barnighausen T, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes Care. 2018;41(5):963–70. https://doi.org/10.2337/dc17-1962.

7. ADA. Guidelines for computer modeling of diabetes and its complications. Diabetes Care. 2004;27(9):2262–5.

8. Mt Hood Diabetes Challenges. Diabetes simulation modeling database. Available from: https://www.mthooddiabeteschallenge.com/registry. [Accessed 2 Dec 2019].

9. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force-2. Med Decis Mak. 2012;32(5):678–89. https://doi.org/10.1177/0272989x12454941.

10. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. Value Health. 2012;15(6):812–20. https://doi.org/10.1016/j.jval.2012.06.014.

11. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force–7. Value Health. 2012;15(6):843–50. https://doi.org/10.1016/j.jval.2012.04.012.

12. CADTH. CADTH pharmaceutical reviews update: issue 12. 2020. Available from: https://cadth.ca/cadth-pharmaceutical-reviews-update-issue-12?utm_source=CONS+List&utm_campaign=19212071f1-lecture-08-06-2019_COPY_01&utm_medium=email&utm_term=0_f3b3313866-19212071f1-263121177. [Accessed 26 Feb 2020].

13. Palmer AJ, Mount H, Modeling G, Clarke P, Gray A, Leal J, Lloyd A, et al. Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. Value Health. 2013;16(4):670–85. https://doi.org/10.1016/j.jval.2013.01.002.

14. Govan L, Wu O, Lindsay R, Briggs A. How do diabetes models measure up? A review of diabetes economic models and ADA guidelines. J Health Econ Outcomes Res. 2015;3(2):132–52.

15. 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(10):e110235. https://doi.org/10.1371/journal.pone.0110235.

16. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU technical support document 15: cost-effectiveness modelling using patient-level simulation. 2014.

17. Lundqvist A, Andersson E, Steen Carlsson K. The costs of diabetes in 2020 and 2030: a model analysis comparing innovative glucose lowering treatments in second line following European and American guidelines compared to current standard of care. In: IHE Report 2016:9, Lund. Available from: https://ihe.se/wp-content/uploads/2016/10/IHE-Report_2016_9_.pdf. [Accessed 3 May 2020].

18. Ericsson A, Fridhammar A. Cost-effectiveness of once-weekly semaglutide versus dulaglutide and lixisenatide in patients with type 2 diabetes with inadequate glycemic control in Sweden. J Med Econ. 2019;22(10):997–1005. https://doi.org/10.1080/13696998.2019.1614009.

19. Ericsson A, Glah D, Lorenzi M, Jansen JP, Fridhammar A. Cost-effectiveness of liraglutide versus lixisenatide as add-on therapies to basal insulin in type 2 diabetes. PLoS ONE. 2018;13(2):e0191953. https://doi.org/10.1371/journal.pone.0191953.

20. Ericsson A, Lundqvist A. Cost effectiveness of insulin degludec plus liraglutide (IDegLira) in a fixed combination for uncontrolled type 2 diabetes mellitus in Sweden. Appl Health Econ Health Policy. 2017;15(2):237–48. https://doi.org/10.1007/s40258-016-0301-y.

21. Johansen P, Hakan-Bloch J, Liu AR, Bech PG, Persson S, Leiter LA. Cost effectiveness of once-weekly semaglutide versus once-weekly dulaglutide in the treatment of type 2 diabetes in Canada.

Pharmacoecon Open. 2019;3(4):537–50. https://doi.org/10.1007/s41669-019-0131-6.

22. Steen Carlsson K, Persson U. Cost-effectiveness of add-on treatments to metformin in a Swedish setting: liraglutide vs sulphonylurea or sitagplitin. J Med Econ. 2014;17(9):658–69. https://doi.org/10.3111/13696998.2014.933110.

23. The Norwegian Medicines Agency (NOMA). Hurtig metodevurdering ved forhåndsgodkjent refusjon §2: Ozempic (semaglutid) til behandling av type 2 diabetes mellitus. 2019. Available from: https://legemiddelverket.no/Documents/Offentlig%20finansiering%20og%20pris/Metodevurderinger/O/Ozempic_T2D_2019.pdf. [Accessed 5 Dec 2019].

24. The Dental and Pharmaceutical Benefits Agency (TLV). Ansökan inom läkemedelsförmånerna: Ozempic. 2018. Available from: https://www.tlv.se/download/18.28d2f835166b654d02f301be/1540800166127/bes181025_ozempic.pdf. [Accessed 5 Dec 2019].

25. CADTH. CADTH common drug review: pharmaeconomic review report: semaglutide (Ozempic) 2019. Available from: https://www.cadth.ca/sites/default/files/cdr/pharmaeconomic/sr0594-ozempic-pharmaeconomic-review-report.pdf. [Accessed 3 May 2020].

26. Palmer AJ, Si L, Tew M, Hua X, Willis MS, Asseburg C, et al. Computer modeling of diabetes and its transparency: a report on the Eighth Mount Hood challenge. Value Health. 2018;21(6):724–31. https://doi.org/10.1016/j.jval.2018.02.002.

27. Lin F, Thomas S, Baldwin M. Microsimulation or cohort modelling? A case study in chronic obstructive pulmonary disease (COPD). Value Health. 2013;16(7):A586. https://doi.org/10.1016/j.jval.2013.08.1617.

28. Kuehne FC, Chancellor J, Mollon P, Weinstein MC. PIN36 Microsimulation or cohort modeling? A comparative case study in HIV infection. Value Health. 2007;10(3):A170. https://doi.org/10.1016/S1098-3015(10)69056-0.

29. Zhou HJ, Zhao YJ, Wojciech AC, Lin L, Caro JJ, Moller J, et al. Differences in cost-effectiveness estimates for chronic hepatitis C treatment among cohort Markov model, Markov microsimulation and discrete event simulation. Value Health. 2016;19(7):A853. https://doi.org/10.1016/j.jval.2016.08.468.

30. Eddy DM. Accuracy versus transparency in pharmacoeconomic modelling: finding the right balance. Pharmacoeconomics. 2006;24(9):837–44. https://doi.org/10.2165/00019053-200624090-00002.

31. 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(5):725–34.

32. 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(12):2140–55. https://doi.org/10.1007/s001250100023.

33. Davis WA, Knuiman MW, Davis TM. An Australian cardiovascular risk equation for type 2 diabetes: the Fremantle diabetes study. Intern Med J. 2010;40(4):286–92. https://doi.org/10.1111/j.1445-5994.2009.01958.x.

34. Ahmad Kiadaliri A, Gerdtham UG, Nilsson P, Eliasson B, Gudbjornsdottir S, Carlsson KS. Towards renewed health economic simulation of type 2 diabetes: risk equations for first and second cardiovascular events from Swedish register data. PLoS ONE. 2013;8(5):e62650. https://doi.org/10.1371/journal.pone.0062650.

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

36. 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(9):1925–33. https://doi.org/10.1007/s00125-013-2940-y.

37. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. 2016;34(4):349–61. https://doi.org/10.1007/s40273-015-0327-2.

38. 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(8):1007–211. https://doi.org/10.3111/13696998.2013.809352.

39. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the economic and health outcomes model of type 2 diabetes mellitus (ECHO-T2DM). Pharmacoeconomics. 2017;35(3):375–96. https://doi.org/10.1007/s40273-016-0471-3.

40. Mt Hood Diabetes Challenge Network. Mt Hood Diabetes Challenge Reference Case simulation. Available from: https://www.mthooddiabeteschallenge.com/refsim. [Accessed 27 May 2019].

41. ADVANCE Collaborative Group. ADVANCE–action in diabetes and vascular disease: patient recruitment and characteristics of the study population at baseline. Diabet Med. 2005;22(7):882–8. https://doi.org/10.1111/j.1464-5491.2005.01596.x.

42. Mt Hood Diabetes Challenges. Challenge session final instructions. Available from: https://www.mthooddiabeteschallenge.com/challenge-sessions. [Accessed Dec 2019].

43. Corro Ramos I, 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 Health. 2017;20(8):1041–7. https://doi.org/10.1016/j.jval.2017.04.016.

44. Zethelius B, Eliasson B, Eeg-Olofsson K, Svensson AM, Gudbjornsdottir 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(2):276–84. https://doi.org/10.1016/j.diabres.2011.05.037.

45. Kengne AP, Patel A, Marre M, Travert F, Lievre M, Zoungas S, et al. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. Eur J Cardiovasc Prev Rehabil. 2011;18(3):393–8. https://doi.org/10.1177/1741826710394270.

46. Adler AI, Stevens RJ, Manley SE, Bilous RW, Cull CA, Holman RR. Development and progression of nephropathy in type 2 diabetes: the United Kingdom prospective diabetes study (UKPDS 64). Kidney Int. 2003;63(1):225–32. https://doi.org/10.1046/j.1523-1755.2003.00712.x.

47. Hoerger TJ, Wittenborn JS, Segel JE, Burrows NR, Imai K, Eggers P, et al. A health policy model of CKD: 1. Model construction, assumptions, and validation of health consequences. Am J Kidney Dis. 2010;55(3):452–62. https://doi.org/10.1053/j.ajkd.2009.11.016.

48. Hoerger TJ, Wittenborn JS, Segel JE, Burrows NR, Imai K, Eggers P, et al. A health policy model of CKD: 2. The cost-effectiveness of microalbuminuria screening. Am J Kidney Dis. 2010;55(3):463–73. https://doi.org/10.1053/j.ajkd.2009.11.017.

49. DCCT. Epidemiology of severe hypoglycemia in the diabetes control and complications trial: the DCCT Research Group. Am J Med. 1991;90(4):450–9.

50. Patel A, MacMahon S, Chalmers J, Neal B, Billot L, Woodward M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Eng J Med. 2008;358(24):2560–72. https://doi.org/10.1056/NEJMoa0802987.

51. Bagust A, Beale S. Modelling EuroQol health-related utility values for diabetic complications from CODE-2 data. Health Econ. 2005;14(3):217–30. https://doi.org/10.1002/hec.910.