VALIDATION OF A DIABETES MODEL:

A PROCESS TO SELECT AND ASSESS THE VALIDITY OF AN EXISTING MODEL FOR ECONOMIC EVALUATION OF TYPE 2 DIABETES IN BRITISH COLUMBIA

by

Kirby Lee Smith

B.Sc. Pharm. The University of Manitoba, 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pharmaceutical Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2017

Abstract

Summary

The growing burden of chronic disease like Type 2 Diabetes Mellitus (T2DM) raises concern about the sustainability of Canada's health system. Consequently, greater emphasis is being placed on economic evaluation using computer simulation models to enhance decision making in health care. This research was initiated to assess if model output for a cost-effectiveness analysis (CEA) in T2DM would be generalizable to a diverse BC population.

The thesis outline is as follows:

- Chapter 1 overviews the epidemiology, risk factors and complications of T2DM, the role of model based economic evaluation and knowledge gaps used to guide the thesis objectives.
- Chapter 2 assesses the methods of existing T2DM models and utilizes criteria to select a model for validation.
- Chapter 3 describes an internal validation of a T2DM model by comparing the expected versus actual impact of changes to input parameters on output.
- Chapter 4 describes an external validation comparing the predicted rate of vascular events to the observed rate by gender, age and ethnic sub-cohorts from a BC population based on the coefficient of determination (R^2) and a 95% confidence interval (CI).
- Chapter 5 summarizes the results, discusses limitations, and highlights future research to enhance the credibility of T2DM models.

Results

The Ontario Diabetes Economic Model (ODEM) was selected and an internal validation demonstrated the simulated rate of vascular events responded as expected to changes in baseline variables in a 10 year simulation. The external validation in cohorts with no history of complications had a modest positive correlation ($R^2 = 0.68$) and a tendency to over predict vascular events in older adults. Adding individuals with previous events improved the correlation ($R^2 > 0.99$) and statistical accuracy of the ODEM. A higher correlation was observed in those of younger age, male gender ($R^2 = 0.71$) and SA ethnicity ($R^2 = 0.77$).

Conclusion

The ODEM was demonstrated to be a functional model with output considered generalizable for the economic evaluation of a diverse BC T2DM population. There were trends in model to overestimate complications in cohorts with no previous vascular events and those of older age that require further research to validate.

Preface

There is an important role for model-based economic evaluation to guide decision making and enhance the value of health interventions for our health system and community. Decision makers therefore must have confidence in the economic output from a decision model that it is relevant for their population of interest. In discussions on this topic with my co-supervisors, Dr. Larry Lynd, Director of UBC's Collaboration for Outcomes Research (CORE) and Dr. Nick Bansback from UBC's School of Population and Public Health (SPPH), it was decided that I would develop an MSc level course syllabus for directed study that would describe and compare existing decision model methodology, structures and design being used to perform economic evaluation in diabetes. As part of this project, I would appraise the current approaches used to validate existing models.

From this directed study, I developed the objectives and methods for this MSc thesis with the intention to advance the critical appraisal and credibility of model based economic evaluation in a complex chronic disease using Type 2 diabetes mellitus (T2DM). I completed the initial search of the literature and applied methodology to define objective criteria to first a) identify a list of appropriate existing models for CEA, then b) to assess and compare the methods used in these T2DM models and finally c) to select a model suitable for cost effectiveness analysis (CEA) in a Canadian population for further validation. My interest in maximizing the efficiency of resource allocation for chronic disease interventions led me to propose a validation project to determine the generalizability of the output of the existing model for use in a diverse BC population.

This Ontario diabetes economic model (ODEM) was a T2DM model which met the criteria for performing comparative CEA and it was provided to UBC CORE with open access for further research by Dr. Daria O'Reilly, Associate Director of Programs for Assessment of Technology in Health (PATH) and Associate Professor at McMaster University. I received guidance from Dr. Lynd and Dr. Bansback on the design of the internal and external validation exercises of the ODEM to assess the functionality of the model and the accuracy of the predicted rate of vascular events in a BC population. Further input on the methods and expert insights on T2DM were provided by my thesis committee members Dr. Mark Harrison and Dr. Nadia Khan. I then used this collective input to define the methods, identify appropriate data to populate the baseline input variables by cohort and identified the statistical assessment to be applied for assessment of model outputs. The implementation and interpretation of the ODEM validation exercises was completed as a core component of my thesis project.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

A-fib: atrial fibrillation
AUC: area under the curve
BC: British Columbia
BMI: body mass index
CADTH: Canadian Agency for Drugs and Technologies in Health
CDC: Centre for Disease Control (USA)
CEA: cost effectiveness analysis
CEAC: cost effectiveness acceptability curve
CHF: congestive heart failure
CI: confidence interval
CV: cardiovascular
CVD: cerebrovascular disease
DES: discrete event simulation
DofD: duration of diabetes
DSA: deterministic sensitivity analysis
ESRD: end stage renal disease
EQ-5D: EuroQOL five dimension questionnaire
HbA1c: glycated hemoglobin
HDL: high density lipoprotein
HTA: health technology assessment
HUI3: Health Utilities Index Mark 3
ICD: International Classification of Disease (9^{th} and 10^{th} revision)
ICER: incremental cost effectiveness ratio
IHD: ischemic heart disease
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
LDL: low density lipoprotein
MAPE: mean absolute percentage error
MI: myocardial infarction
NMB: net monetary benefit
NPHS: National Population Health Survey
ODD: Ontario Diabetes Database
ODEM: Ontario Diabetes Economic Model
PATH: Programs for Assessment of Technology in Health, McMaster University
PHAC: Public Health Agency of Canada
PSA: probabilistic sensitivity analysis
PVD: peripheral vascular disease
QALY: quality adjusted life year
R^2: coefficient of determination
RCT: randomized clinical trial
RMSPE: root mean square percentage error
SA: South Asian
SBP: systolic blood pressure
SD: standard deviation

SE: standard error of the mean
SF-36: Short Form Health Survey (36 item)
SMDM: Society for Medical Decision Making
T2DM: Type 2 diabetes mellitus
TC: total cholesterol
TVRF: time varying risk factors
UBC: University of British Columbia
UK: United Kingdom
UKPDS: United Kingdom Prospective Diabetes Study
WESDR: Wisconsin Epidemiologic Study of Diabetes Retinopathy

Acknowledgements

The completion of this graduate degree provides me with a deep sense of personal gratitude, not just from the designation of my Master of Science, but more importantly from the realization that I am blessed to have a strong network of supporters who were critical in helping me realize this goal.

I wish to extend my deep appreciation to Dr. Larry Lynd whom I met with in 2013 to express my interest in pursuing a part time graduate degree in pharmaco-economics. His early encouragement to pursue this academic training and his gracious acceptance to be my supervisor (although I am not sure that was his original intent) was the catalyst I needed to get started. His ongoing guidance, sage advice and candid feedback were provided with patience and humour. He is an exceptional person and outstanding professor and I am grateful to now be one of his many highly skilled trainees who are/were part of UBC CORE.

I wish to also express my sincerest gratitude and respect to Dr. Nick Bansback, a rising star and co-supervisor for my thesis, for his exceptional scientific insight combined with a very practical perspective on the implications and relevance of this research. He challenged me to step back and look at the forest not just the trees.

Special thanks to Dr. Mark Harrison and Dr. Nadia Khan for sharing their collective expertise, asking insightful questions and bringing a perspective that pushed me gain a deep understanding of the research methods and the clinical relevance of this research.

Finally and most important is my family, who have sacrificed so much of their time and energy to enable me to pursue this degree. My wife, and best friend, Connie who graciously balanced her own career and the demands of raising a young family to allow me to invest the thousands of hours to realize this personal goal. We are blessed with two amazing children, Jonah and Makenna, who have never complained when their dad was too busy with “school” to spend time with them. They are my inspiration that motivates me every day to be the best person I can be so that I feel deserving of their love and support.

To my mother who provides unyielding support and encouragement whenever it is needed. Her family has always been her number one priority and her guidance and love has made us all strong and grounded. To my dad who is no longer with us, thank you for showing me that life is what you make it. You can either accept the cards you are dealt or fight to get a better hand.

Chapter 1: Introduction

In response to the large and growing burden of chronic disease in Canada's aging population, and the potential impact on the sustainability of the health system we are accustomed to, we must work efficiently and collaboratively to maximize the appropriate use of interventions to improve the health of our families, friends and communities to ensure we get maximum value from our health care spend.(1) It has been documented that globally the morbidity and mortality of non-communicable chronic disease is growing and has exceeded all other diseases combined in recent decades.(2) The increasing prevalence is attributed, in part, to poor lifestyle choices including physical inactivity, unhealthy diets and smoking.(3) These are all considered modifiable risk factors and are associated with an increase in cardiovascular disease, T2DM, some cancers, arthritis and chronic obstructive pulmonary disease among other diseases.(1) As a result of the growing burden, it is no longer sufficient to demonstrate to decision makers that health interventions are clinically effective or proven safe, but also these data must consider the cost of care and the impact on quality and/or quantity of life to facilitate a measure of cost effectiveness.(4) This form of evidence, known as economic evaluation, often relies on the use of a decision model in order to compile evidence and data into a structured assessment of a potential intervention, intended to improve the outcomes related to a chronic disease, to provide an output measure of its value to the health care system.(5)

This thesis focuses on T2DM, which is one such chronic disease with a large and growing prevalence in Canadians, that is known to increase the risk for a number of vascular events or diabetes related complications.(6,7) There are many interventions including medicines, devices and services that have been developed to mitigate the risk of these long term complications due to T2DM which have increased the demand for economic evaluations to support efficient resource allocation.(8) Over the last two decades dozens of decision models have been developed by researchers to provide output for decision makers for economic evaluation of primary, secondary and tertiary prevention of T2DM and its management. While a number of T2DM models have been designed for broad assessment of a number of interventions to address a range of T2DM related vascular complications, there are a number of the T2DM models in the literature which have been created for a one-time analysis of a specific intervention or to assess one specific vascular outcome (ie: retinopathy). Typically these single purpose models have not published methods or completed model validation exercises to allow a full transparency of the model structure. Therefore to enable an assessment of the most appropriate T2DM models this thesis focuses on models which have incorporated a broad assessment of T2DM and associated complications and which have been used in the economic evaluations of multiple interventions or have published model methods or validation exercises to demonstrate the functionality and/or validity of their model to predict T2DM complications.

This chapter begins with the background on the epidemiology and growing burden of T2DM that includes a breakdown of the risk factors for both developing T2DM and for experiencing subsequent diabetes related vascular events or complications. There are many risk factors that are relevant to T2DM and include modifiable risks, such as obesity and sedentary lifestyle, and non-modifiable risk factors such as age, gender and ethnicity.(9) Of these risk factors there are noted differences in British Columbia versus other regions in Canada and other countries which may impact the onset of T2DM and the rate of complications. In particular there is a lower rates of obesity and smoking, and a higher proportion of ethnic populations such as those of South Asian (SA), Aboriginal and Chinese descent that may impact the CEA of interventions in T2DM.(6, 9–11, 2) This overview of T2DM is followed by a summary of the growing importance of Health Technology Assessment (HTA) in Canada including the role of economic evaluation of health care interventions using decision models to generate measures of cost effectiveness to enhance the efficiency of allocating resources.(4) The chapter concludes with an overview of identified knowledge gaps and potential areas of future research that were used to guide the development of the objectives of this thesis project.

1.1 Diabetes in Canada

1.1.1 Demographics and epidemiology

The population of BC in 2015 is currently 4.6 million residents with growth projected to outpace the national average, reaching approximately 6.0 million residents or 14.2% of the Canadian population by 2038. This population growth will be supported by an international and interprovincial migration that will be above the national average. The mean age in BC's population is older than the average Canadian, aging faster due to interprovincial migration, with those over the age of 65 estimated to grow from approximately 23% to 34% of the BC population by 2041.(13,14)

The BC population is becoming increasingly ethnically diverse with the largest and fastest growing ethnic populations being of South Asian and Chinese descent. Those identified as SA in the 2011 census in BC totaled 313,440 or 7.2% of the population and those from China equaled 423,435 or 10.1%.(15) The predicted growth of the SA and Chinese population is projected to be greater than the general population growth due to the high immigration rates from these regions of the world.(13, 15)

In the BC population, it is estimated that the prevalence of T2DM will continue to rise from 8.3% in 2013 to 10.3% by year 2020 and will comprise approximately 548,000 citizens.(16) This is growth in prevalence is believed to be due to a combination of a higher incidence of T2DM in older age groups, an aging population, increasingly sedentary lifestyles leading to higher rates of obesity, a growing proportion of ethnic populations who have an increased prevalence of T2DM and an overall trend of people living longer with T2DM after diagnosis.(17)

1.1.2 Burden of disease

In Canada, the cost of health care currently represents between one third and one-half of the annual provincial budgets and is increasing.(18) This trend has been recognized by the BC provincial Ministry of Health who have outlined strategies to enhance the quality, efficiency and value of spend in health care.(19) A significant contributor to the burden on health budgets are the direct costs attributable to the ongoing management and treatment of complications of T2DM. These costs have doubled across Canada between 2000 and 2010 and are estimated to increase from $294 million in 2010 to approximately $472 million by 2020 in BC alone.(17) These direct costs include the cost of medication and devices to manage blood glucose control and the vascular events directly attributed to T2DM such as amputation and renal dialysis, as well as the incremental cost of hospitalizations for vascular disease and physician visits (general practitioner and specialists) over those without a diagnosis of T2DM.(17) There are many long term complications from T2DM that occur as a result of micro and macro vascular disease including ischemic heart disease (IHD), myocardial infarction (MI), cerebrovascular disease (CVD), congestive heart failure (CHF), blindness, amputation and end stage renal disease (ESRD).(6,7) These vascular complications contribute to a higher direct cost in the year of the event and an increase in the direct costs of ongoing treatment in those who survive an event.(20) The growing burden of chronic disease in the aging Canadian population has led to the development of national and provincial strategies to identify the priorities which may reduce the long term impact of the risk factors associated with T2DM.(1, 17, 19)

1.1.3 Risk factors

An area of focus to reduce disease burden in T2DM is to identify and target interventions that mitigate the risk factors shown to either increase the onset of T2DM or are associated with increasing the rate of complications in people with T2DM.(6,9) Importantly, in T2DM these include modifiable risks which create an opportunity to intervene and alter the course of the disease.(6,9) It is widely recognized that T2DM is a complex chronic disease in which the increased rate of vascular events are due to more than elevated blood glucose levels alone but often include comorbid conditions such as high blood pressure and/or elevated blood lipid measures such as total cholesterol (TC), triglycerides and low density lipoprotein (LDL).(6) The presence of these multiple risk factors in an individual diagnosed with T2DM are referred to as metabolic syndrome, a condition which requires numerous interventions and aggressive management due to the high risk of complications.(21) Therefore, the economic evaluation of interventions in T2DM should account for these known modifiable and non-modifiable risk factors to ensure the output of diabetes decision models is reflective of the actual onset or progression of T2DM in the population of interest.

Non-modifiable risk factors

Non-modifiable risk factors are considered as those which are not within the control of an individual to alter or reverse. The most widely recognized of these is the impact of age on the incidence of T2DM which rises rapidly after age 40 and essentially quadruples from 5.1/1000 individuals in the 40 to 45 age category to 20.3/1000 individuals in the 70 to 75 age category. In addition to the increased incidence of T2DM with age is the higher rate of diabetes related events which further compounds the impact of advancing age on the burden of T2DM.(22) Gender is another non-modifiable risk factor where there is a numerically higher incidence of T2DM in males (6.8/1000) than females (5.7/1000) in the overall Canadian population and remains consistently higher in males of all age categories starting from 35 years and continues to 85 years of age and older.(9) Another relevant non modifiable risk factor for the BC population is the higher proportion of certain ethnic populations in the province which are known to have a higher incidence of T2DM and potentially differing risk of vascular complications over time. These differences may be a reflection of biological and behavioral differences including diet and activity levels, type of body fat distribution, age of onset of T2DM and metabolic measures of disease control.(6,9,23) For instance, there is a greater than two fold increase in the odds for developing T2DM in those of SA descent compared to Caucasians as well as differing rates of vascular complications in the SA population with T2DM.(9, 23)

Modifiable risk factors

There are also a number of modifiable risk factors which are known to impact both the incidence of onset of T2DM and the rate of complications related to T2DM. Importantly, an intervention that can delay the onset of T2DM has been shown to reduce the rate of cardiovascular disease and renal failure.(6) For example, a key modifiable risk factor for T2DM is obesity which is known to increase both the incidence of T2DM and the rate of complications in T2DM.(6) In Canada, obesity is defined as a person with a BMI of greater than 30, and overweight is considered as having a BMI in excess of 25.(9) Canadian data shows that the mean BMI in people with T2DM was 29 kg/m2, compared to 25kg/m2 for those without T2DM which results in three quarters (75.6%) of those with T2DM in Canada being considered either overweight or obese.(9) Furthermore, Canadian studies shows self-reported obesity is higher in Caucasians than in those of other ethnic groups such as SA, Chinese and black which suggests the need for ethnic specific cut offs and the inclusion of other measures beyond BMI such as waist circumference to better assess cardiovascular risk.(6, 19, 20)

Another key modifiable risk factor related to the risk of complication is smoking, as those with T2DM who smoke have a significantly higher risk of ESRD, MI, stroke and death.(25) The prevalence of smokers in Canada has declined a mean of 10.3% between 1985 and 1999 and 14.7% in adults between age 45 and 64 years of age however it remains higher in men and the Caucasian population.(12) Of note, in 2005 the rate of current and occasional smokers in BC was approximately 20% (versus Canadian mean of 23%) and was the lowest rate in Canada for both males and females due to a combination of more people who had never smoked and a higher

proportion who had quit smoking. (26) The smoking rate clearly differs across the country by age, gender and ethnic group making it a relevant risk factor to consider in assessing interventions to reduce the burden of T2DM.(11, 23)

In addition to lifestyle related increase in risk factors, the main metabolic measures of T2DM and metabolic syndrome such as elevated blood glucose, blood pressure and cholesterol increase the risk of vascular events over time and have become the primary focus of interventions aimed at reducing the complication of T2DM. Of these metabolic measures, the level of blood glucose control is often measured as glycated hemoglobin (HbA1c) and is a primary focus when assessing the level of control of T2DM.(6) There is evidence of statistically significant reductions in the rate of micro vascular events and major cardiovascular (CV) events with an improvement in blood glucose control and the associated reduction of HbA1c over a number of years.(27–29) In addition to long term control of blood glucose many with T2DM require additional treatment to improve blood pressure and blood lipids due to the greater risk of cardiovascular events in individuals with this metabolic condition, or syndrome, than those with elevated blood glucose alone.(21,25) A retrospective review of the results from a Canadian health survey predicted that the risk of fatal cardiovascular events would be more than fourfold higher in those with metabolic syndrome (4.1%) than those with T2DM alone (0.8%) over ten years. In addition, a recent survey found that only 13% of Canadians with T2DM had all three metabolic targets at or below target levels and the prevalence of metabolic syndrome was higher in the non-Caucasian population.(6) Guidelines for T2DM management in Canada therefore appropriately include target levels for the control of blood glucose, blood pressure and blood lipids which vary depending on the individuals CV risk profile.(6)

1.2 Health Technology Assessment (HTA)

The pressure on health systems from the growing burden of chronic disease has led to increased use of economic evaluation to guide HTA and reimbursement recommendations of new technologies which include prescription medicine, diagnostic tests and devices for surgical, medical or dental procedures. The use of decision models to incorporate available clinical and economic data to provide a structured cost effectiveness analysis are central to the development of HTA reports in many countries.(4,8,30,31) In Canada, HTA is often provided to the federal, provincial and territorial drug plans by the Canadian Agency of Drugs and Therapeutics (CADTH) which developed guidelines in 2006 for economic evaluation which are currently in the process of being updated in 2017. These recent Canadian guidelines are intended to improve the consistency of approaches to complete economic evaluation with recommendations that include defining the decision problem, utilization of a reference case for better comparison of results, appropriateness of data sources and transparency of methods. This report also incorporates recommendations to reflect methodological advancements such as model validation and the incorporation of probabilistic analysis to enable analysis of the joint uncertainty of model parameter estimates on

model output. . These HTA reports generated by CADTH are used by provincial and federal Ministries of Health to guide their reimbursement decisions for their respective jurisdictions.

1.2.1 Decision modelling

The conducting of an economic evaluation can be trial-based or model-based and depends on the evidence available for the required analysis, the scope of the decision and perspective of the decision maker. A large scale clinical trial may be adequate to compare the cost and consequences in the population of interest however, in situations where evidence is captured from multiple sources there may be a need for a decision model to generate measurements of the incremental cost effectiveness of an intervention, policy or program.(32) These models are a simplification of a disease process and require a number of assumptions to reflect the risk factors, rate of events and progression of disease in a defined population over time.(5,32) Decision models allow researchers to explicitly incorporate available evidence from disparate sources to incorporate the clinical and economic benefit of a single or multiple health interventions to inform the decision.(5,32) Decision models are therefore a credible tool with methods that can be critiqued and adjusted based on advancing knowledge, emerging evidence and model validation. There are often concerns about the credibility of the output of decision model which highlights the need for researchers to provide transparency of their model methods, which may be distinctive in structure and assumptions, to assist decision makers in the selection of an appropriate model for assessment of cost effectiveness in complex, chronic diseases such as T2DM. (33)

The primary output measure in a CEA is the incremental cost effectiveness ratio (ICER) which can be compared to the cost effectiveness of other interventions in T2DM or, depending on how effectiveness was measured, other disease areas. This ICER compiles the incremental cost of the intervention, disease and event management, and then divides it by the incremental health gain which is most often measured using life years or quality adjusted life years (QALY).(34) The QALY is a measure derived from the assessment of patient preferences or health utility which can be multiplied by the life years gained to create a generic measure of the effectiveness of an intervention. The QALY is not a disease specific measure and therefore allows the CEA output to be compared within and across different disease states or health conditions. The ICER value generated by the model is then compared to a range of standard cost thresholds set by HTA bodies and health care decision makers to determine if an intervention can be deemed to be cost effective in the population of interest.(34) In addition to an ICER, models are often designed by researchers to generate scatter plots on a cost effectiveness plane, cost effectiveness acceptability curves (CEAC) and measures of net monetary benefit (NMB) to provide more information for the decision maker to better interpret the uncertainty of the CEA output from the model.(4, 32)

1.2.1.1 Model uncertainty

The level of uncertainty previously mentioned is an important consideration for those who use decision models to guide their decisions and is a reflection of the gaps in information for the estimates used to generate the CEA.(32) The methods and structure used to build the diabetes model can result in differing forms of uncertainty which may vary depending on the design, assumptions and quality of evidence.(32) For instance, a model designed to use individual microsimulation may have stochastic, or first order, uncertainty due to the variability between individual outcomes even when using the same probability of events and response to interventions.(35) Parameter, or second order, uncertainty is a reflection of the need to estimate parameters from the available evidence in the model such as the probability of an event or the effectiveness of an intervention. Finally there is inherent structural uncertainty due to assumptions in the model design to simplify the disease process which may not fully reflect the complex evolution of risk factors over time, progression of disease and its comorbidities or the side effects of treatment.(32,35) Methods such as PSA and/or deterministic sensitivity analysis (DSA) have been incorporated into decision models by researchers to enhance the ability to gauge the individual or joint uncertainty of estimated model parameters on the output of a CEA.(35)

Deterministic Sensitivity analysis (DSA)

A common approach researchers use to quantify the impact of parameter uncertainty is via DSA. This can be done by adjusting a single parameter at a time or by multiple parameters at once and observing the impact on the model CEA output. Instead of a random or subjective selection of the range for adjusting the model parameter it is recommended that the adjustment of the input parameter be based on evidence or a statistical measure to reflect an appropriate range of values that could be explained or justified for this model parameter.(34,35) For instance, if the mean and standard deviation of the treatment effect of an intervention is known these values can be used to calculate the upper and lower range of the 95% CI for input and model simulation. The resulting changes in model output in response to range of input variables provide decision makers with additional information to assess the level of uncertainty when using estimated model input variables or parameters.(35) The results of one-way DSA using threshold values for input parameters is often presented graphically via a tornado plot to highlight which model input parameters have the greatest potential impact on the CEA output. This approach can also be used to assess structural uncertainty by capturing the impact of important structural changes on model outcomes such as whether to incorporate a specific risk factor into a T2DM model.(35)

Probabilistic Sensitivity Analysis (PSA)

Another technique which is useful in quantifying the joint uncertainty of decisions based on the model parameter and structure is PSA.(35) Instead of assessing the range of the impact on a point estimate as done in DSA, this approach enables the researcher to apply a probabilistic distribution to the key input parameters (ie: biologic

measures, treatment effect, cost, QALY) to quantify their joint uncertainty on the model CEA output. The selection of the probabilistic distribution around model input parameters (ie: normal, beta, gamma, logistic) is guided by the type of data (ie: continuous or binomial) and the type of distribution (ie: normal or skewed).(32,34) Therefore, probabilistic parameterization allows the assessment of the overall uncertainty even when simulating multiple, non-linear parameters and the range of potential outcomes over time which is very valuable when assessing a CEA for long term outcomes in a complex chronic disease such as T2DM. The ability of a model to incorporate PSA allows researchers to compile and communicate uncertainty of model output in a way that improves comprehension with a visual overview of the results provided via a scatter plot of all outputs on a cost effectiveness plane or through the creation of a CEAC.(35) Based on this, HTA bodies now recommend PSA as the base case to reference when reporting model output and the associated uncertainty to complement the point estimate of an ICER.(4,)

1.2.1.2 Model development: sources of data

The majority of the diabetes models are based, at least in part, on the landmark, long term, randomized clinical trial (RCT) known as the United Kingdom Prospective Diabetes Study (UKPDS).(36–43) Some models have been updated by incorporating results from other large scale RCTs including trials designed to assess interventions in non-diabetic populations who had elevated cholesterol and/or hypertension.(36,39,43) Other models have incorporated epidemiological datasets such as the United States Centre for Disease Control (CDC) (37), Framingham (36), Wisconsin Epidemiologic Data (38,42), Kaiser Permanente diabetes registry (44) and Australian epidemiologic data (45).

These various sources of data used to build the model parameters creates a potential for differences in the predicted vascular event rates such as MI, stroke and ESRD across different models used for CEA in a population treated in a clinical practice to present day standards. For example, the UKPDS assessed diabetes outcomes in a RCT environment in the United Kingdom between 1977 to 1997 and is unlikely to reflect the current metabolic targets or standard of care in Canada.(27) The use of Framingham data to measure cardiovascular risk, while extremely valuable, was based on a population of primarily non-diabetic patients.(46) The Wisconsin Epidemiologic Study of Diabetes Retinopathy (WESDR) used observational data from the early 1980's and may not reflect the rate of retinopathy or blindness in today's T2DM population.(47,48) The models which have been designed based on large scale RCT's often which have restrictive inclusion criteria raises the question of whether the rate of T2DM related events in these models are reflective of a heterogeneous population treated in a real world clinical practice.(44) A number of T2DM models have been tested using observational data to compare the predicted versus observed event rates in a broader population. Ideally, to consider a model output as valid for the

purpose of the CEA the model output should be compared with data reflective of the population in which the intervention is intended to be utilized.(34, 47–49)

1.2.1.3 Model validation

The variety of methods that have been used to create the models used in T2DM highlight the importance of assessing the ability of each model to generate relevant and credible output through ongoing validation exercises. There are different forms of validation that are recommended be utilized to test a model prior to its use by decision makers for economic evaluation. For instance, comparison of model output against the data used to create the model may be done to determine if the model functions as expected is considered an internal validation. An assessment on whether the model has incorporated the relevant risk factors based on expert opinion on the progression of T2DM is known as face validation. An assessment of whether the model generates similar output to other diabetes models when inputting similar baseline variables is a form of cross validation. Finally, and considered most valuable for model validation, is how well the model accurately estimates or predicts T2DM in the real world such as the rate of diabetes complications observed in a study or a population of interest and is an external or predictive validation.(33,47,50) The limitations in the ability of a model to fully capture the risk factors, comorbid condition and progression of a complex, chronic disease like T2DM in the population highlights the value of ongoing validation of a decision model to increase the overall confidence in the CEA output.(33, 35)

There have been significant efforts to create more standardized methods to guide modelling practices. For instance, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and Society for Medical Decision Making (SMDM) have recently published reports which identify a set of best practices which will be utilized to guide validation efforts.(5) These validation exercises can assess the model as a whole or focus on key components such as disease transition or event rates using data from similar, but different, populations.(33) Also, since 2000, computer modellers have participated in regular meetings, known as the Mount Hood Challenges to allow them to share ideas and increase the understanding of their models for use in diabetes. These meetings provide a venue for researchers to perform exercises to compare model output against clinical trials (external validation) or other T2DM models (cross validation).(47,50) The results from these Mount Hood Challenges have been useful for researchers to identify methodological issues when comparing model output and for predicting the absolute benefit of a range of T2DM interventions. It is recommended that validations be performed in a similar population in which the model will be utilized to identify how well the predicted and observed rate of events correlate to further increase the credibility of output. These validation efforts are not typically intended to compare the components of a model which are expected to vary widely by jurisdiction, such as health resource utilization and associated health care costs.(53)

The development of a credible decision model relies on the availability of data to guide the design, assess model functionality and validate disease event and mortality rates.(33) The availability of summary output of vascular events in a cohort of individuals from BC who were diagnosed with T2DM between 1993 and 2006 created an opportunity to perform an external validation exercise that would compare the observed rate of diabetes complications in a local population against the predicted output from an existing T2DM model.

1.3 Knowledge gaps: economic evaluation in diabetes

1.3.1 Model selection for economic evaluation

There are many models that are available and in use for economic evaluation of T2DM in order to simulate the comparative numbers of diabetes related complications, with an intervention versus a standard care, to produce a CEA output. These models have been developed by researchers using differing methods and assumptions that should be considered in the selection of the most appropriate model depending on the scope and purpose of the model simulation. Some of these differences include whether the model is a dynamic population based or closed cohort structure, whether it incorporates individual or group simulation, and whether time varying risk factors have been incorporated into the simulation calculations over time. Furthermore, the guidelines for economic evaluation require that decision models utilize PSA in order to generate output that enables the assessment of uncertainty in model output and not all researchers may have incorporated this into their methods. The decision problem for the economic evaluation, scope of the CEA and demonstration of model validation in the population of interest should all be considered in the selection of an appropriate model.

1.3.2 Relevance for real world populations

While most models have published internal and external validation, in the majority of cases these were based on comparing the simulated diabetes related event rates with the observed number of events from large RCT's. For instance, the results of the RCT may not be reflective of a more diverse, real world cohort of people with T2DM who have differing ethnicity, differing levels of risk factors or mix of pre-existing conditions. Also, the observational data used to create or to validate diabetes models have often come from data that was captured in an era that may no longer reflect the standard of care, lifestyle or behaviors of today's population of people with T2DM. External validation of a T2DM model using a population treated to a real world standard of care will provide valuable insights on the credibility of a diabetes model for its use in the economic evaluation of interventions in a more diverse population.

1.3.3 Generalizability for use in BC

The data used to populate the T2DM models have almost exclusively been captured in patients and populations from the United Kingdom (UK) and United States. Of the numerous models identified as multi-use models for T2DM, none of these were created or validated based on a large scale RCT or observational datasets in a contemporary Canadian population. Considering there are notable differences in the proportion of baseline risk factors such as demographic, ethnic, lifestyle, BMI, smoking rates, and treatment options in different regions it is unclear whether existing models accurately predict the rate of vascular events in a BC population. Model validation in various populations demonstrate that the impact of risk factors and T2DM interventions on the relative rates of vascular complication are consistent across similar populations, however the absolute magnitude of the estimated outcomes have been shown to vary across different populations. Therefore, while external validations have been completed for many of these T2DM models, the overall predicted rate of events results can only be assumed to be applicable to a similar population as the one used for the assessment. Decision makers are likely to have a greater level of confidence in the credibility of model output and the generalizability of CEA results for models which have been validated for use in a BC T2DM population.

1.4 Thesis objectives

There is a growing demand for model-based economic evaluation, as a compliment to clinical evidence, to simulate the predicted impact of an intervention on the progression of T2DM and vascular outcomes experienced by people with T2DM. A key component to the relevance of diabetes model output is whether the simulated outcomes are reflective of outcomes in a real world clinical practice. The methodological differences between T2DM models raises the question of which existing diabetes models will accurately reflect the diabetes trajectory and event rates experienced by people with T2DM in BC. As recommended by the ISPOR-SMDM it is important that model methods and validation results are transparently shared, appropriate for the decision problem being addressed prior to their selection and use for an economic evaluation.

This overarching objective of this thesis is to assess if an existing T2DM model could provide credible and trusted output from a cost-effectiveness analysis (CEA) that would be considered generalizable to a diverse BC population. The initial aim is to overview the methodology of existing T2DM models to apply a set of selection criteria that would advise the choice of an appropriate model for further validation exercises. Once selected, the next aim of the thesis is to demonstrate the functionality of the T2DM model and the responsiveness of model output to differing demographic variables and risk factors via an internal validation exercise. This is designed to assess whether the changes in predicted vascular event rates corresponded with the actual change in response to changing each baseline input parameter one at a time. Finally, an external validation exercise is performed to compare the cumulative incidence of seven diabetes related vascular complications predicted by the model with

the observed rate of complications from a real world BC population dataset of people with T2DM. To allow a full interpretation of the generalizability of the T2DM model in BC's diverse population the cohorts are separated into a general and a SA population, then further split into sub-cohorts based on gender and age.

Completion of these objectives will provide an in depth overview of functionality and absolute predicted rate of vascular events for multiple cohorts to assess the validity of a T2DM model in the overall population and in sub-groups based on gender, age and ethnicity. Economic evaluation using a T2DM model that is shown to provide credible output in a high burden, complex chronic disease will be valuable for guiding efficient resource allocation by decision makers in support of efforts to maintain a sustainable Canadian health care system.

Chapter 2: Assessment of diabetes decision models and methods

2.1 Introduction

The initial aim of this thesis was to implement a structured approach for the assessment and selection of a diabetes model based on a set of objective criteria which were deemed important for providing a CEA from the perspective of a BC health system. There are many models in use for economic evaluation of T2DM that range from broad epidemiological based dynamic models which incorporate incident and prevalent changes in the T2DM population, to more narrow closed cohort models that are designed to compare the value of an intervention to reduce the occurrence of a single complication of T2DM such as retinopathy. Some of these models have been created for a single analysis of a new intervention in specific population while others have been widely used for many evaluations in multiple populations.(54,55) Therefore numerous factors must be considered when selecting a model for use in T2DM due to the complexity of the disease including the associated comorbidities and long term disease progression, the numerous risk factors impacting both incidence of disease and T2DM related events, as well as the range of potential vascular complications related to T2DM. In light of the many factors for developing a CEA in T2DM it is important that the methods for model screening, assessment and selection be based on a set of objective criteria to ensure the results remain relevant and credible for the purpose of the economic evaluation.

Due to the large number of economic models available for evaluation of T2DM a maximum of ten were selected from the literature to allow the inclusion of differing model structures, methods and regional representation to provide a cross section of T2DM models for the more in-depth model assessment. These methods were tabulated for comparison and five objective criteria were applied to identify which of the T2DM models were considered eligible for further validation exercises.

2.1.1 Critical appraisal

The structured approach used to identify appropriate selection criteria and applied to the model methods was based on consensus papers and reference texts for economic evaluation and decision modeling. The references selected to guide this critical appraisal were the “Decision Modelling for Health Economic Evaluation” by Briggs, Claxton and Sculpher (32) and the ISPOR-SMDM Task Force Reports.(5,33,35,56) The definition in the Briggs et al describe decision modelling “as a systematic approach to decision making under uncertainty” which includes the variability of the patient population, estimates of disease related inputs and transition probabilities.(32) This definition and related points highlighted the inherent value of decision models to structure the large volume of available data to capture the complexity of the diseases for which they models are utilized to simulate. This insight demonstrated the need for identifying the standards used to guide the critical appraisal and selection of a model to minimize and measure the uncertainty of the results of model based CEA.

An overarching guideline outlined in these references is for researchers and decision makers to first define the scope of the evaluation that include the perspective of the decision maker (ie: drug plan, health system, societal) and health care setting (ie: community or institution). In the context of T2DM, these included the need to define the purpose of the intervention as it will depend on the intended use for CEA of a primary, secondary or tertiary intervention.(32)

Next, decision makers should define the boundaries of the decision model to simplify the simulation of the disease while balancing the importance of including the relevant baseline input variables and risk factors to adequately capture disease progression.(5) When decision makers are considering the use of a decision model they must define the time span for the economic evaluation, the target population and the desired model outputs prior to determining whether appropriate evidence is available to input into the model.(5) Once these objectives have been decided, there are differing model structures to consider in order to balance the complexity of the disease input variables with the resources and data available for completing the analysis. For instance, straight Markov models are structured to simulate a cohort of people with the same characteristics, providing decision makers with a more simplified simulation of a disease, which does not account for the complexity of tracking the disease history or the time in a particular health state for each individual. In contrast, individual microsimulation provides greater flexibility to incorporate each individual's attributes that may influence model outputs such as time in a particular health state, unique patient baseline characteristics and influence of time varying risk factors.(56,57) To allow this added level of complexity the models which use microsimulation require more data, time and computer processing power to simulate the progression of disease.(58)

Once the scope, population and boundaries of the analysis are identified it is important to ensure that the model methods allow users the ability to generate the output to assess the overall uncertainty of the CEA output.(32,35) There is a growing consensus that researchers should incorporate a probabilistic distribution around the estimated parameter inputs such as treatment effect, health costs and utility measures as a method to apply a PSA on model outputs.(34) This enables decision makers to better assess the level of parameter uncertainty that is inherent in model output, due to the need to estimate the model input parameters, and to better interpret the level of probability that an intervention will be cost effective.(32). Ideally these data outputs can be plotted by quadrant (scatter plot) or against willingness to pay threshold, using a cost effectiveness scatter plot or a CEAC, to demonstrate the joint uncertainty of the probability that an intervention is "cost effective".(32) The inclusion of PSA has been endorsed by HTA bodies involved in developing standardized guidelines for economic evaluation.(4,30)

These differences in the structure of decision models and their methods highlight the need for researchers to transparently share an adequate level of detail to allow decision makers to select an appropriate model for the

purpose of their economic evaluation.(33,35) These references and recommendations in the critical appraisal were used to set the criteria used to identify existing T2DM models that were considered most suitable for economic evaluation in a diverse, real world BC population with T2DM.

2.2 Methods

2.2.1 Identifying existing T2DM models

In order to generate a list of models available for the economic evaluation of interventions in T2DM an electronic literature search was done in MEDLINE using ((computer simulation OR models, economic) AND (cost-benefit analysis OR Type 2 diabetes/economic) AND (Type 2 Diabetes)) then in EMBASE using ("diabetes AND (computer simulation OR economic model) AND (cost-effectiveness OR economic evaluation))". To complement and enhance the search results the review articles and the "similar articles" function in MEDLINE were used to identify additional references that were not captured via the initial literature search. This search was not anticipated to be exhaustive, due to the variation of terminology used in economic evaluation and decision modelling, however it was intended to produce a snapshot of the various economic models most commonly utilized for T2DM. The search results were grouped by model name, lead authors or research institution to identify the available publications by T2DM model. The University of British Columbia (UBC) library website was used as a resource to access peer reviewed publications that enabled the deeper assessment of the methods used in identified models.

2.2.2 Assessing methods of T2DM models

Due to the large number of economic models available for evaluation of T2DM a maximum of ten (10) were selected for further assessment of their structure and methods. The selection of these T2DM models was intended to include various model structures and representation from different regions to provide a cross section of the methods used in economic evaluation. The models which met the initial screening criteria were then further filtered to identify those which researchers or model developers have demonstrated transparency through the publication of their model methods or their participation in model comparison or validation efforts such as the Mount Hood Challenge.(47, 50) Any model that was only designed for a single CEA, specific diabetes related complication or for which published methods could not be located was excluded from further consideration.

This step to perform an in-depth assessment of the ten selected models was done by reviewing published literature, web sites and meeting reports to identify the methods used in each of these models that were then tabulated to enable this comparison. The methods identified to advise the selection of T2DM models for consideration for further validation exercises were as follows:

a) **Structure** – individual or group simulation, dynamic versus closed cohort, type of baseline input variables.

b) **Validation -** sources of evidence used for internal and external validation, methods for model calibration.

c) **Risk Factors** – incorporated to influence the rate of diabetes progression or incidence of events in T2DM such as time varying risk factors and pre-existing events.
e) **Uncertainty** – the type of sensitivity analysis to enable measures of uncertainty in model output.
f) **Output –** breadth of diabetes related complications, health related cost and QALY, discount and adherence rates and economic output measure provided in a CEA.

2.2.3 Selection of a T2DM model for validation

As per the critical appraisal, the selection of the T2DM was based on the perspective of the Ministry of Health for the comparative evaluation of an intervention to reduce the long term risk of complications in those already diagnosed with T2DM. The model baseline input parameters needed to include relevant demographic variables for simulation of T2DM such as age, gender and ethnicity. The model also needed to allow for the inclusion of known risk factors (BMI, smoking,) and the ability to reflect the changes over time of metabolic measures of disease control such as blood glucose, blood pressure and cholesterol in people with T2DM.(27) The model output ideally included the early onset of vascular conditions such as albuminuria and neuropathy or at a minimum had incorporated the common diabetes related vascular events including IHD, MI, stroke, CHF, amputation, blindness and ESRD.

There were some common methods identified in this process that were considered critical for performing a credible and relevant economic evaluation for use in BC. Considering the wide range of risk factors in people with T2DM the models which were designed to include individual microsimulation in the methods were preferred over a more simplistic cohort structure. Also, for the purpose of performing a comparative evaluation of an intervention versus a standard of care, a closed cohort model was selected as the appropriate structure as these types of analysis do not need to incorporate the epidemiological changes in the T2DM population. Finally, the T2DM model required the ability to incorporate probabilistic distribution and provide a PSA in the model output in the form of scatter plots and CEAC to ensure decision makers were able to assess the uncertainty in the CEA. Therefore the selection of an existing T2DM model for additional validation exercises was based on the following five criteria:

1) The ability to perform **micro-simulation** to account for a wide range of risk factors in individuals.
2) A **closed cohort** structure, which models a set number of subjects over time, with baseline input by age, gender and the ability to include **ethnic populations** known to have differing diabetes event rates.
3) The inclusion of **pre-existing events and time varying risk factors** known to influence vascular events.
4) The ability to provide economic output based on **probabilistic sensitivity analysis** including the ability to provide ICER, scatter plots and CEAC as model output for assessing joint uncertainty.
5) Models identified that were available as **open access** for academic purposes.

2.3 Results

2.3.1 Identifying existing T2DM models

An electronic literature search using MEDLINE returned 482 results of articles containing economic data on the cost effectiveness of diabetes interventions. From these search results the T2DM models were screened further by selecting those used for multiple economic evaluations or had participated in validation exercises and had published model methods. The final selection of ten models was completed to ensure a wide geographical distribution of T2DM models from multiple regions which resulted in a number of models being excluded from the assessment exercise which may have been good candidates (Table 1). The IMS CORE, CDC Diabetes, Cardiff and IHE ECHO models were most widely referenced in this literature search for economic evaluation in T2DM.(36,37,39,42) The reports from the Mount Hood challenges identified additional models including the MMD, UKPDS-OM1, ODEM, Archimedes and EBMI models.(37,40,43,47,48,50,59) The Australia model was selected for inclusion of a dynamic population model, which incorporated epidemiologic changes over time, simply to provide a comparison of the methods utilized in the more commonly available closed cohort models.(45)

Table 1: Overview of selected models: country of origin, model source and references for methods

Model Name	Acronym	Country	Source Model	References
IMS CORE (Center for Outcomes Research) Diabetes Model	CDM	Switzerland	Original – based on UKPDS RCT	(22,36,50)
US Centers for Disease Control and Prevention (CDC), Research Triangle Institute Diabetes Model	CDC-RTI	USA	Original – based on Framingham and UKPDS RCT	(37,49)
University of Michigan Model for Diabetes	MMD	USA	Original based on UKPDS RCT and WESDR	(38,48)
Cardiff Diabetes Model	Cardiff	UK	Based on Eastman with updated UKPDS RCT	(39,47,50,54)
United Kingdom Prospective Diabetes Study – Outcomes Model 1	UKPDS-OM1	UK	Original based on UKPDS RCT	(40,52,60,61)
Ontario Diabetes Economic Model	ODEM	Canada	Based on UKPDS-OM1	(41,59,62)
Economic and Health Outcomes Model for type 2 diabetes mellitus	ECHO-T2DM	Sweden	Based on the NIH data, CDM/ DiDACT models	(42,63)
Archimedes Diabetes Model	Archimedes	USA	Original – Kaiser diabetes registry	(43,44,53,64,65)
The Evidence Based Medicine Integrator Simulator	EBMI	USA	Original – Kaiser diabetes registry	(50)
Australia Diabetes Model	Australia DM	Australia	Original – Australian diabetes databases	(31,45,51,66)

UKPDS = United Kingdom Prospective Diabetes Study, RCT = randomized clinical trial, WESDR = Wisconsin Epidemiologic Study of Diabetes Retinopathy, NIH = National Institutes of Health, DiDACT = Diabetes Control and Complications Trial

2.3.2 Assessing methods of T2DM models

The methods of the ten models selected were listed and populated from publicly available data including meeting reports, model manuals and peer reviewed literature to source their methods (Table 2). These results were

compiled into five categories to allow further comparison of the model methods in terms of their structure, validation efforts, risk factors, measures of uncertainty and model outputs for the purpose of economic evaluation. Each of these five categories, further expanded upon in the following sections, are tabulated to allow a more in-depth assessment of the similarities and differences of model methods to guide the final model selection for further validation exercises.

Table 2: **Methods for assessment of ten T2DM models used in economic evaluation**

Methods	CDM	CDC-RTI	MMD	Cardiff	UKPDS OM1	ODEM	ECHO-T2DM	Archi-medes	EBMI (48)(47)	Australia
Structure: 1) Markov 2) Monte Carlo 3) hybrid 4) DES, **Cohort:** dynamic/open (O) or closed (C), **Input variables:** 1) age 2) gender 3) ethnicity										
Structure	3	1	3	4	2	2	3	4	4	1
Cohort	C, O	C, O	C	C	C	C	C	C, O	C,O	O
Demographic input	1-3	1-3	1-3	1-3	1-3	1-3	1,2	1-3	1-3	1-3
Validation 1) RCT 2) observational N) none, **Calibration** (Y/N)										
Internal	1	1	1,2	1	1	1	1	1	2	2(?)
Cross	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
External	1,2	1, 2	1,2	1,2	N	N	1,2	1,2	N	2(?)
Risk Factors: TVRF (Y/N), Pre-existing events (Y/N)										
TVRF	Y	N	Y	Y	Y	Y	N	Y	Y	Y
Pre-existing	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Uncertainty: (Y/N)										
DSA	Y	Y	Y	Y	Y	Y	Y	Y	?	Y
PSA	Y	Y	?	Y	Y	Y	Y	N	?	N
Output: Outcomes 1) micro 2) macro 3) adverse event 4) weight change 5) prevalence, **Cost** Y/N, **QALY** 1) HUI3 2) EQ-5D 3) SF36 4) other, **Discount** 1) cost 2) QALY, **Output** 1) ICER 2) scatter plot 3) CEAC 4) NMB										
Outcomes	1,2	1,2,4	1,2,4	1-4	1,2	1,2	1-3	1-4	1,2	1,2,5
Cost	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
QALY	2	3,4	4	2,3	2	2	2	2	Y ?	4
Discount	1,2	1,2	1,2	1,2	N	1,2	1,2	1,2	?	1, 2
Adherence	N	N	Y	N	N	N	N	Y	Y	N
Economic output	1- 4	1	1,?,?	1,2,3	1,2,3	1- 4	1-4	1	1,?,?	N

DES = discrete event simulation; RCT = randomized clinical trial; TVRF = time varying risk factors; DSA = deterministic sensitivity analysis; PSA = probabilistic sensitivity analysis; QALY = quality adjusted life years; HUI = Health Utilities Index; EQ-5D = EuroQol Five Dimension; SF36 = Short Form 36; ICER = incremental cost effectiveness ratio; CEAC = cost effectiveness acceptability curve; NMB = net monetary benefit

2.3.2.1 Structure

The majority of these models have incorporated individual micro simulation into their methods however the approach used does vary. For instance the UKPDS-OM1 and ODEM models both describe their models as a Monte Carlo simulation using risk equations to determine whether an event occurs over a cycle. The CDM, MMD and more recently the ECHO-T2DM are based on a hybrid approach of Markov states that incorporate micro simulation to enhance the ability to reflect heterogeneity of the individuals with T2DM. In comparison, the Cardiff, EBMI and

Archimedes use a form of DES which allows an event to occur at any moment, as opposed to the number of event over a given time. While these approaches differ they all allowed the incorporation of individual differences over time with T2DM risk factors and greater flexibility in simulating individual progression of a complex chronic disease. The remaining two models, the CDC-RTI and Australia DM were originally structured as straight Markov models have been utilized to measure interventions earlier in the disease process to assess primary prevention diabetes programs.(37,42,66) While they can incorporate time since diagnosis, they treat the cohorts in each state as a homogenous group which does not reflect inter-patient variability or the complexity of diabetes in a diverse population.(57)

Cohort: dynamic versus closed

Further to the basic model structure is whether a model is designed as a dynamic "open" model to incorporate the changing incidence and prevalence of T2DM population over time or as a closed model which simulates T2DM as a set cohort over time. With the exception of the Australia DM, the majority of the models reviewed utilized a "closed-patient cohort" to assess the incremental cost and consequences for a set population.(56) Three of the models the CDC-RTI, Archimedes and the EBMI also allow an open model approach by incorporating epidemiologic data into the simulation. While dynamic open models may be valuable for assessing the overall burden of disease, or the impact of primary prevention interventions in the population, the changing epidemiology of T2DM are not generally incorporated into an economic evaluation of a tertiary intervention versus a standard comparator.(37, 49,61)

Demographic inputs

The key non-modifiable risk factors that impact disease transitions in diabetes, include age, gender and ethnicity.(32,62) All ten models reviewed included the stratification by age and gender as a baseline input parameter for the simulation of T2DM. In comparison, ethnicity was less clearly defined as a baseline input parameter and was only mentioned in the literature for seven of the models reviewed. For instance, the CDM, CDC-RTI and ECHO-T2DM models did not report ethnicity as a baseline input variable in their model design or validation and the Cardiff and Australia DM models mentioned race as a baseline variable however neither model specified the ethnic options for selection. Furthermore, in the MMD, UKPDS-OM1 and ODEM the ethnic inputs limited the selection to either Caucasian or black. The DES models such as Archimedes and EBMI suggested their models have the ability to incorporate race to create an ethnically diverse, simulated population. Archimedes is the most advanced T2DM model for incorporating ethnicity based on the ability to select whether an individual in Hispanic American, African American, Native American, Asian American or white in the baseline input variables.

Overall, the models with methods that incorporated individual micro-simulation preferred due to their ability to better capture complexity and heterogeneity of people with T2DM including pre-existing and time varying risk

factors. In addition, considering the objectives of this thesis was to select models for the economic evaluation of tertiary interventions, the T2DM models which used closed cohort simulation in people diagnosed with T2DM were selected as these provide the desired CEA output. Ideally, the T2DM model used in a diverse population like BC should also allow the inclusion of ethnic sub groups known to have a unique onset and progression of diabetes in a population to improve the generalizability of the model output.

2.3.2.2 Validation

Internal validation

All the models in this review, with the exception of the Australian DM, had published the results of internal validation efforts to demonstrate the ability of the model to replicate the outcomes from the datasets used in their development. The majority of these diabetes models were based, at least in part, on RCT based vascular event rates from the landmark UKPDS study including the CDM, CDC-RTI, MMD, Cardiff, UKPDS-OM1, ODEM, ECHO-T2DM and Archimedes models. Only the EBMI and Australia DM had no published data on internal validation against large scale RCT, but instead relied solely on observational datasets to create and verify model function such as the Kaiser Permanente diabetes registry and Australia epidemiologic datasets. A number of the T2DM models incorporated additional data from other large scale RCT's including people with Type 1 and Type 2 diabetes (Archimedes, ECHO-T2DM) or from observational datasets such as the CDC database (CDC-RTI), Framingham (CDM) and WESDR (MMD). Further validation is required before it is known whether the comparison to observational data will improve the accuracy of T2DM models to predict diabetes related events in the real world.

Cross validation

All T2DM models reviewed, again with the exception of the Australia DM, have been compared by researchers using a reference dataset via the numerous Mount Hood challenges.(47,50) In these organized events where a structured analysis of model performance has been completed, the researchers have input similar baseline characteristics into each model and the outcomes were simulated to determine how well the models reflected trial outcomes (external validity) and how well they correlated to each other (cross validity). While no reference model has been selected as a benchmark for other T2DM models in the cross validation exercises the results of these exercises show there are discrepancies between model outputs. This is in part due to differing definitions of diabetes related outcomes, uncertainty in parameter estimates and differences in the methods related to pre-existing and time varying risk factors, which can all impact the absolute risk of predicted events over time between different models.

External validation

It is relevant to note that the publication of external validation for these T2DM models increased markedly since 2013.(36,38,39,42,52). There was however a lack of consensus on which correlation or statistical measures researchers should be used when assessing the quality or credibility of model output or to quantify the changes required to update the rate of vascular events or disease progression based on the results of the external validation.(69–71) An assessment of the ten models selected showed that eight had evidence of external validation and were primarily compared against outcomes of large scale RCTs (Table 3). Seven of the models claimed that their external validation was performed against observational or registry data; however, not all these analyses have been published to allow transparency of review. Some models have been internally validated or calibrated to observational data, such as the ten year open label extension study of the UKPDS RCT (39,42,52) or a diabetes database, which may increase their credibility for real world economic evaluation.(49–51) At the time of the search, there was no literature referencing the use of observational datasets for the ODEM, UKPDS-OM1 and EBMI in an external validation exercise. There is therefore limited evidence that has assesses the vascular event rates in a Canadian sub populations that are known to have a higher prevalence of T2DM, or greater risk of complications, such as those of advancing age or ethnicity.

In external validation exercises in the selected T2DM models there were a number of different methods that have been used to measure the goodness of fit of the predicted model output to the observed data (Table 3).(53,72) The approach was not consistent and varied depending on the model structure and type of output (IE: parametric, proportional hazards, survival curves, linear or log linear regression) to be used in comparison with the external data.(72) The **Coefficient of Determination (R^2)** was the method most consistently utilized to quantify the linear correlation of the observed versus predicted data in the external validation of these models. This R^2 measure represents the proportion of the variance in the outcome which can be explained by the input parameters in the regression equation and is typically used as a relative measure of correlation and were not referenced with an absolute threshold to demonstrate model validity. To provide some guidance, one source suggests a correlation coefficient above 0.7, which when squared to equal R^2 of 0.49, was set as a threshold to demonstrate positive linear correlation of the predicted versus observed model output.(73) Correlation Plots are often used in conjunction with the R^2 to graph the observed versus predicted endpoints with a point falling on the 45 degree line, or an R^2 of 1.0, was reflective of an exact match between the predicted and observed outcomes. While the R^2 provided an objective measure of the goodness of fit of model output these did not include or assess statistical significance of the model fit. Therefore, additional methods were included in the validations to assess the statistical significance of the output.

There were a number of statistical measures used in the validation of the ten selected models however none were consistently used by all researchers. Another measure used in multiple model validations was the **MAPE** (mean

absolute percentage error) which provided a measure of the predictive accuracy of a model by calculating the sum of the differences between actual and forecasted values divided by the number of points fitted. The MAPE was used as a measure in the CDM, Cardiff and UKPDS-OM1 model validations as it was considered to be an easily interpretable measure of the percentage error. That said, the MAPE should be used with caution when interpreting low volume outcomes which occur infrequently, such as ESRD, as the calculation can produce widely variable extreme values in response to a small absolute changes in values which reduces its usefulness in this analysis. The associated RMSPE (root mean square percentage error), calculated by squaring the percentage errors, averaging and square rooting to create a measure of fit and was only used by the CDM. Only the CDM used the paired t-test, a statistical measure comparing two sets of output data in the same subjects, which was not a common measure used in external validations based on the validation exercises which were completed in these T2DM models. The UKPDS-OM1 used a variety of statistical measures including a **C-statistic,** described as a rank order statistic or measure of the area under the curve (AUC), to assess the goodness of fit of a logistic regression model. This test provides a measure between 0 (no fit) and 1 (perfect fit). The general interpretation is that a C-statistic above 0.7 is required to demonstrate a model has value, and above 0.8 reflects a strong model. The C-statistic provided a quantitative measure for comparison to other models but it did not provide an absolute measure of statistical validity of the model output. The validation of the UKPDS-OM1 also used a simple comparison of the predicted incidence of events from the model with the mean and CI of the observed events. This was an intuitive and quantifiable measure of statistical relevance of the model output for each of the diabetes related events predicted by the UKPDS-OM1. The UKPDS-OM1 and the Archimedes researchers also used a **log rank test** as a quantitative statistical measure of the predicted versus observed data. A key assumption for the log rank test, when comparing survival curves, is that there are proportional hazards over time and data can be from non-parametric distribution. The log rank tests the null hypothesis of no difference between the curves by utilizing hazard functions to calculate a test statistic (X^2). Log rank tests were used to compare survival curves of predicted versus actual output, such as **Kaplan-Meier curves**, to test the statistical validity of the model. The ECHO-T2DM validation was the only model that utilized an F-test, which is a ratio of the explained (measured by R^2) versus the unexplained variability (or 1 - R^2), to quantify the usefulness of the model. A higher F-test score in the validations suggested greater fit of the model but were not considered a statistical measure of the models predicted events versus observed events.

The selection and inclusion of an appropriate statistical measure of predicted versus actual model output was determined to be an important consideration to provide quantifiable measures of the model output to be considered validated for use in the population of interest. While no single measure or approach was used to demonstrate whether a model was considered as validated, the inclusion of a statistically based validation exercise in addition to measures of correlation added further confidence in the validity of model output.(53)

Table 3: Published T2DM model validations, source of evidence and statistical measures

Model name	Year of publication	Source of evidence	Measure for model validity
CDM (36)	2014	RCT, Obs	C of D (R^2), paired T test, RMSPE, MAPE
CDC-RTI (49)	2009	RCT, Obs	C of D (R^2)
MMD (38)	2015	RCT, Obs	C of D (R^2),
Cardiff (39)	2015	RCT, Registry	C of D (R^2), MAPE
UKPDS-OM1 (52)	2013	10 year Obs	C of D (R^2), MAPE, C statistic, log rank test (X^2), predicted versus the mean and CI of observed
ODEM	N/A	None	N/A
ECHO-T2DM (63)	2013	RCT, Obs	C of D (R^2), F test
Archimedes (44)	2003	RCT	C of D (R^2), log rank test (X^2),
EMBI	N/A	None	N/A
Australia DM (66)	N/A	Obs	N/A

RCT = randomized clinical trial; Obs = observational data; C of D (R^2) = coefficient of determination; RMSPE = root mean squared error; MAPE = mean absolute percentage error; N/A = not available

2.3.2.3 Uncertainty

Deterministic sensitivity analysis were utilized and reported on by all T2DM models in this review with the exception of the EBMI. This is a common approach utilized by researchers and decision makers to assess the uncertainty of model input parameters. The ability to perform DSA was not a differentiating criterion in the analysis of model methods as it is a standard feature that can be applied to all T2DM models simply by adjusting model parameters and assessing the change in model output.

Probabilistic sensitivity analysis requires the incorporation of probabilistic distribution of model parameters and was not commonly reported in CEA research or model validation studies. There was mention of PSA incorporated into the methods of the CDM, UKPDS-OM1, ODEM, ECHO-T2DM and Cardiff or reported in the CEA output. The user manual of the MMD described an ability to select probabilistic distributions for model parameters however PSA was not described in any of the published economic evaluations which utilized the MMD. There were no published studies identified which had used the EBMI and therefore it could not be determined whether PSA was incorporated into the model methods. It was therefore unclear from the literature whether the MMD or the EBMI have incorporated PSA into their methods. The publications of the CDC-RTI, Archimedes, EBMI and Australia DM did not report the use of PSA in their methods or CEA outputs.

There was limited information in the literature on the type of probabilistic distributions which were utilized for PSA in these T2DM models; however the ability to incorporate PSA into the output was a distinguishing feature in a number of the T2DM models. The inclusion of PSA in the model methods should be considered a standard requirement for any T2DM model used by decision makers to guide health resource allocation as it enables researchers to provide a measure -of the uncertainty of model output.

2.3.2.4 Risk factors

Time varying risk factors (TVRF)

The long term studies in T2DM demonstrated there was a gradual increase in biologic measures of disease control such as HbA1c, systolic blood pressure (SBP) and TC levels observed over time.(27) This trend has been incorporated into the methods of these ten T2DM models with the exception of the CDC-RTI. The Cardiff, Archimedes and Australia DM models have also incorporated an increasing BMI over time in those with T2DM. In addition to CEA of a tertiary intervention, this change in BMI would also be relevant for an analysis of a primary prevention intervention due to the positive correlation between BMI and the incidence of T2DM. It was not determined whether the T2DM models which incorporated BMI as a TVRF was better able to predict the risk of diabetes related vascular events over a longer term simulation. It would be insightful for future research to include an assessment to determine if TVRF's lead to a statistical improvement in a model's predictive accuracy versus observed events.

Pre-existing risk factors

The CDC-RTI and the Australia DM have not incorporated pre-existing risk factors into the disease transition calculations. All other models incorporated a standard range of pre-existing risk factors such as smoking and history of vascular events into the baseline input parameters. The UKPDS-OM1 and the ODEM specifically mentioned the inclusion of A-fib and PVD as baseline risk factors while the ECHO-T2DM mentioned the inclusion of A-fib. Other models may have included A-fib and peripheral vascular disease however these were not identified in the literature reviewed.

2.3.2.5 Output

Disease measures

The comparison of model methods included an overview of which of the T2DM model outputs included the early signs and symptoms of disease (ie: albuminuria) or only the diabetes related vascular events (ie: ESRD).(55) Models that have captured the early signs of vascular disease, such as albuminuria, neuropathy or retinopathy, could be expected to provide a more complete picture of the absolute costs and quality of life, versus only capturing the vascular event; however it is unclear whether this would alter the relative comparative measures used in a CEA of an intervention. While the inclusion of these early signs of vascular disease may be desired for the purpose of completing a robust economic evaluation of T2DM this will also add complexity and require additional data to capture the early signs of vascular disease.(56) While these early signs of vascular disease will be relevant in an economic evaluation they can alternatively be captured as part of the ongoing cost and QALY measures associated with T2DM management (Table 4).(40)

Eight of the assessed T2DM models incorporated the main diabetes related micro vascular (ie: amputation, ESRD and blindness) and macro vascular events (ie: IHD, MI, stroke, CHF) with the exception of CDC-RTI and the Australia DM which have not included CHF in the vascular event output. The CDM, ECHO-T2DM and Archimedes models captured the broadest range of diabetes related outcomes including the early signs of vascular disease which may offer an advantage to decision makers for the purpose of an economic evaluation. While the Cardiff, UKPDS-OM1, ODEM and EBMI models have not included measures of the early signs of vascular disease they all predict the seven main vascular events for the economic evaluation in T2DM.

Table 4: Overview of T2DM related vascular outcomes predicted by the ten models

Outcome	CDM	CDC-RTI	MMD	Cardiff	UKPDS-OM1	ODEM	ECHO-T2DM	Archimedes	EBMI	Australia DM
MI	X	X	X	X	X	X	X	X	X	X
IHD	X	X	X	X	X	X	X	X	X	
Stroke/CVD	X	X	X	X	X	X	X	X	X	X
CHF	X		X	X	X	X	X	X	X	
Amputation	X	X	X	X	X	X	X	X	X	X
Blindness		X	X	X	X	X	X	X	X	X
ESRD	X	X	X	X	X	X	X	X	X	X
Albuminuria	X	X	X				X	X	X	X
Proteinuria	X	X	X				X	X		
Neuropathy	X	X	X				X			X
Skin ulcers	X						X	X		X
Retinopathy	X	X	X				X		X	
Macular edema	X		X				X	X		
Cataracts	X							X		
Hypoglycemia	X			X			X	X		

X = outcome predicted by model, [ ] = outcome not predicted by model,
MI = myocardial infarction; IHD = ischemic heart disease; CVD = cerebrovascular disease; CHF = congestive heart failure; ESRD = end stage renal disease;

Costs

Costs associated with treatment of T2DM and related events were a core component of all T2DM with the exception of the initial version UKPDS-OM1 which only provided a predicted disease transition or vascular event rates in T2DM.(40) The UKPDS-OM1 has since been updated with revisions that incorporated costs to allow its use for economic evaluation.(60) While closed cohort decision models typically capture the "incremental cost" of an intervention presented as an ICER, scatter plot and CEAC, it was noted that the methods used to capture costs in these T2DM models varied depending on the structure of the model. For instance, the cohort based CDC-RTI, ECHO-T2DM and Australia DM models calculated costs based on the average cost of the resources used to manage each cohort over a cycle and included the cost of diabetes related event and the ongoing treatment of T2DM. The CDM, MMD, Cardiff and ODEM micro simulation models offered more flexibility in tracking and timing of resource utilization by per individual. Archimedes may have an advantage over other T2DM models when detailed health utilization and cost data is available as it was designed to simulate a highly customized path of the resources

utilized for each individual.(43) The models used for CEA must be updated to the costs in the currency of the region in which the model will be utilized to reflect the local standard of care and health resource utilization.

Health utility

The incorporation of health utility measures was used for the calculation of the QALY by multiplying a utility value to the life years gained or lost with an intervention. This is a critical component to the measure of the incremental effectiveness in the ICER. The use of utility measures allows model output measures of effectiveness to be compared across different health conditions as it is a generic measure of health outcomes.(34)

Of the ten models reviewed, six have used the utility measures based (all or in part) on the EuroQol Five Dimension Questionnaire (EQ-5D) including the CDM, Cardiff, UKPDS-OM1, ODEM, Archimedes and ECHO-T2DM models. The Short Form 36 (SF-36) was the next most commonly utilized health utility measure that was incorporated into parts of the CDC-RTI and Cardiff models. The MMD and Australia DM models used their own version of health utility measures in the calculation of the CEA output from these models.

The EQ-5D measures of health utility for diabetes related complications or events in these T2DM models were mostly based on data from the RCT portion of the UKPDS study which reflected utility measures in a UK population treated to standards prior to 1996. (74) The UK has since issued guidance that recommended the EQ-5D as the preferred instrument to measure health utility for the calculation of QALY in a CEA.(75) There was research that validated the Health Utilities Index Mark 3 (HUI3) as a basis to measure QALYs in Canadians with a broad range of conditions, including diabetes, however it was not utilized in the QALY calculations in any of the T2DM models assessed.(76)

There was not a standardized approach in the use of health utility measures in T2DM models for the calculation of QALY which could lead to differences in the ICER from CEA performed in similar populations.(77,78) The T2DM models utilizing the EQ-5D measure of utility should be updated to reflect the Canadian utility measures when performing CEA in Canada and models using another utility measure should ideally be compared against the QALY measures using EQ-5D to assess uncertainty from differing sources of utility measures.

Discount rate

The discount rate has been incorporated into many of the model methods to reflect the present day value of an intervention measured over multiple years.(34) All models, with the exception of early versions of the UKPDS-OM1, allowed the application of discount rates to both the measured cost and QALY. The UKPDS-OM1 has been updated to add costs and has included discount rate in the model. There was limited data available to fully assess

the EBMI on whether discount had been incorporated into the model. The models could not be differentiated based on whether their methods included an ability to apply discount rates in a CEA.

Adherence

The vast majority of models were designed and validated using data from RCT's the relative benefit of treatment are likely to reflect adherence rates well above those demonstrated in real world studies.(79) Of the ten models reviewed, only the MMD, Archimedes and EBMI have mentioned the ability to incorporate real world adherence rates into a model simulation. For the remaining T2DM models which do not have measures of adherence in their methods, the effectiveness of the intervention in the real world would have to be adjusted via a sensitivity analysis to assess the impact of lower adherence on the ICER output.

Economic output

For the purpose of providing output of an economic evaluation to HTA and decision makers it was recommended to provide at least a point estimate of the ICER and ideally include a cost effectiveness scatter plot and a CEAC to assess the probability an intervention is cost effective and the joint uncertainty of model output.(4) All of the T2DM models, with the exception of the Australia DM, provided an ICER as a standard output in a CEA. Of the remaining nine models, the CDM, Cardiff, UKPDS-OM1, ODEM and ECHO-T2DM have scatter plots and CEAC included in the publications of model output. It could not be determined from the literature whether the MMD or the EBMI provided additional output measures beyond an ICER. These additional model outputs were determined to be important methods in the criteria for consideration for selecting a T2DM model for CEA.

2.3.3 Selection of a T2DM for validation

The criteria for screening these T2DM models was based on a perspective of the Ministry of Health for the comparative evaluation of tertiary interventions in a population already diagnosed with T2DM. It is important to highlight that the results of this screening exercise were based on what had been identified in the published literature at the time of the assessment these results may not be fully reflective of each model's current capabilities.

These analyses demonstrated that there were many T2DM models which have been developed and evaluated, in some cases over many years, using quality evidence and numerous model validations to evolve and enhance the T2DM model functionality and credibility. The methods used in the T2DM models were screened in this research by applying five objective criteria from the recommendations in the critical appraisal section which enabled the selection of appropriate T2DM models which were then considered for further validation as outlined below (Table 5):

1) The model must be based on a closed cohort structure with age, gender and ethnicity as key baseline input variables for use in diverse T2DM populations: All models with the exception of the Australia DM had the methods to perform a closed cohort simulation. The requirement to allow the selection of ethnicity as a baseline input was not as clearly identifiable and there were limited ethnic options even in T2DM models which had included this in their methods. To apply this, models which did not have any ethnic options in the methods were excluded namely the CDM and ECHO-T2DM. After applying this initial criterion the CDM, CDC-RTI, MMD, Cardiff, UKPDS-OM1, ODEM, Archimedes and EBMI remained for further consideration.

2) The model methods must have included an option to run individual **micro-simulation**: The T2DM models which have been developed using a Markov based cohort design, such as the CDC-RTI, were therefore excluded. The seven of the remaining models, the MMD, Cardiff, UKPDS-OM1, ODEM, ECHO-T2DM, Archimedes and EBMI were based on either a Monte Carlo, or a hybrid of Markov and Monte Carlo, or DES methods which were all considered adequate to allow individual microsimulation.

3) The model methods must have allowed the inclusion of pre-existing events and TVRFs: Of the remaining seven T2DM models, all were determined to include the pre-existing risk factors believed to influence the future risk of events, with the exception of the ECHO-T2DM, such as such as history of vascular complications and smoking. An important consideration to be included in the model methods was the ability to capture changes in TVRF's such as blood glucose, blood pressure and blood lipids due to their known impact on the rate of future diabetes related events. Therefore, the remaining six T2DM models continued to be considered based on meeting these criteria.

4) The researchers must have included PSA in model methods and the capability to generate a CEAC for assessing joint uncertainty of model output: There was no documentation in published literature of the MMD, Archimedes or the EBMI models that identified methods which included PSA or the creation of a CEAC as part of the model output in a CEA. Of the remaining T2DM models there was evidence of the use of PSA and the creation of CEAC in the literature for the CDM, Cardiff, UKPDS-OM1 and ODEM which resulted in these being the preferred models for follow up to request open access.

5) The T2DM models needed to be available as open access for academic purposes: The CDM, Cardiff, the UKPDS-OM1 and the ODEM were the only models which had documentation of methods which supported all of the four previous criteria. The research sites of these model developers were searched to determine whether their models were offered for academic research purposes at no charge. The UKPDS-OM1 and the ODEM were identified as open access models and the researchers at Oxford University (UKPDS-OM1) and McMaster University (ODEM) were contacted to request access to their T2DM model with an explanation of the intended validation exercises.

The ODEM was the only model which was provided as a fully functioning, open access T2DM model and was the therefore selected for further internal and external validation.

Table 5: Model selection using five objective screening criteria

Model name	Cohort Closed/ Ethnic	Individual simulation	Risk Factors Pre/ TVRF	PSA/ CEAC	Open Access
CDM					
CDC-RTI					
MMD					
Cardiff					
UKPDS-OM1					
ODEM					
ECHO-T2DM					
Archimedes					
EBMI					
Australia DM					

Red box – methods did not meet stated criteria, Green box – methods met stated criteria, Orange – model updated, depends on version used
TVRF = time varying risk factors; PSA = probabilistic sensitivity analysis; CEAC = cost effectiveness acceptability curve

2.4 Discussion

The selection of a T2DM model was a result of a step wise approach to identify a list of available T2DM models. A preliminary screening was then applied to select group of ten T2DM models for deeper assessment of the model methods. Finally, five objective criteria were applied to identify which of these models were designed with the methods considered suitable for the economic evaluation of a closed cohort from the perspective of the Ministry of Health. These criteria were based on the guidance provided by the reference text and ISPOR-SMDM reports as an objective way to interpret the methods used in the T2DM model. This approach builds upon these recommendations for transparency in the selection of model methods and importance of ongoing validation against data from the population of interest to improve trust and confidence in the output of decision models used for Canadian HTA.(35,53)

There were a number of models which have an appropriate structure and methods to justify their selection and use in a diverse BC population. Some of the models may also have new methods incorporated in more recent versions which were not identified in the available literature. While the ODEM met the five selection criteria there were some methods which could be enhanced with new evidence or revisions to the ODEM. For instance, the ODEM does not record the early sign of vascular disease, adverse events of treatment or provide a mechanism to adjust for lower adherence to treatment. The ODEM is a model which has been used for the economic evaluation of a number of T2DM interventions in a Canadian health care setting however the risk equations which determine the rate of events remain based on the original data from the UKPDS RCT. One distinguishing feature of the ODEM was that the health care utilization and associated costs have been updated by researchers, using a validated

Ontario Diabetes Database (ODD), to reflect the Canadian results for the ongoing management of T2DM. To determine if the output of ODEM output can be considered generalizable for use in the BC health system requires an internal validation of its functionality and an external validation of its ability to predict diabetes related outcomes in a real world BC population with T2DM. These were the primary aims of the next steps of this thesis research.

Chapter 3: Internal validation of the ODEM

3.1 Introduction

The search of available models in T2DM, assessment of their methods and the eventual selection using a set of objective criteria resulted in gaining access to a model which met the needs for use in a Canadian HTA. In order to build trust and confidence in the output of an economic evaluation using the ODEM model required additional work on model validation. The ISPOR-SMDM report on "Model Transparency and Validation" highlighted the different internal validation techniques used to assess the models ability to reflect disease progression. These reports encouraged transparency for others to interpret methods and assess the uncertainty of the output of a decision model.(53)

The first step in model validation was to perform an in-depth assessment of the performance of the ODEM. The ISPOR-SMDM recommendations highlighted different methods for internal validation either aimed to assess the mathematical functionality of the model with data used to build the model or as a verification of the impact of changes to input parameters on model output.(53) This analysis used the latter approach by observing changes in model output when adjusting individual input parameters. This model validation will centre on the performance of the risk equations in ODEM to determine model functionality and deepen the working comprehension of ODEM's methods.

3.1.1 Structure

The ODEM is a probabilistic discrete-time, state transition, Monte Carlo micro simulation model based on the methods from the UKPDS-OM1.(27) The diabetes related events are generated each cycle using regression equations which are based on characteristics of each simulated individual. The ODEM uses parametric proportional hazards (Weibull) to incorporate demographic characteristics (age, gender and ethnicity), time varying and pre-existing risk factors (HbA1c, SBP, cholesterol, A-fib, PVD) and history of diabetes related events to predict the cumulative incidence of events for up to a 40 year time horizon.(41)

3.1.2 Model assumptions and validation

The ODEM is designed as a closed-cohort simulation, meaning that it used a set number of individuals over the time span of the simulation, and does not account for changes in the incidence or prevalence of T2DM. The model simulates a control and treatment arm to compare the incremental cost and consequences of an intervention versus the existing standard of care using a Ministry of Health perspective. The methods of ODEM incorporated random number sampling, TVRFs and PSA on treatment effects, costs and QALYs to generate model output. This output included an ICER as a base case estimate of the cost effectiveness and a CEAC to provide a probability that an intervention is cost effective at different thresholds that are based on the maximum cost decision makers are

willing to pay for a year at full health. The scatter plot and CEAC are valuable as model outputs as they allow an assessment of the joint uncertainty of the ODEM input parameters for decision makers.

A critical component to the credibility of the output from the ODEM for use in a real world population in BC depended on the relevance of the risk equations used in the model to determine the cumulative incidence of vascular events. These risk equations and predicted diabetes related events were a primary contributor to the cost and consequences in the economic output of the ODEM and were the methods assessed in this internal validation exercise.

3.2 Methods

This internal validation provided a comparison of the expected impact on the cumulative incidence of events from changing an input parameter, based on the coefficients in the ODEM risk equations, compared to the actual changes observed in events after changing the input parameter. The outputs of interest were the ten year cumulative incidence of seven diabetes related events, diabetes related mortality, and the associated QALYs. This step wise approach enabled a deeper understanding of the functionality of the ODEM and determined whether the risk equations in the ODEM were responsive and properly linked with the baseline input parameters.

3.2.1 Development of a base cohort

A base cohort was created by defining the base parameters for input into the ODEM to produce a benchmark of the cumulative incidence of diabetes related events, life years and QALYs over 10 years. The base cohort was defined as all Caucasian, male, with a mean age of 65 years and average duration of diabetes (DofD) of 5 years, with a mean BMI of 26.9 kg/m^2. The surrogate measures of disease control were set at an estimated population average for those with T2DM and included HbA1c at 7.28%, mean SBP at 129 mm/Hg, TC of 4.95 mmol/l and high density lipoprotein (HDL) of 1.13 mmol/l. The percent with baseline risk factors of smoking, A-fib, PVD and history of previous vascular events were all set to zero. These baseline input variables were added into the ODEM and simulated to capture the model output for diabetes related events, life years and QALYs for this base cohort.

3.2.2 Statistical evaluation for validation

There was a need for an objective measure to determine whether the magnitude of the difference in the output with each simulation was, due to the change of the baseline input variable or, in response to the natural variability in model output from the use of a random number generation to predict an event in the ODEM simulations. There was not a standardized approach identified in the literature to define an effect size or a statistical measure to compare output in validation or verification exercises.(71,80) The availability of cumulative incidence data for each vascular event data from the ten iterations over a ten year period in the ODEM simulation provided an mechanism to calculate a mean, variance and standard error of the mean (SE) for the seven vascular events in the baseline

cohort (Equations 1 – 3). These data were used to calculate a 95% CI to define an objective measure to assess whether differences in the incidence of the seven vascular events after changing a single input parameter was due to the natural variance of the simulation or if it exceeded the expected range.(70) This was a similar approach as one used to demonstrate the internal validation of the UKPDS risk engine to predict the risk of stroke.(81)

Equation 1: Mean cumulative incidence for each vascular event

$$\mu_o = \frac{1}{N}\sum_{i=1}^{N} x_{iot}$$

Equation 2: Calculation of the cumulative incidence standard error of the mean = σ

$$\sigma_o = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{iot} - \mu_o)^2}$$

Equation 3: Calculation of the cumulative incidence 95% CI

$$95\%\ CI = \mu_o \pm 1.96\ \sigma_o/\sqrt{N}$$

where N = number of years with i = iteration, o = vascular outcome, t = year.

3.2.3 Single parameter variation

The initial step in the internal validation was a qualitative assessment to document the expected direction of change in diabetes related events, life years and QALY's, versus the base cohort results, when adjusting each of the baseline variables. These input variables included the demographic characteristics (ethnicity, gender, age, BMI), the pre-existing and TVRF's at baseline (A-fib, PVD, smoking, HbA1c, SBP, TC and HDL) and the history of diabetes related vascular events (IHD, MI, CHF, stroke, amputation, blindness and ESRD). The level that each variable was set at for each simulation was defined in the table of results, and the expected direction of change of the cumulative incidence of events was predicted based on the coefficients within the risk equations in the ODEM (Table 7). These included both the direct and indirect impact of a variable in the risk equations. A direct impact was an input variable that was included in the risk equation for that diabetes related event, while an indirect impact was defined as a variable that influenced the risk of an event indirectly by increasing one of the direct variables for that vascular outcome. The expected direction of the direct impacts on the model output for each parameter change versus the base cohort were captured using arrows (↑, ↓), and the indirect impacts were captured with a positive (+) or negative (–) sign.

The next step was to change a baseline input variable to an extreme value, run a simulation, and then record the actual cumulative incidence of diabetes related events, life years and QALY's. The size of any differences in the actual events versus the events for the baseline cohort were assessed to determine if the magnitude of change fell within or outside of the lower and upper 95% CI limits of the base cohort simulation. The direction of the actual events was then compared to the expected direction that was recorded in the first step to provide an objective

assessment of the model functionality based on each individual baseline input parameter. These results were tabulated and color coded for ease of comparison.

For the purpose of this validation exercise baseline patient variables were set as a constant so that the point estimate of the input variable was applied equally to all individuals. This approach was used instead of a random number generated range of input variables to reduce the statistical variability in the model input parameters that determined the cumulative incidence of events in the ODEM. Each cohort simulation was done for 1000 virtual patients simulated through the model ten times for a total of 10,000 patient level simulations for ten years.

3.2.4 Expected direct and indirect impact of risk equations

To ensure this assessment reflected the expected changes to baseline variable on model outputs, it was deemed relevant to incorporate both the direct and indirect impact of each variable based on the ODEM's risk equations (Table 6).(40,41) The direct impacts were anticipated to have a greater and more immediate influence on the model outputs compared to indirect impacts, as these indirect variables must first cause a vascular event which is included in the risk equation for a different vascular event. For example, the ODEM the risk equations, an increasing BMI is not included in the risk equation for MI, however it can indirectly increase the rate of MI due to the increased incidence of CHF in those with high BMI. Therefore, an individual with a history of CHF has a direct impact, while BMI has an indirect impact, on increasing the risk of an MI in that individual.

Table 6: The direct and indirect impacts included in the ODEM risk equations by vascular event

Event	Direct Impact	Indirect Impact only
IHD	Gender, Age, TC:HDL, SBP, HbA1c	none
MI	Ethnic, gender, age, smoking, TC:HDL, SBP, HbA1c, **IHD, CHF**	BMI (via CHF)
CHF	Age, BMI, SBP, HbA1c	none
Stroke	Gender, Age, A-fib, smoking, TC:HDL, SBP, HbA1c, **CHF**	BMI (via CHF)
Amputation	PVD, SBP, HbA1c, **Blind**	Age (via blindness)
Blind	HbA1c, Age	none
ESRD	SBP, **Blind**	HbA1c, Age (via blindness)
Mortality	**MI, Stroke, ESRD, Amputation**	IHD, CHF, Blind (via MI, Stroke, Blind and Amputation)

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Blind = blindness in one eye; ESRD = end stage renal disease; TC/HDL = total cholesterol to high density lipoprotein ratio; SBP = systolic blood pressure; HbA1c = glycated hemoglobin A1c; BMI = body mass index; A-fib = atrial fibrillation; PVD = peripheral vascular disease; Age = age at diagnosis of diabetes

3.3 Results

This internal validation showed the ODEM performed as expected for the vast majority of diabetes outcomes. These positive results were represented by the grey boxes, in the right columns, where outputs directionally match and reflect a consistency between expected and actual cumulative incidence of events (Table 7). The yellow boxes highlighted the results where the actual outputs did not respond to changes to the input parameters that were

expected based on a direct or indirect impact. The cumulative incidence outputs, marked in orange and red boxes, highlight where the magnitude of the expected and actual output measures over and under exceeded the 95% CI of the vascular events predicted in the base cohort respectively. The unexpected results of the validation exercise, marked in yellow, orange and red boxes, were then further investigated to assess the reason for the discrepancy in the ODEM output.

The expected impact of input variables on life years and QALY's was also captured in the model to allow an assessment of the risk equations for mortality which included event mortality (death at the time of a vascular event), diabetes mortality, (increased risk of death due to a diabetes related event or complication) and other death (standard death rate for people without T2DM). The measure of QALY's were included to determine if the decrease in health utility due to a diabetes related event was reflected in the calculation and compilation of the overall QALY in each cohort.

3.3.1 Discrepancy between expected and actual model output

a) Yellow boxes

There was no change in the cumulative incidence of stroke in the cohort with a BMI of 36kg/m^2 when compared to the base cohort. This was different than the expected impact as the risk equations in the ODEM reflected a direct increase on the incidence of CHF of approximately 7% for every unit of increase in BMI above 28kg/m^2. In turn, every additional case of CHF resulted in a 5.7 fold increase in the risk of stroke per cycle. The actual output from the ODEM high BMI cohort did increase the rate of CHF however this did not lead to an increase in the rate of stroke over 10 years. The timing of these events was considered relevant as the model requires CHF to occur first before increasing the risk of stroke with elevated BMI. In the simulation, the mean time to a CHF event was 5.7 years and the mean time to a stroke event was 5.9 years. Therefore, in a ten year simulation this left only 4 years for these additional cases of CHF to in turn increase the incidence of stroke. It is feasible to assume that the 10 year time frame may not be adequate and a longer term simulation may be required for this to be assessed.

There was also no change in the cumulative incidence of renal disease in the cohort with a mean elevated HbA1c of 11% when compared to the base cohort with a mean HbA1c of 7.28%. This was different than expected as the risk equations in the ODEM reflected a direct increase on the incidence of blindness of approximately 25% for every percent increase in HbA1c above 7.09%. Based on the coefficients in the model risk equations, every additional case of blindness was associated with an 8 fold increase in the risk of ESRD per cycle. In the ODEM simulation of the high HbA1c cohort, the rate of blindness was double the rate in the base cohort however this did not result in any increase in the rate of renal disease over 10 years. The timing of the event was again considered relevant as the model requires blindness to occur first before increasing the risk of renal disease as a result of elevated HbA1c. In the simulation, the mean time to an event of blindness was 5.2 years and the mean time to

ESRD was 5.9 years. Therefore, in a ten year simulation only 5 years remained for additional cases of blindness to increase the incidence of ESRD. Similar to the discussion on the indirect impact of BMI on stroke, it is feasible to assume that the 10 year time frame may once again not be adequate and a longer term simulation would allow this to be further assessed.

The duration of diabetes (DofD) was a baseline input parameter that was not directly incorporated into the risk equations, but was used to calculate the age of diagnosis of T2DM which is included as part of the risk equations. The actual output of the simulation for the cohort with a DofD of 20 years aligned with what was expected based on the risk equations in the ODEM however the calculation of age of diagnosis in the risk equations meant that those with a longer duration of T2DM have a reduced risk of diabetes related events. This impact was considered counter intuitive to what would be assumed in a test of face validity of the ODEM and these boxes were marked in yellow for further investigation. Considering the ODEM was based on the risk equations from the UKPDS-OM1, the literature for the development of that model was reviewed and this approach was the same in both models. Upon further investigation, it was apparent that DofD was linked directly to age at diagnosis and current age, and therefore if one of these variables changed then others would automatically be adjusted. Therefore, the DofD was dependent on current age and age at diagnosis, which appeared in other parts of the model, including risk equations related to mortality calculations and was a baseline input variable that possibly should not be assessed in isolation.

b) Orange boxes

There was no consistent trend observed in the cases where the cumulative incidence of events was of a higher magnitude than expected and above the 95% CI. Most of these fell just outside of the statistical range set by the above equations and are most likely attributed to random variation in the model output.

c) Red boxes

A consistent trend was observed in cohorts who demonstrated a higher mortality rate, as measured by a decreasing number of life years in the ten year simulation, such as those with 100% MI, CHF, amputation and ESRD at baseline. The total life years in these cohorts were near or below 7 also resulted in a positive correlation with the predicted incidence of vascular events falling below the 95% CI for IHD, MI, CHF, stroke and blindness. This result is most likely due to competing events in the simulation, where more people who die in the ten years, the fewer who are alive to experience an event.

3.3.2 Mortality

The actual change in life years correlated positively with the expected direction of change in all cohorts included in this internal validation. There was a noticeable trend toward a lower cumulative incidence of diabetes related

events in the cohorts with high mortality, resulting in life years of less than seven over the ten year simulation, such as those with 100% CHF, amputation or ESRD at baseline. Upon further investigation, it was likely that the higher mortality rate in these cohorts reduced the incidence of diabetes related events as a result of death causing fewer individuals to be alive to experience an event. This was evident in the cohort of 100% of individuals with ESRD at baseline, which resulted in the lowest measure of life years of 6.0 in a ten year simulation, and the vascular event rate fell below the lower range of the 95% CI. There was an inverse correlation between higher mortality and lower incidence of events that became stronger with each subsequent year as more individuals in the cohort experienced a fatal event.

3.3.3 Quality adjusted life year (QALY)

The actual measures of QALYs also performed as expected based on the changes in each input parameter one at a time by cohort. The changes in QALYs were generally modest, and remained above 6.0 for the cohorts which had fewer incremental diabetes related events such as the high BMI, A-fib and PVD cohorts. The cohorts which had a higher risk of vascular events, such as those with a baseline of all smokers, elevated SBP, TC/HDL ratio and HbA1c had a corresponding decrease in QALY which fell below 6.0. As expected, the impact on QALY was the greatest in the cohorts with 100% baseline history of events as these events are all associated with a utility decrement in the ODEM methods. The impact of 100% CHF, stroke, amputation and ESRD at baseline also created the largest decrease in QALY's, with all cohorts ten year total below 5.0, which is aligned with the larger utility decrements with these specific events in the ODEM QALY coefficients.

Table 7: Results of the ODEM internal validation of single parameter variation

Revised baseline input parameter	Direction of predicted output versus base case (based on risk equations)									Actual output from the ODEM (10 year cumulative incidence)								
	IHD	MI	CHF	Stroke	Amp	Blind	ESRD	LY's	QALY's	IHD	MI	CHF	Stroke	Amp	Blind	ESRD	LY's	QALY's
Base case*	–	–	–	–	–	–	–	–	–	0.058	0.160	0.042	0.046	0.006	0.039	0.004	7.93	6.12
Lower CI										0.053	0.153	0.038	0.041	0.004	0.034	0.002	n/a	n/a
Upper CI										0.063	0.166	0.045	0.051	0.007	0.043	0.006	n/a	n/a
Ethnicity - 100% Afro-Caribbean	0	↓	0	0	0	0	0	↑	↑	0.064	0.048	0.045	0.048	0.007	0.037	0.005	8.13	6.28
Gender – 100% female	↓	↓ –	0	↓	0	0	0	↑	↑	0.041	0.081	0.045	0.028	0.006	0.036	0.006	8.21	6.37
Mean Age - 45 yrs	↓	↓ –	↓	↓	0	↓	0	↑	↑	0.041	0.057	0.006	0.010	0.006	0.010	0.005	8.67	6.75
DofD - 20 yrs	↓	↓ –	↓	↓ –	_	↓	_	↑	↑	0.047	0.097	0.022	0.022	0.008	0.015	0.010	8.44	6.55
Mean BMI – 36 kg/m2	0	+	↑	+	0	0	0	-	↓	0.059	0.172	0.076	0.046	0.006	0.038	0.004	7.91	6.10

Revised baseline input parameter	Direction of predicted output versus base case (based on risk equations)									Actual output from the ODEM (10 year cumulative incidence)								
	IHD	MI	CHF	Stroke	Amp	Blind	ESRD	LY's	QALY's	IHD	MI	CHF	Stroke	Amp	Blind	ESRD	LY's	QALY's
% with A-fib (100%)	0	0	0	↑	0	0	0	-	↓	0.065	0.170	0.043	0.081	0.007	0.036	0.005	7.87	6.05
% with PVD (100%)	0	0	0	0	↑	0	0	↓	↓	0.062	0.163	0.045	0.042	0.061	0.036	0.005	7.86	6.01
Smoking rate (100%)	0	↑	0	↑	0	0	0	↓	↓	0.059	0.222	0.043	0.063	0.006	0.035	0.003	7.61	5.86
TC: HDL @ 6/0.5mmol/l	↑	↑ +	0	↑	0	0	0	↓	↓	0.217	0.470	0.038	0.096	0.005	0.034	0.003	7.05	5.32
SBP @ 150 mmHg	↑	↑ +	↑	↑ +	↑	0	↑	-	↓	0.082	0.228	0.059	0.101	0.008	0.035	0.014	7.72	5.89
HbA1c @ 11%	↑	↑ +	↑	↑ +	↑	↑	+	↓	↓	0.092	0.254	0.073	0.070	0.030	0.078	0.004	7.59	5.78
IHD (100%)	0	↑	0	0	0	0	0	-	↓	0	0.354	0.042	0.043	0.007	0.036	0.004	7.50	5.1
MI (100%)	0	0	0	0	0	0	0	↓	↓	0.058	0	0.040	0.038	0.005	0.037	0.004	7.01	5.04
CHF (100%)	0	↑	0	↑	0	0	0	-	↓	0.048	0.524	0	0.080	0.004	0.029	0.004	6.76	4.42
Stroke (100%)	0	0	0	0	0	0	0	↓	↓	0.058	0.159	0.042	0	0.005	0.034	0.005	7.59	4.63
Amputation (100%)	0	0	0	0	0	0	0	↓	↓	0.053	0.144	0.034	0.040	0	0.027	0.003	6.70	3.32
Blindness (100%)	0	0	0	0	↑	0	↑	-	↓	0.062	0.164	0.042	0.044	0.030	0	0.033	7.88	5.46
ESRD (100%)	0	0	0	0	0	0	0	↓	↓	0.043	0.129	0.031	0.034	0.004	0.025	0	6.00	3.08

Legend: expected ↑or ↓ is a direct effect, + or – is an indirect effect, 0 is no effect

Grey – observed and predicted fell within 95% CI of base cohort
Yellow – not responsive to expected impact of input variable
Orange – magnitude above expected and exceeded 95% CI
Red – magnitude below expected and exceeded 95% CI

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Amp = amputation; Blind = blindness in one eye; ESRD = end stage renal disease; LY = life years; QALY = quality adjusted life years; CI = confidence interval; DofD = Duration of Diabetes; BMI = body mass index; A-fib = atrial fibrillation; PVD = peripheral vascular disease; TC/HDL = total cholesterol to high density lipoprotein ratio; SBP = systolic blood pressure; HbA1c = glycated hemoglobin A1c

3.4 Discussion

This internal validation exercise provides decision makers with an objective analysis and additional evidence that the risk equations in ODEM are functional and responsive to changes in a range input parameters one at a time which includes demographic variables such as age, gender and ethnicity, pre-existing and TVRF's and a baseline history of seven different T2DM related vascular complications. This exercise also demonstrates that the ODEM is capable of creating a virtual population of thousands of individuals for microsimulation that is reflective of the mean and range of baseline input parameters in a cohort. There are other methods in the ODEM, used in the criteria for model selection which were not assessed in this internal validation, including the probabilistic distributions related to treatment effects, the relevance of the health costs and QALY's for people with T2DM in BC, or the functionality of the ICER, scatter plots and CEAC output.

There were some potential concerns identified in the model functionality related to the lack of responsiveness of the indirect impacts of risk equation coefficients on diabetes related events over ten years. The issue raised is related to the face validity of the DofD which predicted a decreased risk of events with a longer time since diagnosis. Another trend observed was due to the lower than expected cumulative incidence of vascular events in populations with a high mortality rate which reflects the competing outcomes in a model simulation. These results show that mortality must be adjusted for in a comparison of the predicted versus observed outcomes in a model validation exercise. These findings can be assessed with longer term simulation, beyond ten years, and with an adjustment of the cumulative incidence by multiplying the incidence rate by the same number of life years to calculate a cumulative incidence for comparison to the observed events. Overall, this internal validation demonstrates a very positive correlation between the expected and actual changes in model output for all the model outcomes assessed including the diabetes related events, total mortality and QALY's as measured in a 10 year simulation. While valuable, this internal validation exercise does not demonstrate the generalizability of the ODEM for use in a real world BC population with T2DM.

Chapter 4: External validation of the ODEM

4.1 Introduction

As described in Chapter 2, there were many models available for use in the economic evaluation of T2DM. The structure, methods and evidence used to create these models differs which influences the relevance of the model depending on the perspective of the decision makers, the intended population and purpose of the model output. The criteria used in the selection of a T2DM model was for providing a CEA, from the Ministry of Health perspective, of the comparative evaluation of tertiary interventions in a closed cohort, in individuals already diagnosed with T2DM. This process identified a number of potential models which met these criteria and led to the eventual selection of the ODEM. The internal validation of the ODEM demonstrated that the model's risk equations are functional and responsive to changes in baseline input parameters based on observed changes in the cumulative incidence of diabetes related events, life years and QALY outputs.

An important assumption in a model based economic evaluation is that the predicted rate of diabetes events and related mortality are reflective of the rate of events in the population of interest as these have the greatest influence on the costs and QALYs in T2DM. The ODEM was designed based on the methods and risk equations used in the UKPDS-OM1, which itself was developed using the baseline input parameters, risk factors and coefficients to match the rate of diabetes related events in the UKPDS RCT data.(27) A recent validation of the UKPDS-OM1, against the rate of events from the open label extension of the UKPDS trial, demonstrated that after 10 years post UKPDS RCT that the UKPDS-OM1 predicted events were within the observed 95% CI for MI, IHD, blindness and renal failure. The UKPDS-OM1 model slightly over-predicted the probability of heart failure but under-predicted stroke and amputation in this UK population in the open label extension.(52) These results are encouraging and give confidence to policy makers in the UK that the output of an economic evaluation using the UKPDS-OM1 is relevant and credible for their T2DM population. These results in the UK however did not demonstrate that the ODEM was an externally validated model for predicting a rate of vascular events in BC, until it is also shown that these results are generalizable to a BC T2DM population.

This external validation exercise included both a general cohort and a SA cohort, each further divided into sub-cohorts based on gender and age (under 60 and 60 plus years of age) to provide a more in-depth analysis of the impact of these important demographic variables on the predictive accuracy of the ODEM in the real world. These cohorts and sub cohorts were selected due to the differences in the onset and progression of T2DM by gender, age and ethnicity. The availability of aggregate data from a BC population with T2DM that tracks the observed rate of events by age, gender and ethnicity was used to complete this validation exercise and determine the correlation and statistical significance of the ODEM output for use in a BC T2DM population.

4.2 Methods

4.2.1 Cohort structure

To address the gap in the evidence for model output related to demographic risk factors, the external validation was completed for the total cohort, known as the "general" cohort, and a separate cohort of individuals identified of SA descent. These 2 main cohorts were then further divided into sub-cohorts based on gender and age to account for the difference in the rate of vascular events in observed data with these two demographic variables (Figure 1). The age of 60 years was selected as an appropriate method to split in the BC data as it is an age where the prevalence rate and the rate of T2DM related vascular events markedly increase.(9) This allowed a deeper analysis of the results of the external validation to assess whether these demographic coefficients influenced the goodness of fit of the ODEM and to identify areas requiring additional evidence or to guide future calibration of the ODEM risk equations.

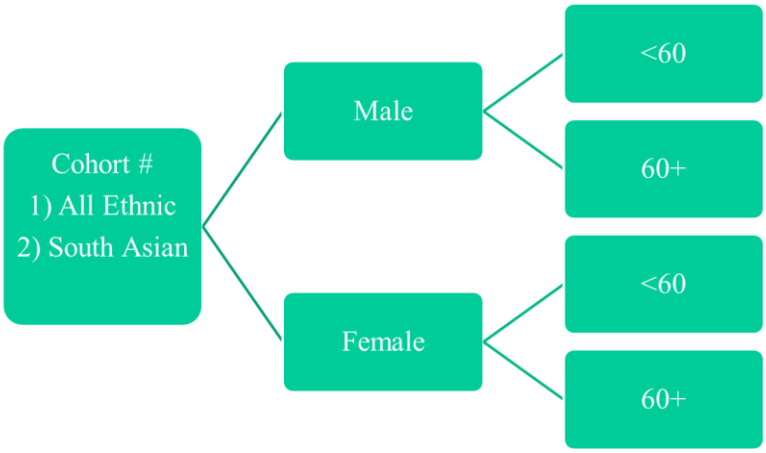


Figure 1: Cohorts and sub cohorts from a T2DM population for the external validation

4.2.2 BC diabetes cohort

British Columbia has administrative data on all citizens who are registered for provincial health insurance including demographic, population health and health service utilization information. The BC data is a compilation of an individual's data using their unique individual, depersonalized, health identification number. The BC data was not accessible for this project however summary results of the rate of vascular events was provided from a previous analysis of a linked, longitudinal and aggregated BC diabetes data of a cohort of 16,056 people with a diagnosis of T2DM and followed from 1993 to 2006. This aggregated results of the data included demographic data such as age and gender, pre-existing risk factors (A-fib and PVD) and the history of previous events as well as the rate of new vascular events over this time period. In this aggregated results there were 1839 people of SA descent (ie: from Pakistan, India, or Bangladesh) identified in the original data using a validated surname analysis. These results were also provided as an aggregate rate of events which enabled an analysis of this ethnic cohort.(82)

The BC diabetes results provided in summary were the actual, real world rate of events per 1000 person years and was organized and tabulated to reflect each sub cohort (Table 8 and Table 9). The BC diabetes data was provided for those who had no history of previous events (pre-event) as well as those who had a history of vascular events (post-event). The pre-event group had data for the incidence rate for six of the diabetes related vascular outcomes with the exclusion of amputation. For the post-event group the rate of incidence vascular events was only provided for those with a pre-existing event that was included in the ODEM risk equations. Therefore, the BC data was available for calculating the observed rate of MI in those with a previous history of IHD or CHF, the observed rate of stroke in those with baseline A-fib or previous CHF, and the observed rate of ESRD in those with a history of blindness. To assess the ability of the ODEM to predict diabetes related vascular events by cohort, two separate simulations were run that included the baseline history of events in a pre-event population and then a mixed pre and post event population. The observed incidence rate by each cohort from the BC data was then multiplied by the life years generated by the ODEM simulation for that cohort to convert this value to an observed cumulative incidence which enabled direct comparison with the ODEM predicted output (Equation 1).

Equation 4: Rate of Events (observed)/1000 x Life Years (predicted) = Cumulative Incidence (observed)

Table 8: BC data: observed incidence rate of vascular events per 1000 in the general population

General Cohorts	Pre-event only						Pre and post event		
	IHD	MI	CHF	Stroke	Blind	ESRD	MI	Stroke	ESRD
All	10.09	13.10	4.10	5.62	0.59	1.00	20.52	6.97	1.00
Male	12.57	16.38	4.01	5.99	0.78	1.25	25.25	7.26	1.23
Male <60	11.49	13.45	1.75	2.79	0.91	0.95	11.37	1.63	1.24
Male 60+	13.96	20.23	6.94	10.18	0.60	1.63	32.24	12.42	1.60
Female	7.36	9.49	4.20	5.22	0.39	0.71	15.27	6.63	0.73
Female <60	6.17	6.66	1.64	2.17	0.42	0.66	4.80	2.17	0.65
Female 60+	8.52	12.29	6.70	8.22	0.36	0.77	20.13	10.50	0.80

Table 9: BC data: observed incidence rate of vascular events per 1000 in the SA population

SA Cohorts	Pre-event only						Pre and post event		
	IHD	MI	CHF	Stroke	Blind	ESRD	MI	Stroke	ESRD
All	11.27	13.10	2.96	2.83	1.41	1.12	18.85	3.78	1.10
Male	14.72	17.99	2.40	2.69	1.87	1.33	25.05	3.84	1.29
Male <60	12.22	14.58	1.09	1.47	2.20	1.46	15.24	2.90	1.31
Male 60+	21.89	27.89	5.98	6.11	0.98	0.98	37.41	9.16	0.92
Female	7.57	7.89	3.57	2.99	0.89	0.89	12.03	3.72	0.89
Female <60	5.08	5.08	2.52	0.50	1.01	0.50	4.94	0.49	0.50
Female 60+	11.23	12.07	5.11	6.66	0.72	1.45	17.64	8.21	1.44

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Blind = blindness in one eye; ESRD = end stage renal disease

Considering the fact that the BC diabetes results did not provide the required baseline input variables to populate the simulation of the cohorts in the ODEM, the missing baseline input data was estimated using the literature that was deemed to be most reflective of the characteristics of British Columbians with T2DM in the 1990's and early

2000's (Table 14 and Table 15). This resulted in the need to estimate the mean and a distribution around the mean for the baseline variables in order to create a virtual population for simulation in the ODEM. This estimation of baseline variables created additional stochastic uncertainty in the ODEM simulation that was assessed via sensitivity analysis to determine the potential impact on model outputs (see section 4.3.3).

4.2.3 Diabetes related vascular outcomes

The T2DM related vascular events that were included in the external validation of the ODEM were IHD, MI, CHF, stroke, ESRD and blindness over a 10 year period . This duration of the simulation was selected to match the duration of time in the linked, longitudinal BC diabetes data. The incidence of amputation was not assessed due to an absence of the code to measure the occurrence of this complication in the BC diabetes dataset. In order to identify the occurrence of these vascular events the International Classification of Disease (ICD) codes were used to identify an event in the BC data. This search required the inclusion of both ICD version 9 and 10 codes as BC transitioned from ICD9 to ICD10 in 2001/02 and the available data spanned from 1996 to 2003. The ICD codes used to identify vascular events from the BC data were compared to the codes used to identify vascular events in the UKPDS data which was used to develop the risk equations in the ODEM (Table 10). From this comparison it was noted that the ICD codes used in the UKPDS were very similar, but not identical, to the codes used in the BC data. Specifically the UKPDS data included additional codes for fatal MI (some which overlap with IHD) and fatal stroke however there were more codes used in the BC data to identify CHF and renal failure.

Table 10: ICD9 and ICD10 codes to identify T2DM outcomes in the ODEM and the BC data

Description	UKPDS(40)	BC ICD-9	BC ICD-10
MI	410, Fatal ≥410 to≤414.9	410 and 410.0	I21.0 to I21.9
IHD	≥411 to ≤414.9	411to 414.9	I20, I22, I24, I25
CHF	≥428 to ≤428.9 (if prior to MI)	425.4 to 425.9, 428	I50 to I50.9, I099, I110, I130, I132, I255, I420, I425, I426, I427, I428, I429, I43, P290
Stroke	≥430 to ≤434.9, 436 (Fatal: ≥430 to ≤438.9)	430 to 436	I60, I61, I63, I64, G45, H341
Renal Failure	250.3, 585 to ≤586, Fatal ≥580 to ≤593.9	582 to 588.0, 403, 404, V420 , V451, V56	N18, N19, N052 to N057, N250, I120, I131 to N037, Z490, Z491, Z492, Z940, Z992
Blindness	≥369 to ≤369.9	369 to 369.9,	H54 (H54.0 to H54.9), E11.3

UKPDS = United Kingdom Prospective Diabetes Study; BC = British Columbia; MI = myocardial infarction; IHD = ischemic heart disease; CHF = congestive heart failure; ICD = International Statistical Classification of Disease (9th and 10th revision)

4.2.4 Baseline input variables by cohort

The baseline input variables required to populate each of the ODEM cohorts for simulation include demographic characteristics (ethnicity, gender, age), duration of diabetes, weight and height, the percent with pre-existing A-fib

or PVD, current smokers and mean measures of TVRF's including TC (mmol/l), HDL (mmol/l), SBP (mmHg) and HbA1c (%). The ODEM baseline input on the percent of individuals by cohort who had a history of each of the seven diabetes related events was only populated for simulations of a combined pre and post event population.

As previously mentioned the BC diabetes results did not include many of the actual measures for the baseline input variables from the population with T2DM. The missing variables were the DofD, BMI, smoking rates and TVRF's by cohort. Therefore, these input variables were identified and extrapolated from available population data and published literature. Considering the BC diabetes data represented the incidence rate of diabetes related events from a real world population in the 1990's and early 2000's, the sources of data were selected which best reflected the traits of the BC population during this era.(53) Therefore, the selection of missing input variables followed a step wise approach which placed the order of priority of data selection in order of a) observational data from the BC population post 1990, followed by b) observational data from Canadian data from 1990 onward and finally by c) identifying evidence from a Canadian RCT post 1990 onward (Table 11). The baseline input variables for the SA cohorts were extrapolated from the variables used for the general population cohorts when there was Canadian literature to quantify the difference for use in the SA cohorts.

Table 11: Data sources for the ODEM baseline input variables for the general population cohorts

Data source	Baseline input variable	Timing
BC diabetes results	Demographic: age, gender, SA ethnicity Pre-existing risk factors – A-fib, PVD History of vascular events	1993 to 2006
National Population Health Survey (NPHS)	BMI (Height and Weight) Smoking rate	1996/97
PHAC 2011 Diabetes in Canada	Duration of diabetes	2008/09
Canadian primary care chart audit and primary care registry	HbA1c, SBP, TC and HDL	2002/03 and 2005/06

PHAC = Public Health Agency of Canada; BMI = body mass index; A-fib = atrial fibrillation; PVD = peripheral vascular disease, HbA1c = glycated hemoglobin; SBP = systolic blood pressure; TC = total cholesterol; HDL = high density lipoprotein

4.2.4.1 Demographic data

The baseline input for the total number of individuals, their mean age and proportion by gender were calculated from the BC diabetes aggregate results. Considering the known influence of age on the rate of diabetes related events, an appropriate age distribution was required to be generated by cohort that approximated the age distribution from the BC diabetes data. Using Canadian public health data as a guide, and assuming a normal distribution, a cohort of 1000 virtual individual patients was generated in the ODEM using the minimum and maximum age with an estimated variance to recreate an approximate age distribution. This was based on Canadian population data which shows the age range for T2DM is typically between 30 years and 85 years, with the greatest number of people in the 60 to 64 age category, and approximately 68% between the age of 49 and 77

years.(83) Therefore, a mean age of 63 years from the BC diabetes data and an SE of 13 years, with a minimum age of 30 and a maximum age of 85, were added to baseline input variables to create a virtual cohort of individuals reflective of the T2DM cohort. This approach was repeated for each sub cohort using the actual mean age calculated from the BC aggregated results.

4.2.4.2 Duration of diabetes (DofD)

There was no information on the DofD in the BC diabetes dataset or literature in the BC population. Therefore, the DofD for the general cohort and sub cohorts were calculated from Canadian population data using the prevalence and incidence of T2DM by gender and age category (Equation 5).(9) The Canadian population data included those with type 1 diabetes in the prevalence and incidence estimates and typically is diagnosed in people under the age of 20 years while T2DM is typically diagnosed over the age of 40 years.(83) Therefore based on expert recommendation, and to ensure the calculation of the DofD was primarily representative of individuals with T2DM, the prevalence rate for diabetes for the under 30 years of age category in the Canadian dataset were subtracted from the prevalence rate in each age category prior to calculating DofD. The removal of individuals under the age of 30 years represented a prevalence rate of 0.69/100 in the Canadian diabetes population and this value was used to adjust the prevalence in each age category prior to calculating the estimated DofD for each cohort.

Equation 5: DofD by age category = [Prevalence (age category) – 0.69 / Incidence (age category)

The DofD calculated for each age category based on estimated prevalence and incidence data in the Canadian population was separated into 10 year age groups by gender (Table 12). As there was no population data to calculate the DofD for the SA cohorts the estimated DofD from the general cohort was increased by 4.6 years to represent the estimated DofD for each SA cohort by gender and age.(82)

Table 12: Canadian prevalence and incidence to calculate duration of diabetes by age and gender

Age (years)	Prev	Inc	DofD (years)	Prev	Inc	DofD (years)	Prev	Inc	DofD (years)
	Female			Male			Total		
30's	2.35	0.29	5.46	2.21	0.32	4.86	2.23	0.31	4.95
40's	4.53	0.52	7.30	5.39	0.73	6.54	4.97	0.62	6.87
50's	9.22	0.98	8.63	11.99	1.36	8.33	10.56	1.17	8.45
60's	15.94	1.53	9.93	21.29	2.08	9.94	18.57	1.79	9.98
70's	22.18	1.78	12.01	27.74	2.31	11.73	24.76	2.02	11.94
80+	21.73	1.55	13.53	26.02	1.91	13.29	23.35	1.69	13.43
All	6.40	0.57	9.92	7.20	0.68	9.64	6.80	0.63	9.70
<60			7.13			6.58			6.76
60+			11.82			11.65			11.78

Prev = prevalence; Inc = incidence; DofD = duration of diabetes

4.2.4.3 BMI: using height and weight

The data for BMI for each general cohort was derived from self-reported height and weight data from a Canadian health survey in the mid 1990's.(84) It was deemed important to gain an accurate estimate of the BMI in the BC population during this era due to the observed trend of an increasing mean BMI in the population over the last 3 decades.(9) The Canadian survey data was from the general population and therefore it was adjusted to better reflect the BMI in those with T2DM by increasing the mean BMI by 4 kg/m^2 above the average value for the population of Canadians without T2DM.(9)

The ODEM baseline input did not use BMI but instead used the height and weight that was calculated from the estimated BMI in people with T2DM. The mean height was sourced from Canadian population data, and was used to calculate the mean weight by gender for the baseline input variables in the ODEM (Table 14). To do this, the mean BMI was divided by the average squared height (in meters) for each cohort (Equation 6). Furthermore, to generate the range of heights and weights that would be representative of a Canadian population for each cohort a normal distribution was created around the mean using the approximate range for the minimum and maximum height in order to estimate a SE of the mean. The mean height and weight for each cohort was then entered into the baseline parameters along with the estimated standard deviation (SD) minimum and maximums to generate a cohort of 1000 individuals with a height and weight distribution representative of the Canadian population with T2DM in the 1990's. For example, this calculation for the general cohort used the mean height for the population of 1.71 meters and estimated a SD of 15 centimeters based on a range between 150 to 190 centimeters that was input into the baseline input variables to generate a distribution of height and weight for the microsimulation.

Equation 6: Weight = BMI / $Height^2$

The literature for SA's living in Canada showed they have a lower mean BMI and a lower prevalence of obesity (ie: BMI >30) than their Caucasian counterparts.(9) This data was used to estimate a baseline mean BMI of 28 kg/m^2 for the overall SA cohort.(85,86) The published data also reflected a BMI in SA men which was slightly higher than the BMI for SA women and this was used to adjust the BMI for the baseline input parameters of the SA cohorts by gender (Table 15).

4.2.4.4 Time varying risk factors (TVRF's)

The levels of control of the biomarkers relevant for T2DM, HbA1c, SBP and TC/HDL are important coefficients in the long term risk of diabetes related events.(6) The baseline HbA1c, SBP and cholesterol levels were not available from the BC diabetes data and therefore this needed to be extrapolated from Canadian primary care data from T2DM patients who had a similar mean age, DofD and history of diabetes related events. Two studies, that were

considered reflective of the BC T2DM population in the era of the BC data capture, were selected and the results were used to populate the baseline HbA1c, SBP and TC/HDL input variables by cohort. A 2005/06 study of a Canadian primary care diabetes registry of 3002 patients was the source the baseline input for SBP, TC and HDL and a Canadian primary care chart audit from 2002/03 was selected to source the baseline HBA1c (Table 13).(87,88)

The primary care chart audit included data that guided the estimation of the baseline HbA1c by age cohort as it provided a breakdown of the measures of HbA1c that varied with the DofD. In this data from the chart audit the percent of patients with optimal (<7%), sub optimal (7.0 to 8.4%) and inadequate (>8.4%) control of their HbA1c were split into five categories based on their DofD (<2 years, 3 to 5 years, 6 to 9 years, 10 to 14 years and 15+ years). There was an observed trend which showed a decreasing proportion of patients with optimal HbA1c control as the duration of T2DM increased. These data allowed the mean HbA1c to be calculated for each DofD category using the proportion of people that were identified at ideal, sub-optimal or inadequate levels of control (Equation 7). The mean HbA1c value that was assigned by level of control was 6.0 in the ideal, 7.7 in the suboptimal and 9.4 in the inadequate and each was then multiplied by the proportion of people at each level of control. This was repeated based on the changing proportion of people at the various levels of control by DofD category. This calculated level of HbA1c by DofD category was then matched to the estimated DofD in each cohort for the ODEM simulation. This resulted in an estimated baseline HbA1c in the under 60 years of age cohort of 7.2%, and in the 60 plus cohort of 7.6% (Table 13). These appeared to be reasonable estimates when cross referenced to the mean HbA1c of 7.3% from Canadian primary care registry.(76,77)

Equation 7: Calculation of mean HbA1c using level of control and duration of diabetes by cohort:

Mean A1c = [6.0 *(% ideal] + [7.7 *(% suboptimal)] + [9.4 *(% inadequate)](87)

There was no literature identified to estimate the baseline HbA1c level for the SA cohorts versus a comparable group in the general population, which had been adjusted for confounding variables, and therefore no adjustments were made to the baseline HbA1c for the SA cohorts. There were comparative data that showed a slight increase in the mean SBP by 1 mmHg and a higher proportion of SA adults with hypertension.(85) The TC used for the SA cohorts were the same as the general cohort however the HDL was lowered by 0.19 mmol/l to reflect the literature that resulted in an increase in the TC/HDL ratio for the SA cohort.(23,85)

Table 13: Comparison of baseline characteristics for the estimation of HbA1C, SBP, TC and HDL

Source of evidence	Patient Demographics	Gender/ BMI	Pre-existing events (%)	Comments:
BC diabetes data	N = 16056 Age = 63 yrs DofD= 9.7 yrs	54% male BMI 29	PVD = 2.7 IHD = 20.3 MI = 24.4 CHF = 6.5 Stroke = 6.1	Actual HbA1c, SBP and cholesterol were not available
Canadian Primary care registry 2005/06 (87)	N = 3002 Age = 64 yrs DofD = 6 yrs	59% male, BMI 30 kg/m2	PVD = 9 CAD = 23 CHF = 5 CVD = 8	HbA1c = 6.9% SBP = 130 mmHg TC = 4.3 mmol/l HDL = 1.2 mmol/l
Canadian Primary care chart audit 2002/03 (88)	N = 2473 Age = 63 yrs DofD = 7.8 yrs	54% male, BMI 31	PVD = 6 Angina = 11 MI = 11, CHF = 7 Stroke = 5	Mean HbA1C = 7.3%, with breakdown based on level of control based on DofD

DofD = duration of diabetes; BMI = body mass index; PVD = peripheral vascular disease; IHD = Ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; CVD = cerebrovascular disease; CAD = coronary artery disease; HbA1c = glycated hemoglobin; SBP = systolic blood pressure; TC = total cholesterol; HDL = high density lipoprotein

4.2.4.5 Pre-existing risk factors: A-fib, PVD and smoking

The aggregated BC diabetes results included a measure of the proportion of individuals with an ICD9 code for A-fib or PVD in their records at baseline. The numbers for PVD were consistent with recent literature using BC data that captured a number of pre-existing risk factors in people with T2DM (Table 14 and Table 15).(82) These were used in the baseline input for the simulations in the ODEM output that was done in those with a history of pre-existing events.

It has been documented that smoking rates in Canada were higher in the 1990's and has decreased over the last three decades.(12) Data from the 1996/97 National Population Health Survey (NPHS) were used to estimate the mean smoking rate by age and sex. Those included in the baseline percent were "current smokers", defined as those who smoked occasionally or were daily smokers. It has been observed that BC has the lowest overall smoking rate in Canada therefore using Canadian data may have overestimated the actual prevalence of smoking in BC's diabetes population in the simulation.(12,26) The rate of smoking also differed by gender with male rates approximately 10% higher and women 10% lower than the population mean. These estimated differences were applied to the smoking rate used for the simulation of the general cohort to reflect the gender differences in smoking rate for each cohort (Table 14).(12) In the SA cohorts the baseline input for the percentage of smokers was based on pooled data of a number of Canadian health surveys showing a smoking rate in the Canadian SA population of 8.6% which was well below the smoking rate in a general Canadian population of a similar age.(Table 15).(11)

4.2.4.6 Pre-existing diabetes related events

The baseline proportion of people with pre-existing vascular events was captured directly from the aggregated BC Diabetes results and stratified by age, sex and ethnicity (Tables 14 and 15). These data were compared with other published data on prevalence rates which confirmed the estimates were within a reasonable range.(87–89) The rates of amputation was not available from the BC diabetes data and therefore the percent of people with an amputation in the T2DM population was instead estimated from the Canadian literature.(87,88)

Table 14: General Population: baseline input variables for T2DM cohort by gender and age

Cohort Parameters*	All general	All male	Male < 60 yrs	Male 60+ yrs	All female	Female <60 yrs	Female 60+ yrs
Mean age (yrs)	63.0	62.0	50.9	71.2	64.1	50.9	72.4
% female	46	0	0	0	100	100	100
DofD (yrs)	9.7	9.6	6.8	11.8	9.9	7.1	11.8
BMI (kg/m2)	29	30	30	30	29	29	29
Weight (kg)	85	93	93	93	77	77	77
Height (m)	1.70	1.76	1.76	1.76	1.63	1.63	1.63
TC (mmol/l)	4.3	4.3	4.3	4.3	4.3	4.3	4.3
HDL (mmol/l)	1.2	1.2	1.2	1.2	1.2	1.2	1.2
SBP (mmHg)	130	130	130	130	130	130	130
HbA1c (%)	7.3	7.3	7.2	7.6	7.3	7.2	7.6
% Smoker	24.5	27.0	26.2	13.5	22.1	21.6	11.2
% PVD	2.7	3.0	0.9	4.6	2.5	1.3	3.2
% A-fib	2.5	3.2	0.9	5.1	6.3	0.3	10.0
History IHD	20.3	21.4	13.4	27.9	19.0	10.4	24.4
History MI	24.4	25.6	17.2	32.4	22.8	13.52	28.6
History CHF	6.5	6.5	2.3	9.9	6.6	1.3	10.0
History Stroke	6.1	6.1	2.3	9.2	6.1	1.6	8.9
History Amp	1.0	1.0	1.0	1.0	1.0	1.0	1.0
History blind	0.7	0.7	0.4	0.9	0.8	0.4	1.0
History Renal	1.7	1.9	1.1	2.5	1.4	0.9	1.7

*SE input into the ODEM to generate a simulated population has not been included in the table
DofD = duration of diabetes; BMI = body mass index; TC = total cholesterol; HDL = high density lipoprotein; SBP = systolic blood pressure; HbA1c = glycated hemoglobin; PVD = peripheral vascular disease; A-fib = atrial fibrillation; IHD = ischaemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Amp = amputation; Blind = blindness

Table 15: South Asian: baseline input variables for T2DM cohort by gender and age

Cohort Parameters*	All SA	All SA male	SA male <60 years	SA male 60+ years	All SA female	SA female < 60 years	SA female 60+ years
Mean age (yrs)	58.0	56.8	50.2	69.1	59.4	51.1	69.1
% female	45	0	0	0	100	100	100
DofD (yrs)	14.3	14.2	11.4	16.4	14.5	11.7	16.4
BMI (kg/m2)	28	28.5	28.5	28.5	27.5	27.5	27.5
Weight (kg)	82	88	88	88	73	73	73
Height (m)	1.71	1.76	1.76	1.76	1.63	1.63	1.63
TC (mmol/l)	4.3	4.3	4.3	4.3	4.3	4.3	4.3
HDL (mmol/l)	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Cohort Parameters*	All SA	All SA male	SA male < 60 yrs	SA male 60+ yrs	SA female	SA female <60 yrs	SA female 60+ yrs
SBP (mmHg)	131	131	131	131	131	131	131
HbA1c (%)	7.3	7.3	7.2	7.6	7.3	7.2	7.6
% Smoker	8.6	13.8	14.4	7.4	3.6	3.2	1.7
% PVD	0.7	0.9	0.3	2.0	0.4	0.4	0.7
% A-fib	1.0	1.0	1.1	1.0	1.1	0.8	1.4
History IHD	16.6	18.4	13.6	27.2	14.3	10.2	19.2
History MI	20.4	21.9	17.1	30.7	18.6	14.9	22.8
History CHF	4.0	4.3	2.1	8.4	3.6	1.2	6.4
History Stroke	2.8	3.3	0.8	7.9	2.1	0.4	4.1
History Amp	0.5	1.0	0.0	1.0	1.0	0.0	0.5
History blind	0.9	1.4	1.1	2.0	0.2	0.0	0.9
History Renal	1.5	2.1	0.2	3.0	0.8	0.4	1.4

*SE input into the ODEM to generate a simulated population has not been included in the table
SA = South Asian; DofD = duration of diabetes; BMI = body mass index; TC = total cholesterol; HDL = high density lipoprotein; SBP = systolic blood pressure; HbA1c = glycated hemoglobin; PVD = peripheral vascular disease; A-fib = atrial fibrillation; IHD = ischaemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Amp = amputation; Blind = blindness in one eye

4.2.5 Design of the simulation

In addition to the effort to identify appropriate baseline input parameters for the external validation, it also required the selection of the number of individuals to include in each simulation, the number of iterations on each individual in the cohort, and the number of cycles (years) to run the simulation. Therefore, each simulation was based on creating a virtual cohort whose individual characteristics varied to reflect the baseline input parameters for each cohort. The baseline patient characteristics for continuous variables including age, DofD, height and weight, TC, HDL, SBP and HbA1c were randomly selected from a distribution based on the mean, SD, minimum and maximum values that were input into the ODEM. The baseline input variables to determine whether an individual had a pre-existing risk factor such as A-fib, PVD, smoking or history of a vascular event were identified from a random number generator that assigned a discrete number (0 or 1) based on whether the random value exceeded the cohort mean. The model simulation was designed to utilize the probabilistic distribution and the progression of TVRF's functions to calculate the cumulative incidence of vascular events in 1000 individuals run through ten iterations over ten years, to equal 10,000 individual micro simulations. This external validation exercise focused on comparing the diabetes related vascular events predicted by the ODEM's risk equations and did not include any assessment of cost or QALY's in BC's population.

4.2.6 Correlation and statistical analysis

The predicted cumulative incidence of diabetes related events from the ODEM simulation was compared against the observed incidence of events from the aggregated BC diabetes results for each cohort (Tables 16 to 19). The objective measures used to assess the accuracy of the ODEM predicted events were the coefficient of determination (R^2) and whether the predicted incidence were under, within or overestimated versus the mean and 95% CI of observed data as a measure of the statistical significance of the output. A goodness of fit measure of R^2

above 0.49 was considered as the threshold to assess whether the validation reflects a correlation between predicted and observed events at 10 years. The R^2 for each cohort was then complemented by a correlation plot to provide a further visual assessment of the absolute cumulative incidence of each vascular outcome and how closely they correlated with the BC diabetes data. The 95% CI of the observed data was provided as part of the aggregate results from the BC cohorts similar to the approach used in the validation of the UKPDS-OM1. The use of MAPE was not incorporated into the comparison of the 10 year cumulative incidence of predicted versus observed data due to the wide skewing of the output measures when assessing low volume data such as the incidence of ESRD and blindness. Other measures such as RMSPE, 2 sided T-test and C statistic were not included based on their infrequent use in external validations of T2DM models. The log rank test (X2) could not be considered as the data provided was not sufficient enough to create a survival curve of the events over time that is required for this measure.

4.3 Results

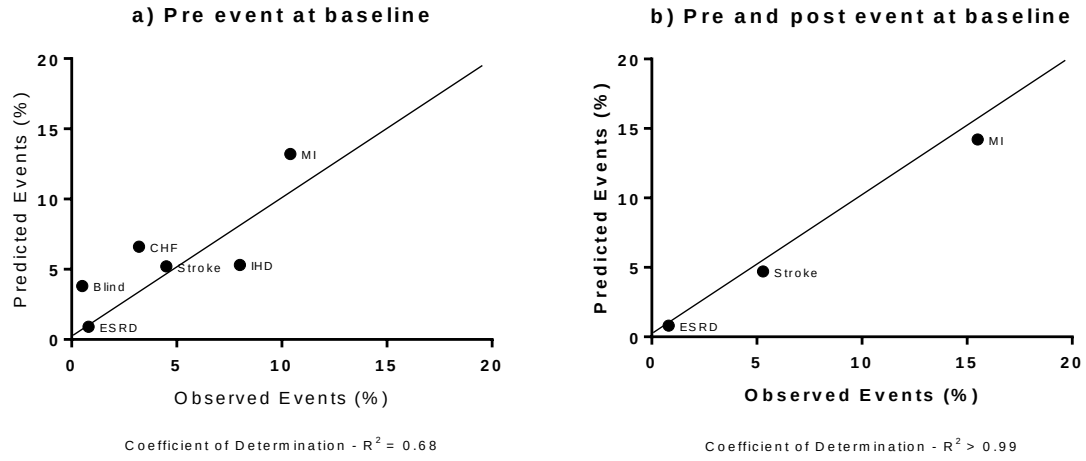
The external validation results of the ODEM simulation for the general population were first completed and assessed and then followed by a similar approach by cohort for the SA population. The analysis included each sub cohort and a separate simulation based on presence or absence of previous diabetes related events, gender and age. This approach was taken as it provided an in-depth evaluation of the generalizability of the ODEM for use in CEA of a diverse, real world BC population with T2DM.

4.3.1 General population cohort

This cohort represented a full mix of people with T2DM in the BC population regardless of gender, age or ethnic descents in adults aged 40 and over. The results in this cohort were therefore the considered the most relevant for decision makers as it represents the T2DM population of interest for an economic evaluation T2DM interventions. The simulation was first run only for those who were identified as having no previous history of events which showed the ODEM predicted incidence of events were equal to or higher than the incidence of all observed vascular events with the exception of IHD (Figure 2a, Table 16). The correlation of determination in this pre-event group was positive and resulted in a R^2 of 0.62 and the estimated events for stroke and ESRD falling within the statistical bounds of the 95% CI of the actual events. Of note, the occurrence of blindness estimated by the ODEM was more than 7 fold higher than actual in the overall population while all other outcomes were modestly above or below the observed rate. Considering blindness is included as an indirect coefficient in the ODEM risk equations for ESRD, this could in theory have indirectly increased the predicted incidence of ESRD, however the internal validation found that these indirect coefficients had little impact on the predicted rate of events in a ten year simulation. Similar to the finding of the internal validation, the ten year time span for this simulation may have been too short to allow the increase in blindness to result in an increase in ESRD.

Considering that any real world T2DM population is likely to include those who have experienced a diabetes related event, a further simulation was done for the general cohort that combined results for those who were pre and post-event. These observed results were only available for comparing the incidence of three of the vascular events namely MI, stroke and ESRD (Figure 2b, Table 17). In this combined pre and post-event cohort the ODEM predicted incidence of events were highly correlated with an R^2 at 0.9997 and all three outcomes predicted within the 95% CI range of the observed events. Upon further investigation the increased correlation was not due to changes in the values from the predicted incidence of events in ODEM but due to an increase in the observed incidence of events in the BC results. These results support the use of the ODEM as a model for consideration for future economic evaluations in a general BC T2DM population that include a mix of individuals with a history of T2DM related complications.

Figure 2: General cohort: Plot of the cumulative incidence (%) of vascular events, predicted versus observed

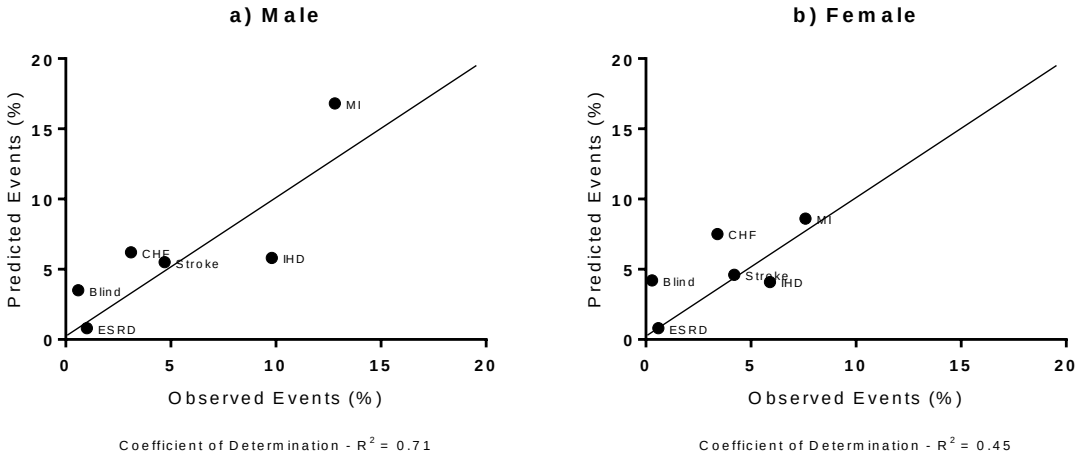


4.3.1.1 General population sub-cohorts by gender

The next step in the external validation was to run simulations by gender to determine if the female coefficient in the ODEM risk equation influenced the models predictive performance. This simulation included all vascular outcomes however only IHD, MI and stroke included a specific coefficient to influence outcomes for a female cohort. Therefore, any differences in the predicted incidence of events by gender could have been either due to the gender coefficient in the risk equation or differences in the female versus male baseline risk factors. A look at the results in the pre event male cohort showed a similar overall trend as the general population with the ODEM again over estimating the incidence of MI, CHF and blindness and under estimating the incidence of IHD (Figure 3a, Table 18). As with the general population the incidence of stroke and ESRD were within the 95% statistical bounds of the observed rate for the male cohort with the R^2 improving to 0.71. The results for the pre event female cohort resulted in a similar trend for each of the six outcomes with a noted improvement in the estimated versus

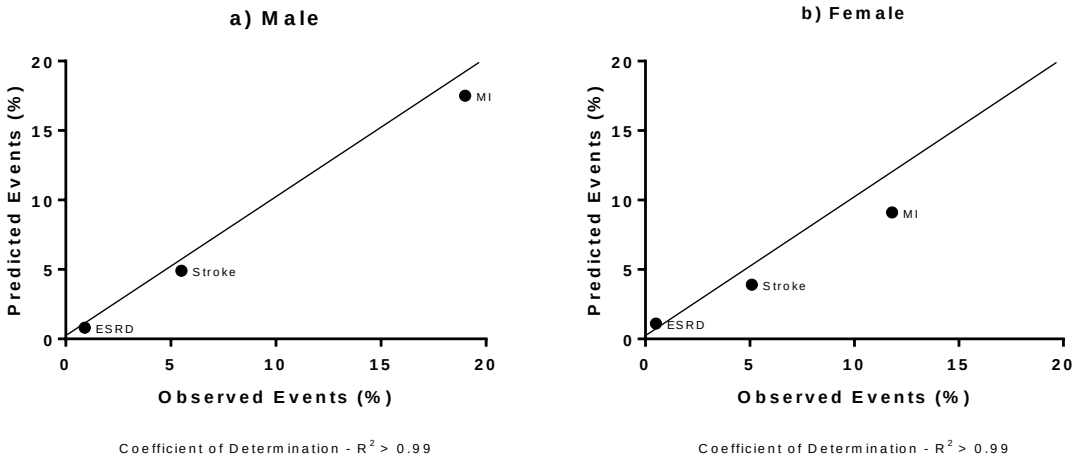
observed outcomes for MI. This resulted in the MI, as well as stroke and ESRD, also falling within the 95% statistical bounds of the observed rate for female (Figure 3b, Table 19). The ODEM again underestimated the predicted incidence of IHD for the female cohort however it moved closer to the lower bound of the 95% CI of the observed rate. While the predicted incidence of MI and IHD improved in the female versus the male cohort, the ODEM over-estimation of the incidence for CHF and blindness for females was greater than in males which resulted in an overall decrease in the R^2 to 0.45. Considering there was not a coefficient for female gender in the risk equations for CHF and blindness, the greater over estimation of predicted events must have a result of the lower incidence of observed event in the aggregated BC results. Overall, the separation of the cohorts by gender did not directionally change the results of the predicted versus observed outcomes but did impact the magnitude of these differences based on gender for the vascular events of MI, IHD, CHF and blindness.

Figure 3: General cohort, by gender (pre-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed



Next, the proportion of individuals identified as post-event were once again included in the baseline input parameters for the male and female cohorts and simulated in the ODEM. The results for these two gender specific cohorts of a mixed pre and post event were consistent, and showed that the ODEM predicted and observed rate for MI, stroke and ESRD were again highly correlated with an R^2 above 0.99 in both the male and female plots (Figure 4a and 4b). Furthermore, statistically speaking, all of the predicted outcome measures fell within the 95% CI of the actual incidence in both cohorts with the exception of MI in the female cohort where the model slightly underestimated the observed events (Table 20 and 21). The ODEM demonstrated a high predictive ability of the 10 year incidence of MI, stroke and ESRD in both men and women when compared to the results in a BC T2DM population that included those with a history of vascular events in the baseline data.

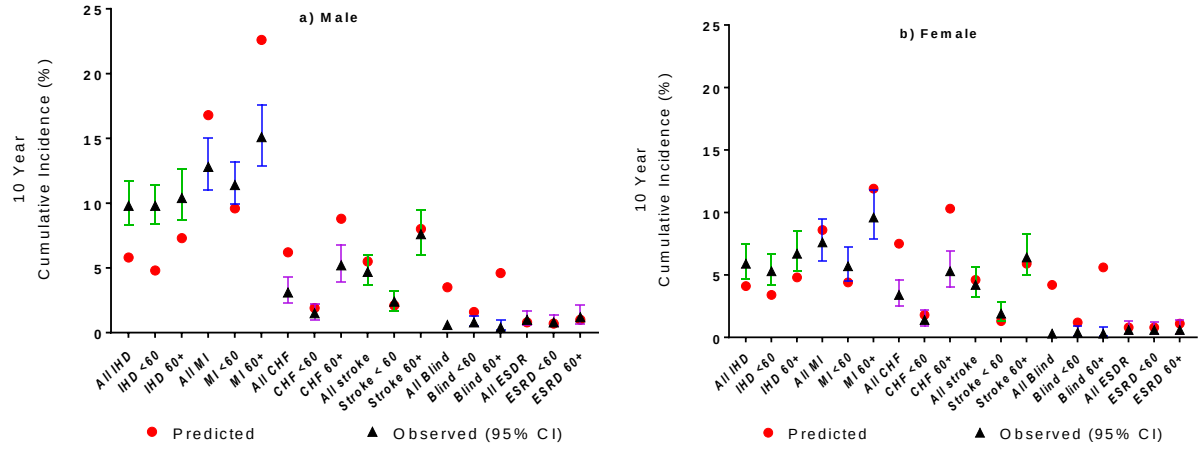
Figure 4: General cohort, by gender (pre and post -event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed



4.3.1.2 General population gender sub-cohorts by age

To assess the influence of age on the predictive ability of the ODEM the male and female gender cohorts were further divided into those under 60 years and another 60 years plus. The baseline input variables were adjusted to reflect differences in age, DofD, baseline HbA1c and a percent of smokers in the age cohorts by gender. The results of these simulations were insightful and demonstrated that both the predicted and observed results appropriately reflected an age dependant difference in the rate of vascular complications (Figure 5a and b). As expected the younger cohorts had fewer vascular events than the older cohort in both the predicted and observed data. This demonstrated the responsiveness of the ODEM to the age and DofD variables that will become more relevant when conducting future CEA in an aging population. It was also clear that the over estimation of the predicted rate of events from the ODEM for MI, CHF and blindness was essentially a result of the high predicted rate of events in the age 60 years plus cohort for both genders, while the under 60 age results were close to, or within, the 95% CI range of observed events for MI, CHF and blindness (Table 18 and 19). Interestingly, an older age at baseline did not result in an over estimation of the predicted versus observed outcomes for IHD, stroke or ESRD. As previously outlined, the risk equations for vascular events in the ODEM did not include coefficients for current age or the DofD but only included the age at the diagnosis of diabetes. The coefficient for current age was however used in the equations which calculated the changes in time varying risk factors of HbA1c and SBP. It can therefore be determined that the higher predicted rate of vascular events in the older age cohort was likely a result of increasing HbA1c and SBP with older age and the longer DofD. This was an assumption in the ODEM methods based on the data from the UKPDS RCT on the natural progression of T2DM over time.(40) Of the TVRFs in the ODEM, only HbA1c was included in all of the risk equations for MI, CHF and blindness while SBP was included in the calculation of MI and CHF. The coefficients of the progression of HbA1c, SBP, TC and HDL should be validated in the ODEM to ensure these are reflective of the real world changes with age and the DofD in the population.

Figure 5: General cohort, by gender and age (pre-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed (with 95% CI)



As a final assessment of the ODEM for predicting the incidence events in the general population the baseline input variables were once again revised, using the proportion of the T2DM population with history of complications in the aggregated results of the BC data, to include those who were post event at baseline. Once again, the simulation of a cohort with the mixed history events in the ODEM resulted in a high correlation between the predicted and actual incidence of MI, stroke and ESRD (Figures 6a and 6b). Furthermore, the over estimation by the ODEM in the 60 plus age cohort that was apparent in the pre event cohort were no longer evident and all predicted outcomes fell within the statistical range of the observed incidence of events (Table 20 and 21). This higher correlation of the ODEM and BC data, particularly in these 60 plus age cohorts, was mainly as a result of an increase in the rate of observed events in individuals who have a history of a vascular event and was not due to major changes in the absolute predicted rate of events in the ODEM output.

Figure 6: General cohort, by gender and age (pre and post-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed (with 95% CI)

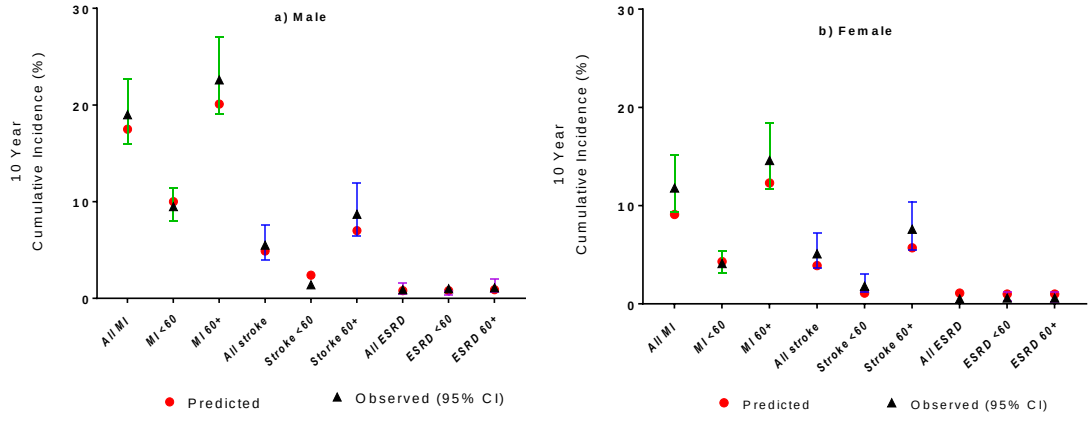


Table 16: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in a pre-event general cohort

Type of vascular event	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed
IHD	0.080	0.066	0.098	0.053	under
MI	0.104	0.088	0.120	0.132	over
CHF	0.032	0.024	0.044	0.066	over
Stroke	0.045	0.034	0.058	**0.052**	within
Blindness	0.005	0.002	0.010	0.038	over
ESRD	0.008	0.004	0.015	**0.009**	within

Table 17: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in a post-event general cohort

Type of vascular event	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed*
MI	0.155	0.128	0.188	**0.142**	within
Stroke	0.053	0.038	0.074	**0.047**	within
ESRD	0.008	0.004	0.015	**0.008**	within

Table 18: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in males by age for pre-event only

Type of vascular event	Male cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed*
IHD	All Age	0.098	0.083	0.117	0.058	under
	< 60	0.098	0.084	0.114	0.048	under
	60+	0.104	0.087	0.126	0.073	under
MI	All Age	0.128	0.110	0.150	0.168	over
	< 60	0.114	0.099	0.132	0.096	under
	60+	0.151	0.129	0.176	0.226	over
CHF	All age	0.031	0.023	0.043	0.062	over
	< 60	0.015	0.01	0.022	**0.019**	within
	60+	0.052	0.039	0.068	0.088	over
Stroke	All Age	0.047	0.037	0.060	**0.055**	within
	< 60	0.024	0.017	0.032	**0.021**	within
	60+	0.076	0.060	0.095	**0.080**	within
Blindness	All Age	0.006	0.003	0.011	0.035	over
	< 60	0.008	0.005	0.013	0.016	over
	60+	0.004	0.002	0.010	0.046	over
ESRD	All Age	0.010	0.006	0.017	**0.008**	within
	< 60	0.008	0.005	0.014	**0.007**	within
	60+	0.012	0.007	0.021	**0.010**	within

Table 19: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in females by age for pre-event only

Type of vascular event	Female cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
IHD	All Age	0.059	0.047	0.075	0.041	under
	< 60	0.053	0.042	0.067	0.034	under
	60+	0.067	0.053	0.085	0.048	under
MI	All Age	0.076	0.061	0.095	**0.086**	within
	< 60	0.057	0.045	0.072	0.044	under
	60+	0.096	0.079	0.118	0.119	over
CHF	All age	0.034	0.025	0.046	0.075	over
	< 60	0.014	0.009	0.022	**0.018**	within
	60+	0.053	0.040	0.069	0.103	over
Stroke	All Age	0.042	0.032	0.056	**0.046**	within
	< 60	0.019	0.013	0.028	**0.013**	within
	60+	0.064	0.050	0.083	**0.059**	within
Blindness	All Age	0.003	0.001	0.008	0.042	over
	< 60	0.004	0.001	0.009	0.012	over
	60+	0.003	0.001	0.008	0.056	over
ESRD	All Age	0.006	0.003	0.013	**0.008**	within
	< 60	0.006	0.003	0.012	**0.008**	within
	60+	0.006	0.003	0.014	**0.011**	within

Table 20: Comparison of the cumulative incidence (%) of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in males by age for combined pre and post-event

Type of vascular event	Male cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
MI	All Age	0.190	0.160	0.227	**0.175**	within
	< 60	0.095	0.080	0.114	**0.100**	within
	60+	0.226	0.191	0.270	**0.201**	within
Stroke	All Age	0.055	0.040	0.076	**0.049**	within
	< 60	0.014	0.010	0.021	0.024	over
	60+	0.087	0.065	0.119	**0.070**	within
ESRD	All Age	0.009	0.005	0.016	**0.008**	within
	< 60	0.010	0.003	0.008	**0.008**	within
	60+	0.011	0.007	0.020	**0.009**	within

Table 21: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in females by age for combined pre and post-event

Type of vascular event	Female cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
MI	All Ages	0.118	0.094	0.151	0.091	under
	< 60	0.041	0.031	0.054	**0.043**	within
	60+	0.146	0.117	0.184	**0.123**	within
Stroke	All Ages	0.051	0.037	0.072	**0.039**	within
	< 60	0.018	0.012	0.030	0.011	under
	60+	0.076	0.055	0.104	**0.057**	within
ESRD	All Ages	0.005	0.002	0.011	**0.011**	within
	< 60	0.006	0.003	0.012	**0.010**	within
	60+	0.006	0.002	0.012	**0.010**	within

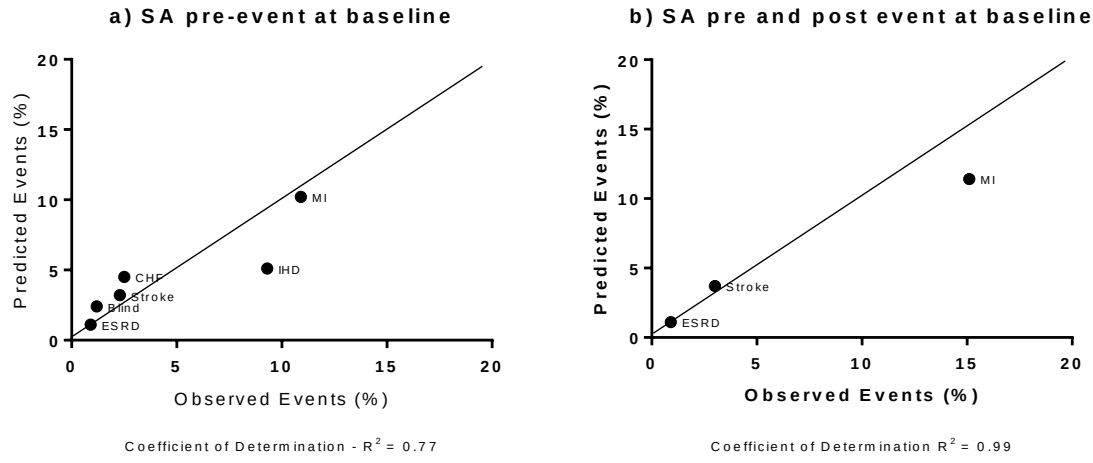
*Captures whether the ODEM underestimated, overestimated or were within the 95% CI (confidence interval) of the observed BC data. IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; ESRD = end stage renal disease.

4.3.2 South Asian cohort

The SA population had a differing epidemiology, level of baseline risk factors and rate of vascular events in T2DM and therefore it was not known whether the predicted outcomes from the ODEM were applicable to the observed incidence in a SA population. First, the BC diabetes aggregated results of observed events showed that those of SA descent had higher rates of IHD, MI and blindness however a lower rate of CHF and stroke compared to the general population (Table 8 and 9). It is important to note that, while the ODEM was designed to allow a baseline input variable for ethnicity, the coefficient in the risk equations used to identify an individual as SA was the same coefficient used for a Caucasian cohort. Therefore, as shown in tables 14 and 15, the differences in predicted outcomes for the SA cohort in the ODEM simulation were due to differences in the baseline input parameters for the SA cohorts such as a younger mean age (58 years versus 63 years), longer duration of T2DM (14.3 versus 9.7) and lower smoking rates (8.6% versus 24.5%). This meant that the increased correlation of the ODEM's predicted outcomes versus actual events was not a reflection of a coefficient for SA ethnicity in the risk equations.

The predicted outcomes from the ODEM in the SA cohort showed an increased correlation with an R^2 of 0.77 with the predicted rate of vascular events were within the statistical range of the observed rate with the exception of IHD (Figure 7a, Table 22). The ODEM's lower predicted incidence of MI and stroke in the SA cohort was not unexpected based on the lower proportion of smokers versus the general population and the coefficient for smoking in the ODEM risk equations for these two vascular complications.

Figure 7: South Asian cohort: Plot of the cumulative incidence (%) of vascular events, predicted versus observed

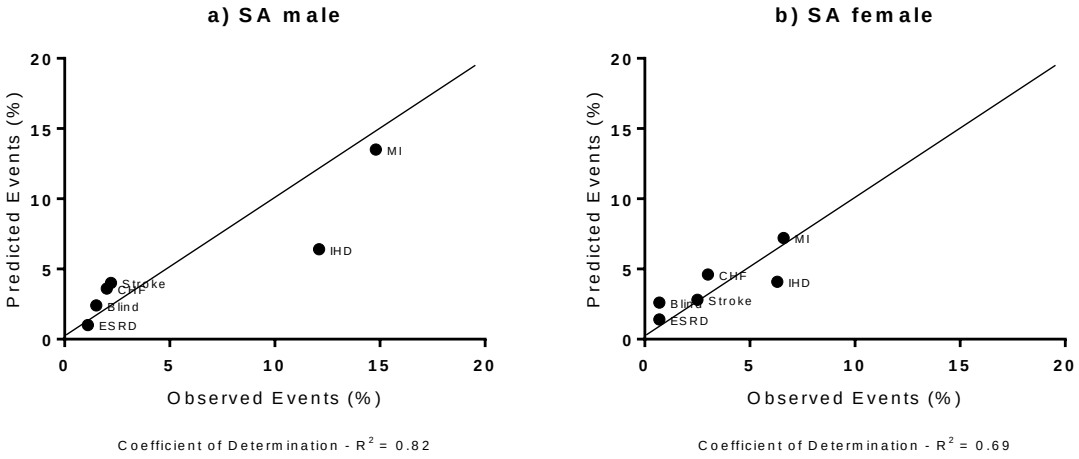


Similar to what was demonstrated in the general cohort, the addition of individuals to the SA cohort who were post event at baseline increased the correlation between the predicted and observed outcomes for MI, stroke and ESRD to an R^2 above 0.98 (Figure 7b). In the SA cohort, all three measures were within the wider 95% statistical range of the observed rate of vascular events (Table 23). There was a noticeable decrease in the predicted incidence of MI in the SA simulation compared to the general cohort which was most likely reflected the reduced smoking rate in the baseline input parameters. This higher correlation of predicted and observed events in the SA cohorts was considered a promising indication for the use of the ODEM in a BC T2DM population which includes people of SA descent and a history of T2DM related events.

4.3.2.1 South Asian sub-cohorts by gender

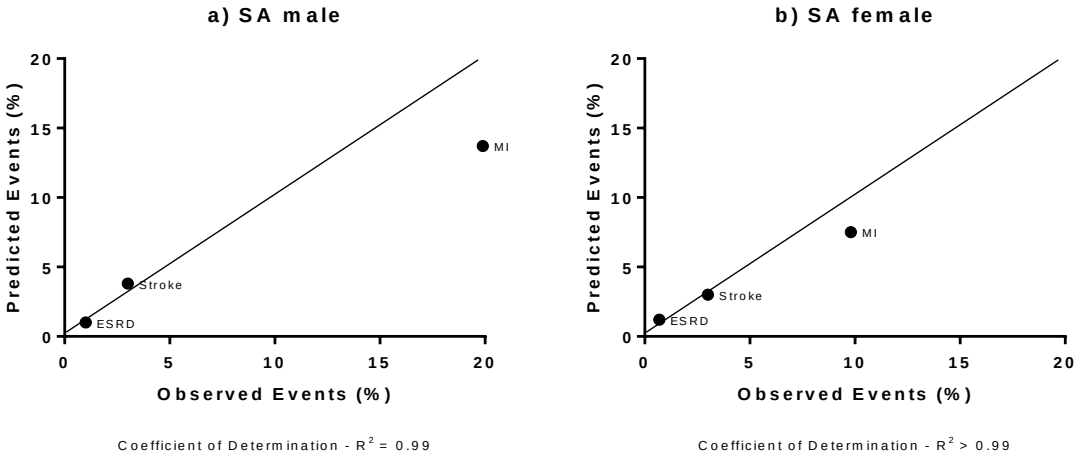
Generally the predicted outcomes from the ODEM were consistent for both male and female cohorts (Figure 8a and 8b). The simulation slightly overestimated the incidence of CHF, stroke and blindness while it underestimated the incidence of IHD, in particular in males. The predicted rate of MI in the male and female SA cohorts were much more closely correlated with the observed rate and both were well within the 95% CI statistical range of observed events (Table 24 and 25). Overall, the enhanced correlation between predicted and observed in the male SA cohort, with an R^2 of 0.82, was due to a combination of a higher observed rate of IHD, MI and blindness in the SA cohorts and a lower predicted rate from the ODEM. The enhanced correlation in the SA female cohort (R^2 of 0.69), versus the general female cohort (R^2 of 0.45), was a reflection of the lower predicted rate of CHF and blindness from the ODEM simulation in SA women compared to the general female cohort while the observed rate remained similar. This enhanced correlation in the female SA cohort was therefore attributed to the longer DofD and a younger mean age used in the risk equations.

Figure 8: South Asian cohort, by gender (pre-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed



As observed in the general cohort simulation, adding those with a history of pre-existing events by gender into the simulation once again resulted in a high correlation of the predicted versus observed rate of MI, CHF and ESRD for both the male and female, with the predicted measures of all vascular events being within the 95% CI statistical range set by the BC results (Figure 9a and 9b, Table 23 and 24). There is a noticeable decrease in the ODEM's predicted rate of MI in the SA cohorts compared to the general cohorts which again was determined to be due to lower estimated smoking rate and the younger mean age of SA with T2DM. While there is a slight increase in the observed incidence of MI in the aggregated BC results for SA males, the difference in the predicted and observed smoking rate and younger mean age at diagnosis of T2DM in the baseline input parameters.

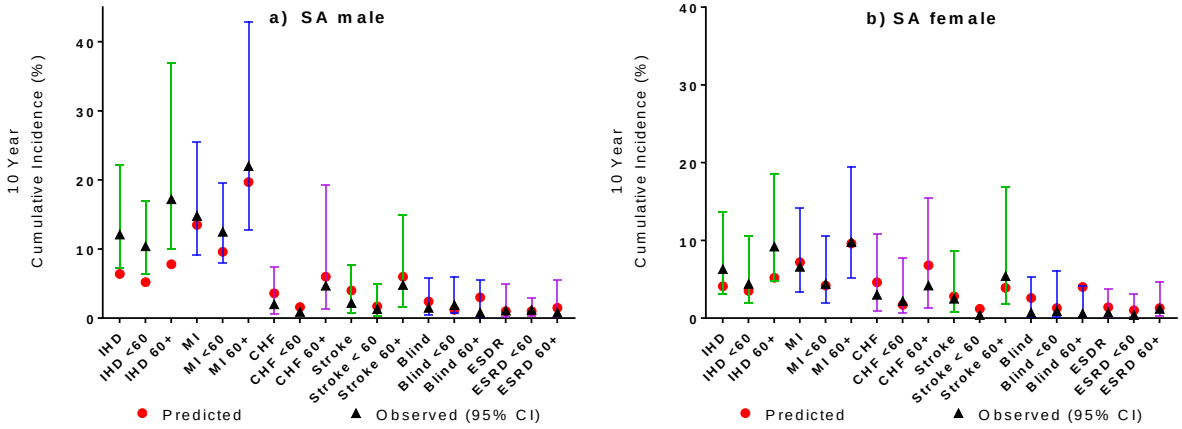
Figure 9: South Asian cohort, by gender (pre and post -event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed



4.3.2.2 South Asian sub-cohorts by age

The next analysis compared the ODEM predicted versus observed events in the SA gender based sub cohorts separated by age, starting with only those with no history of a T2DM related complication (Figures 10a and 10b). The results demonstrated that there were few statistical outliers for any of the vascular events in the SA sub cohorts except for the under estimation of IHD in the SA males (Table 24 and 25). Furthermore, the impact of age in the predicted rate of events in the general population age cohorts that resulted in large over-estimation of the rate of MI, CHF and blindness was not demonstrated in the results of the SA cohort simulation. By contrast, the relative impact of older age in the SA cohorts for all predicted vascular events closely mirrored the relative changes in the observed rate of events. This difference in these age results were in part due to an increase in the actual event rates of MI and blindness combined with a lower predicted event rate for MI, CHF and blindness. These results were showed there was a strong correlation in the ODEM predicted versus observed events by gender and age for a SA population in BC even when there was no history of T2DM related complications at baseline.

Figure 10: South Asian cohort, by gender and age (pre-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed (with 95% CI)



A final simulation included those with a history of T2DM events by gender and age demonstrated similar results as the SA cohort of only pre event comparison (Figure 11a and 11b). The predicted events for MI, stroke and ESRD were highly correlated with the observed rate and within the statistical range set by the 95% CI of the BC results. The ODEM simulation once again showed the model risk equations were responsive to changes to the baseline input variables and the output by age and gender closely mirrored the overall and relative rate of events from the BC real world results (Tables 26 and 27).

Figure 11: South Asian cohort, by gender and age (pre and post-event): Plot of the cumulative incidence (%) of vascular events, predicted versus observed (with 95% CI)

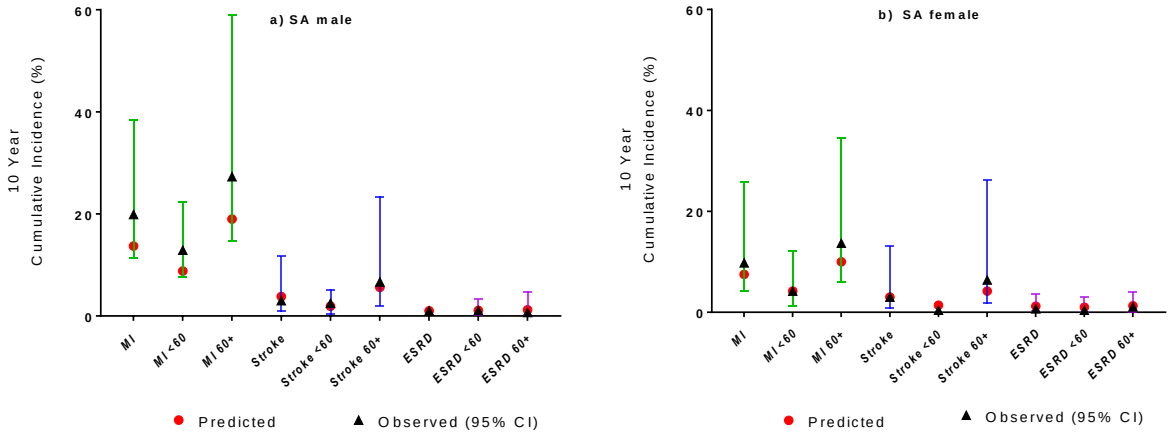


Table 22: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in a SA population pre-event cohort

Type of vascular event	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
IHD	0.093	0.053	0.182	0.051	under
MI	0.109	0.063	0.200	**0.102**	within
CHF	0.025	0.007	0.090	**0.045**	within
Stroke	0.023	0.007	0.082	**0.032**	within
Blindness	0.012	0.003	0.055	**0.024**	within
ESRD	0.009	0.002	0.044	**0.011**	within

Table 23: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in a SA population post-event cohort

Type of vascular event	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
MI	0.151	0.079	0.318	**0.114**	within
Stroke	0.030	0.008	0.124	**0.037**	within
ESRD	0.009	0.002	0.042	**0.011**	within

Table 24: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in SA males by age for pre-event only

Type of events	SA male cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed*
IHD	All Age	0.121	0.073	0.222	0.064	under
	< 60	0.104	0.064	0.170	0.052	under
	60+	0.172	0.100	0.369	0.078	under
MI	All Age	0.148	0.091	0.255	**0.135**	within
	< 60	0.125	0.080	0.195	**0.096**	within
	60+	0.220	0.128	0.429	**0.197**	within
CHF	All age	0.020	0.006	0.074	**0.036**	within
	< 60	0.009	0.003	0.029	**0.016**	within
	60+	0.047	0.013	0.192	**0.060**	within
Stroke	All Age	0.022	0.007	0.077	**0.040**	within
	< 60	0.013	0.003	0.050	**0.017**	within
	60+	0.048	0.016	0.149	**0.060**	within
Blindness	All Age	0.015	0.005	0.058	**0.024**	within
	< 60	0.019	0.007	0.060	**0.012**	within
	60+	0.008	0.001	0.055	**0.030**	within
ESRD	All Age	0.011	0.002	0.050	**0.010**	within
	< 60	0.012	0.003	0.029	**0.010**	within
	60+	0.008	0.001	0.055	**0.015**	within

Table 25 : Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in SA females by age for pre-event only

Type of events	SA female cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus Observed*
IHD	All Age	0.063	0.031	0.137	**0.041**	within
	< 60	0.044	0.020	0.105	**0.035**	within
	60+	0.092	0.048	0.186	**0.052**	within
MI	All Age	0.066	0.033	0.141	**0.072**	within
	< 60	0.044	0.020	0.105	**0.042**	within
	60+	0.098	0.052	0.195	**0.096**	within
CHF	All age	0.030	0.009	0.108	**0.046**	within
	< 60	0.022	0.007	0.077	**0.017**	within
	60+	0.042	0.013	0.155	**0.068**	within
Stroke	All Age	0.025	0.008	0.087	**0.028**	within
	< 60	0.004	0.001	0.031	**0.012**	within
	60+	0.054	0.018	0.169	**0.039**	within
Blindness	All Age	0.007	0.001	0.053	**0.026**	within
	< 60	0.009	0.001	0.061	**0.013**	within
	60+	0.006	0.001	0.042	**0.040**	within
ESRD	All Age	0.007	0.002	0.037	**0.014**	within
	< 60	0.004	0.001	0.031	**0.010**	within
	60+	0.012	0.003	0.047	**0.013**	within

Table 26: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in SA males by age for combined pre and post-event

Type of events	SA male cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed*
MI	All Age	0.199	0.114	0.384	**0.137**	within
	< 60	0.129	0.077	0.224	**0.088**	within
	60+	0.273	0.147	0.589	**0.190**	within
Stroke	All Age	0.030	0.009	0.118	**0.038**	within
	< 60	0.0245	0.0025	0.0515	**0.019**	within
	60+	0.0668	0.0203	0.2340	**0.056**	within
ESRD	All Age	0.010	0.002	0.047	**0.010**	within
	< 60	0.0111	0.0022	0.0355	**0.011**	within
	60+	0.0067	0.0009	0.0478	**0.012**	within

Table 27: Comparison of the cumulative incidence of vascular events from observed (BC data with 95% CI) versus the predicted (ODEM) in SA females by age for combined pre and post-event

Type of events	SA female cohort by age	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Predicted versus observed*
MI	All Age	0.098	0.0422	0.2590	**0.075**	within
	< 60	0.042	0.0175	0.1208	**0.042**	within
	60+	0.137	0.0604	0.3459	**0.100**	within
Stroke	All Age	0.030	0.0083	0.1311	**0.030**	within
	< 60	0.004	0.0006	0.0298	**0.014**	within
	60+	0.064	0.0183	0.2617	**0.042**	within
ESRD	All Age	0.007	0.0015	0.0362	**0.012**	within
	< 60	0.004	0.0006	0.0304	**0.010**	within
	60+	0.011	0.0008	0.0396	**0.013**	within

*Captures whether the ODEM underestimated, overestimated, or were within the 95% CI (confidence interval) of the observed BC data. IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; ESRD = end stage renal disease.

4.3.3 Uncertainty analysis

The ODEM was selected in part due to the ability to incorporate probabilistic analysis in the economic evaluations to assist measures of uncertainty and enable creation of relevant model output. The components of the ODEM which utilize PSA are treatment effects, QALY and costs which were not assessed in this external validation exercise. This research focused on assessing the credibility of the risk equations and predicted event rates from the ODEM which do not utilize probabilistic distributions or PSA. Therefore PSA output while an important component in an economic evaluation could not be utilized in this sensitivity analysis.

To assess the influence of estimated baseline input parameters on the predicted incidence of T2DM related events a deterministic sensitivity analysis was completed on the baseline input variables included in the risk equations for the vascular events that were under or overestimated by the ODEM simulation. This DSA was therefore based on

adjusting the baseline inputs for BMI, smoking rate and level of HbA1c and the duration of diabetes to reflect possible alternative input values for the general, pre-event BC cohort with T2DM in the mid 1990's.

The BMI was a coefficient in the ODEM's risk equation for estimating CHF which was a vascular event that was overestimated in the model simulations. The baseline input for BMI were estimated using self-reported measures in the 1996 NPHS survey in the general Canadian adult population. To adjust the baseline BMI to reflect the higher measure in those with T2DM the estimate from the survey was increased by 4 kg/m^2 based on the literature. There is evidence that BMI may be underestimated in self-reported measures, however Canadian data may overestimate the actual BMI in a BC T2DM population due to evidence of lower provincial BMI rates. Therefore a DSA analysis adjusted the mean BMI up or down by 3 kg/m^2 (Table 28). As anticipated from the risk equations the changes to the baseline BMI was correlated with changes in the predicted incidence of CHF. If the BMI for the general cohort was in actuality lower than the BMI estimated in the ODEM simulation the predicted incidence of CHF would decrease from 0.66 to 0.54 however this was still an overestimation of CHF versus the observed incidence in the BC data of 0.032 (95% CI of 0.024 - 0.044)

Table 28: Sensitivity analysis for BMI ($kg/m^{2)}$

ODEM output by event	@ BMI of 29	@ BMI of 26	@ BMI of 32	Correlation
CHF	0.066	0.054	0.083	yes

CHF = congestive heart failure

Smoking is a coefficient in the risk equation for predicting MI in the ODEM which was a vascular event that was consistently overestimated in the model simulation. The smoking rate used as baseline input parameters in the BC population was also estimated from the NPHS survey in 1996 and captured the national smoking rate in the adult population. The baseline smoking rate was not adjusted to reflect the potential lower smoking rate in BC population due to a lack of BC data identified on smoking in adults from the 1990's. In addition the smoking rates in Canada dropped significantly for both genders from the early 1990's into the mid 2000's. There is a risk equation in the ODEM to reflect a reduced smoking rate with increasing age and increasing duration of T2DM however the link to this time varying smoking rate in the risk equations for stroke could not be located in the ODEM. Based on this, it is probable that the smoking rate in the BC population with T2DM in the 1990's was actually lower than the baseline rate input into the ODEM simulation and did not decrease over time in the simulation. A sensitivity analysis was done which adjusted the smoking rate up or down by an absolute rate of 10% to assess the impact on the predicted incidence of MI (Table 29). Surprisingly the predicted incidence of MI did not decrease in response to the 15% smoking rate at baseline however it did increase with higher smoking rates of 36%. Therefore if the smoking rates in the BC population were in fact lower than estimated for the ODEM simulation it would not have resolved the over-estimation of MI versus the observed incidence in the BC data of 0.104 (95% CI of 0.088 to 0.120).

Table 29: Sensitivity analysis for smoking rate

ODEM output by event	@ base rate of 24.5%	@ rate of 15%	@ rate of 36%	Correlation
MI	0.132	0.134	0.148	mixed

MI = myocardial infarction

A time varying measure of T2DM control, HbA1c, was included as a coefficient in the risk equations used to estimate IHD, MI, CHF and blindness. The actual baseline measures of these TVRF's were not available in the aggregated BC diabetes results and required extrapolation from Canadian primary care data in the early 2000's. These data were anticipated to be a reasonable estimate of the HbA1c in the BC diabetes population in the mid 1990's, however due to selection bias of physicians well versed and motivated to control T2DM in their patients; it is possible that the measure of HbA1c in reality could have been higher. The HbA1c estimates from two Canadian primary care databases were used in this sensitivity analysis. One study showed a mean HbA1c in a primary care registry of 6.9% while another from an observational Ontario dataset from 2000/01 had a mean HbA1c of 8.14% (Table 30).(41) This sensitivity analysis showed the changes in HbA1c positively correlated with the predicted incidence of IHD, MI, CHF and blindness and therefore further highlighted the importance of an accurate mean baseline HbA1c. Therefore, if the actual HbA1c levels in the BC T2DM population at baseline were lower than the mean HbA1c of 7.3% level used in the ODEM simulation the correlation of the predicted and observed events for MI, CHF and blindness from the ODEM would improve but would still not have been statistically within the 95% CI.

Table 30: Sensitivity analysis for baseline HbA1c

ODEM output by event	Base = 7.3%	Lower = 6.9%	Higher = 8.14%	Correlation
IHD	0.053	0.051	0.054	yes
MI	0.132	0.132	0.143	yes
CHF	0.066	0.060	0.071	yes
Blind	0.038	0.034	0.039	yes

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Blind = blindness in one eye

The DofD in the baseline input variables is directly linked to calculating the age in years at the diagnosis of diabetes and the current age used in the ODEM simulation. Therefore a change to one of these variables automatically influences the others. The internal validation exercise (Chapter 3) demonstrated that increasing the DofD decreased the estimated incidence of all T2DM related complications. Considering the duration of T2DM was estimated from Canadian population data using the prevalence and incidence by age it is possible that the mean estimated DofD (Table 12) differed from the actual mean duration in the BC population from the early 1990's. To assess this, the mean DofD was adjusted up and down by 3 years, based on the differences seen between the 9.7 year mean calculated for baseline versus the mean duration of 6 to 7.8 years in the Canadian primary care data used to capture HbA1c in the baseline and sensitivity analysis (Table 31).(87,88) The longer DofD of 12.7 years reduced the predicted incidence of MI, CHF and blindness and put these results closer to the observed incidence of

events however this did not result in the predicted being within the 95% CI of the observed rate of MI, CHF or blindness. Therefore, if the mean duration in the BC T2DM data was longer than 9.7 years, the correlation of the predicted events from the ODEM and the observed events from the BC data would have improved but would still be within the statistical range.

Table 31: Sensitivity analysis for duration of diabetes

ODEM output by event	DofD = 9.7 years	DofD = 6.7 years	DofD – 12.7 years	Correlation
IHD	0.053	0.049	0.050	no
MI	0.132	0.146	0.130	yes
CHF	0.066	0.069	0.063	yes
Blind	0.038	0.039	0.034	yes

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Blind = blindness in one eye

The sensitivity analysis raised a number of insights when considering the results of this external validation. While the literature was searched to identify the most relevant data to estimate the baseline input variables it was likely that the actual baseline variables in the BC population in the 1990's did differ somewhat from these estimated baseline values. Considering the complexity of the ODEM methods, input variables and risk equations it was possible that changes to one variable at a time in the sensitivity analysis would not have a large impact on the predicted results. Therefore a multivariate sensitivity analysis which adjusted all three of the estimated baseline BMI, HbA1c and DofD input parameters to determine if all baseline inputs adjusted at once would reduce the over estimation of the predicted incidence of MI, CHF, stroke and blindness. This multivariate sensitivity analysis showed that the correlation of the predicted and observed measures of MI, CHF and blindness improved to the point that the predicted incidence of MI was no longer overestimated but within the statistical range of the BC data (Table 32). This multivariate sensitivity analysis supported the need for an external validation to be ideally performed with real world data which includes the actual baseline input parameters from the population in which the observed rate of T2DM vascular events had been captured.

Table 32: Multivariate DSA of three input parameters

Type of Event:	Observed incidence (BC data)	Lower bound of the 95% CI	Upper bound of the 95% CI	Predicted incidence (ODEM)	Multivariate DSA (ODEM)
IHD	0.080	0.066	0.098	0.053	0.047
MI	0.104	0.088	0.120	0.132	**0.117**
CHF	0.032	0.024	0.044	0.066	0.050
Stroke	0.045	0.034	0.058	**0.052**	**0.044**
Blindness	0.005	0.002	0.010	0.038	0.031
ESRD	0.008	0.004	0.015	**0.009**	**0.010**

IHD = ischemic heart disease; MI = myocardial infarction; CHF = congestive heart failure; Blind = blindness in one eye; ESRD = end stage renal disease

4.4 Discussion

The external validation of the ODEM demonstrates a strong overall correlation of the predicted and observed rate of vascular events when including the overall total population of all individuals in the T2DM cohort. These results alone would provide researcher and decision makers with a high level of confidence in the model output for the purpose of an economic evaluation of T2DM in a BC population. It is important to recognize that the predicted output in cohorts that included individuals with a history of events at baseline would be more reflective of a diverse, real world BC population of people with T2DM. Additional positive outcomes highlighted in the validation results include the high predictive correlation of the absolute incidence of stroke and ESRD in all cohorts regardless of gender or age, as well as the predicted rate of MI and CHF in the younger age cohorts that is within, or close to within, the statistical range. The model also was sensitive to gender and age inputs parameters that changed the relative rate of events between male and female, younger and older, in a way that mirrored the relative change in the observed rates from the BC results.

Interestingly, a clear trend that led to the lower correlation in ODEM versus BC results was due almost solely to the over prediction of MI, CHF and blindness in the older age cohort. This is an encouraging result of the external validation as it narrows the potential need for model calibration to a few potential risk variables and associated coefficients in the ODEMs risk equations. This trend toward an over estimation of vascular events has been noted in other external and cross validation studies in T2DM models which utilize the UKPDS risk equations. (36,42,49,50,52,63). This could be a reflection of improved treatment standards which, over a longer term of treatment, lead to a significant improvement in the rate of vascular events in an older cohort which may not be evident in the younger population. This theory is supported by an identified trend toward a lower rate of MI and the decreased age standardized mortality from heart disease and stroke, noted in data from the Canadian Institute for Health Information and Statistics Canada, with BC consistently showing it has the lowest rate in the country. As previously stated, this over prediction of events could also be due to the influence of time varying increases in metabolic measures of control which are not reflective of dynamic treatment programs and incremental interventions over time in the real world T2DM population. The use of progressive time varying risk factors in the ODEM should be validated upon the availability of a linked and longitudinal BC dataset which includes actual metabolic measures over time.

In the validation studies of all of the T2DM reviewed there was no threshold of the measure of correlation defined, between predicted and observed events, for a model to be considered validated for use in the population of interest. However, if a correlation coefficient of 0.7 which calculates to an R^2 of 0.49 was set as a threshold to determine a positive correlation, then the cohorts assessed in the ODEM by gender and ethnicity would exceed these requirements with the exception of the female only cohort with no previous history of events ($R^2 = 0.45$). Of

note in this female cohort the predicted rate of MI was within the 95% CI of the observed data, which was an improvement compared to a similar male only cohort, however measure of correlation of the predicted versus observed events was lower in women due primarily to the over prediction in the rate of CHF.

Finally, this inclusion of the SA cohorts was meant to provide greater insight on the predictive ability of the ODEM in an ethnic sub population that had evidence of a differing epidemiology, risk factors and progression of T2DM. There were some important differences noted in the baseline input parameters, including a younger mean age, longer duration of diabetes and lower smoking rate which would influence the predicted rate of events. This external validation of the ODEM in the SA population was quite encouraging and somewhat surprisingly resulted in a higher correlation between the predicted and observed incidence of vascular events versus the general population (R^2 of 0.77 versus 0.68 respectively). In the SA cohort the predicted incidence of all of the vascular events were within the statistical bounds of the observed BC data with the exception of an under estimation of IHD in SA males. A partial reason for the improvement in the correlation of the ODEM in the SA cohort was due to a higher incidence of IHD, MI and blindness and a decreased rate of CHF and stroke in the real world results for those of SA descent in BC. In summary, while the results of the validation in the SA cohorts should be interpreted with caution due to the reduced number of individuals and fewer absolute number of T2DM related events, the ODEM demonstrated the methods used in this model effectively incorporated the differing baseline input variables to predict the absolute and relative changes in vascular events in an SA population.

Table 33: Summary of the external validation results by event, gender, ethnicity and age cohorts

Cohort	Coefficient of determination = R^2	Overestimated by the ODEM	Underestimated by the ODEM	# predicted within 95% CI of observed
All pre event	0.68	MI, CHF, blindness	IHD	2/6
All - Pre and post	0.99	none	none	3/3
Male	0.71	MI, CHF, blindness	IHD	2/6
Female	0.45	CHF, blindness	IHD	3/6
SA – all pre	0.77	none	IHD	5/6
SA – all pre/post	0.99	none	none	6/6
Male under 60	N/A	blindness	IHD, MI	3/6
Female under 60	N/A	blindness	IHD, MI	3/6
Male 60+	N/A	MI, CHF, blindness	IHD	2/6
Female 60+	N/A	MI, CHF, blindness	IHD	2/6

SA = South Asian, MI = myocardial infarction, CHF = congestive heart failure, IHD = ischemic heart disease, N/A = not calculated

Chapter 5: Summary

The growing prevalence of T2DM and its associated vascular complications has led to greater interest from decision makers on the role of economic evaluation to enhance decisions and the appropriate allocation of our finite health care resources. The development of credible and trusted CEA requires access to validated decision models to measure the incremental cost and effectiveness of new interventions versus an existing standard of care. There are multiple existing T2DM models available which meet the criteria for the comparative evaluation of T2DM interventions in Canada and researchers continue to revise their models to incorporate new evidence and to improve the functionality and validity of output. There is a trend toward developing models with the ability to run individual micro-simulation and probabilistic analysis to allow researchers to measure the joint uncertainty of model output. These models should be screened and selected based on their perspective, scope, methods and structure that are considered to be most suitable for the purpose of the economic evaluation. The availability of a validated model provides local decision makers with a valuable tool in their efforts to support a sustainable health system.

The selection and internal validation of the ODEM demonstrates that this is a highly functional model that is sensitive to individual changes to the model input parameters. This internal validation exercise supports the use of risk equations in the ODEM as an effective method to capture a wide range of individual characteristics including time varying and pre-existing risk factors required for modelling a complex, chronic disease such as T2DM. While this internal validation increases the confidence in the functionality of the ODEM as a simulation tool it does not provide insight on how well the model replicates reality in T2DM.

The overarching objective of this thesis is to identify and assess an existing T2DM model using an external validation against real world data to enhance the credibility of CEA output for use in British Columbians. Of the ten T2DM models selected for further analysis of their methods none were created or validated using large scale RCT or observational data from a contemporary Canadian population. The availability of real world aggregate results from a BC T2DM population provides a unique opportunity to address two previously identified knowledge gaps. These include the ability to assess whether the ODEM's absolute predicted rate of vascular events are reflective of the real world rate of vascular events and to assess the generalizability to a BC T2DM population. For this reason the external validation of the ODEM incorporates a number of sub cohorts based on gender, age and ethnicity to enable an in-depth analysis of populations with unique risk factors and differing rates of vascular events.

5.1 Strengths and limitations of this research

Model based economic evaluation is intended to provide a valuable tool to measure and enhance the efficient allocation of resources. Therefore, it is relevant that decision makers understand the strengths and limitations of the decision model being used and the uncertainty of output provided in order to properly interpret the results. The ODEM is one of the T2DM models with methods which meet the criteria outlined as optimal for developing model based CEA for use in a diverse, Canadian population. Additionally, a primary strength of the ODEM is that the predicted rates of vascular events are based on the risk equations from the UKPDS RCT. The UKPDS is recognized as a landmark study which has provided invaluable, high quality evidence of the impact of pre-existing and time varying risk factors on the rate of vascular events in T2DM. These risk equations are now utilized in many existing T2DM models and therefore have been assessed via numerous internal and external T2DM model validations. The use of the UKPDS-OM1 risk equations in the ODEM is considered the best available evidence for predicting T2DM, however these equations may require revision or recalibration to reflect today's standard of care and the resultant improvement in the prevention of vascular complications related to T2DM.

Another strength of the ODEM is that the costs of T2DM have been updated based on a large, validated Ontario diabetes database that has captured the health resource utilization associated with treating T2DM in Canada. This data includes the costs associated with the management of T2DM for individuals with no previous complications, the cost of a fatal or non-fatal vascular event and the incremental cost of management of the complications of T2DM after a vascular event. This is an important enhancement for the ODEM and provides a measure of incremental cost in a CEA based on data from a Canadian health system.

Many of the limitations of the ODEM and the results of this external validation are a reflection of the limited data, methods and assumptions required to simplify the complexity of T2DMwhich results in uncertainty in the model output. These limitations may not be completely overcome however they can be mitigated with greater access to more robust, real world datasets to enable ongoing validation and calibration of the ODEM risk equations.

Several limitations are related to the structure and methods used in the ODEM which are typical of decision models created for use in T2DM. Firstly, as the ODEM is a closed cohort model it does provide valuable evidence for the comparative evaluation of T2DM interventions, however the model methods do not account for the epidemiological trends such as increasing prevalence of T2DM over time. Secondly, the ODEM does not explicitly capture the early signs of the onset of vascular disease such as neuropathy, cataracts and albuminuria. These events are likely to reduce QALY and increase costs of care and can be indirectly captured in the general cost and QALY measures for the ongoing management of T2DM. A third limitation in the ODEM methods is that the model was not designed to capture subsequent events beyond the first event and therefore may underestimate the

cumulative incidence of events. This has been previously identified as a limitation of the UKPDS-OM1 however further analysis shows this does not have a major impact on validation results. (40) Finally, with respect to the ODEM methods, there are potential limitations in the use of the UKPDS-OM1 risk equations, as previously outlined, to calculate the rate of vascular events and the progression of time varying risk factors in a different population with a different standard of care. These risk equations require calibration against larger, more recent data which have captured a larger number of less common events such as ESRD and longer follow up to ensure these calculations are reflective of today's management of T2DM.

A number of limitations of this thesis are due to the limited data availability to design and run this external validation. Firstly, the available BC data only included the aggregate results of the mean rate, and associated confidence intervals, of vascular events by cohort. In addition a number of baseline input variables were not available from the aggregate BC results including the DofD, weight, height, HbA1c, SBP, TC, HDL or smoking rate and therefore are estimated and extrapolated from available public data and published literature. The estimation of mean and random number distribution of baseline input variables to create a virtual cohort of individuals for each micro simulation adds to the first order uncertainty in the ODEM's predicted rate of vascular events. Based on the results of a sensitivity analysis, this uncertainty with the input variables could be partially responsible for the over estimation of MI, CHF and blindness if the estimated baseline mean BMI, smoking rate and HbA1c were in reality lower, and the duration of diabetes was in reality longer, in the BC T2DM cohort. Secondly, there may also be limitations in interpreting the results due to differing definitions of vascular complications used in the formation of the risk equations in the ODEM and the definitions to compile the aggregate BC data. For instance, the ICD9 codes described in the UKPDS RCT differ slightly from the ICD9 codes used to identify events in the BC diabetes data based on the comparison (Table 10). It is possible that misallocation of ICD9 codes could contribute to the under prediction of IHD and the over prediction of MI however this scenario appears unlikely as the ODEM over estimation of the incidence of MI was only evident in an older cohort and not pervasive in all cohorts. Thirdly, a limitation of the available data is due to the impact of competing events such as mortality on the incidence of subsequent vascular events, as mortality rates could not be directly compared. It is feasible that a lower predicted mortality in the ODEM output could be in part responsible for the higher incidence of MI and CHF versus the observed data. This outcome was not consistently observed across vascular events, for instance with IHD or stroke, which limits the likelihood that mortality difference impact the accuracy of the ODEM. Finally, there is limitation in the interpretation of the results for the SA cohorts due to the fewer number of individuals and low number of certain vascular events which results in a wide 95% CI. The wider 95% CI may overstate the predictive accuracy of the ODEM in the SA T2DM population. Finally, the baseline input parameters for the SA population which are required to be extrapolated from the estimated baseline inputs of the general population also potentially increases the first order uncertainty of the model output. This should be further assessed using an observational dataset with the actual baseline input values for all sub-cohort simulations.

The structure of this external validation also presents some limitations on the interpretation of the ODEM output. Firstly, while the ODEM is designed to allow the application of probabilistic distributions to treatment effects, costs and QALY's to produce a PSA on model outputs, these variables are not assessed in this external validation. Therefore, the validation does not provide a measure of the joint uncertainty of the economic output of the ODEM. Secondly, the external validation of the ODEM utilized the methods to calculate the progression of HbA1c, SBP and cholesterol over time which may contribute to the over-prediction of vascular complications. To measure the potential impact of non-progressive time varying risk factors, the general cohort was re-simulated in the ODEM to determine if this would improve the predictive ability of the ODEM (Table 34). Of note, the over prediction of MI is no longer above the 95% CI of observed data and the over prediction of CHF is not as extreme in the results of this simulation when based on non-progressive metabolic risk factors.

Table 34: Comparison of progressive vs non progressive TVRF's on predicted versus observed events

Vascular outcome	Predicted, progressive TVRF	Predicted, non-progressive TVRF	Observed (BC Data)	Predicted vs Observed
IHD	0.053	0.041	0.080	below
MI	0.132	0.108	0.104	within
CHF	0.066	0.057	0.032	above
Stroke	0.052	0.038	0.045	within
Blindness	0.038	0.034	0.005	above
ESRD	0.009	0.007	0.008	within

TVRF = time varying risk factor, IHD = ischemic heart disease, MI = myocardial infarction, CHF = congestive heart failure, ESRD = end stage renal disease

5.2 Questions for future research

There are a few additional components to assess and to update in the ODEM prior to broad use in BC for the economic evaluation.

1) Will a second validation using real world BC data confirm under estimation of the rate of IHD, and an over estimation in the rate of MI, CHD and blindness to support re-calibration of the ODEM risk equations?
2) Do the actual measures of the TVRF's, specifically HbA1c and SBP, reflect the actual progression of these metabolic measures in a real world, BC T2DM population?
3) Do the cost inputs used in the ODEM reflect BC health system costs for the ongoing treatment of T2DM, vascular events and subsequent management of complications post vascular event including a fatal and non-fatal outcome?

4) Do the EQ-5D measures currently used to calculate the QALY in the ODEM provide the same measure of incremental effectiveness as the updated EQ-5D Canadian based utility measures? Do other utility measures, such as HUI3 in the ODEM impact the economic output of the ODEM?
5) Does a recalibration of the ODEM's risk equations, based on this external validation of the absolute incidence of vascular events, impact the results of a CEA which measures the incremental cost and effectiveness based on the relative change in the incidence of vascular outcomes in response to a T2DM intervention?

5.3 Next steps

This is the first known validation of the ODEM using real world data from a Canadian population with T2DM. The cohorts that include all pre and post-event individuals by age, gender and ethnicity demonstrate a very high correlation and statistical accuracy however sub cohort results identified some notable trends. There is a weaker correlation, and lower statistical accuracy of the ODEM predicted events versus the BC observed events, in the older cohort and in those without a history of vascular events that must be confirmed or resolved via calibration of the model risk equations.

While the overall external validation results are supportive of the generalizability of the ODEM for use in a BC population these assessments should be replicated using another observational dataset prior to adjusting or re-calibrating the risk equations in the ODEM. Ideally, this next dataset would include greater access to actual baseline metabolic measures, time varying risk factors and pre-existing conditions of the individuals in each cohort to reduce first and second order uncertainty in the ODEM baseline input variables. If the trends identified from this real world external validation remain consistent, adjustment of the ODEM risk equations to reflect the BC T2DM population would provide a highly credible decision model for local decision makers responsible for resource allocation in the BC population diagnosed with T2DM.

This external validation of the predicted absolute rate of vascular events for a real world BC T2DM population is an important first step toward the creation of a BC Diabetes Economic Model (BCDEM). Next, the model parameters which incorporate probabilistic distribution and enable a PSA in the model output should be validated and revised as follows. The results of a contemporary RCT could be used to update the treatment effects, including the time varying changes in metabolic measures, to validate the model's ability to predict the relative rate of vascular events as a result of a T2DM intervention. To ensure the economic output is reflective of a BC environment, the health utilization and associated costs of diabetes management, vascular events and post event management could be updated to represent BC health care costs in 2017 Canadian dollars. Next, the QALY decrements in the ODEM could be revised to reflect EQ-5D and /or HUI3 health utility measures which have been measured or validated for use in a contemporary Canadian population. Finally, this newly formed BCDEM model could be

utilized as a reference model for BC researchers and decision makers to provide economic evaluation of new intervention or to cross validate the CEA generated by other T2DM models.

5.4 Conclusion

The growing concern with the sustainability of the current Canadian health care system is leading to an increased role for economic evaluation of health interventions to guide the allocation of our finite resources. Model based economic evaluation is an efficient and effective mechanism to enhance the ability of decision makers to utilize an evidence based, systematic approach to enhance their decision making under uncertainty. It is critical that decision makers have trust and confidence in these decision models to provide economic output that is relevant for their population. This requires greater transparency on behalf of the researchers who develop these models, to allow others full insight into the current methods used and to enable greater access for ongoing validation of the models relevance.

A number of observations became increasingly evident during this research project. Firstly, T2DM is a very complex condition that can be effectively replicated, using a mathematical representation, to predict vascular events over time. Considering the complexity of T2DM, the consistency and quality of the results of these long term simulations were impressive and demonstrate the credibility of models used in economic evaluation of T2DM. That said, there are other conditions related to elevated blood glucose and cardio metabolic disease which are not captured in T2DM models, including a higher rate of infections and malignancy, which may need to be incorporated to assess the overall value of interventions as more evidence emerges. Secondly, model methods used in T2DM are becoming more aligned over time, with trends toward increasing use of microsimulation and PSA. This will be further supported by new Canadian HTA guidelines which recommend more consistency in describing the decision problem and use of a reference case to enhance the trust and credibility in model performance as approaches to model based economic evaluation become more standardized.

As for model validation, it was evident in these analyses that a standard approach to comparing the predicted output to observed events needs to be adopted. There are many statistical approaches and goodness of fit measures used to assess the correlation of the events. The use of the coefficient of determination (R^2) was the most widely utilized; however, this may not provide the most relevant result and can possibly be misleading for non- linear output. There is a growing interest in intra-class correlation (ICC) as a measure of agreement, which was not used in the validation of assessed T2DM models, for comparing continuous outputs which may be worthy of consideration for future analyses.(90)

In summary, to continue to build trust and confidence in the role of decision modeling methods, output and overall utilization for guiding efficient resource allocation, model developers are required to allow greater access to their models. In this analysis of ten T2DM models, and application of objective criteria for the selection of an appropriate model for CEA, only the ODEM was provided in response to our request for full access for in depth analysis and model validation. While the general understanding of the model methods can be gathered from published literature, it was only with greater access to full model functionality that we could fully appreciate the models capabilities. It is understandable that model developers require some proprietary protection for the time, energy and resources invested in developing a credible model; however, this needs to be balanced with greater transparency and access to advance the role of model-based economic evaluation. Ideally, these high quality models would be provided as open access models for research purposes to deepen the understanding of the methods used, allow broader comparison of model performance and enhance the generalizability of model output for use in contemporary, diverse, real world populations.

Bibliography

1. Public Health Agency of Canada. Chronic Disease Risk Factor Atlas. Ottawa, ON: Government of Canada; 2008.

2. Światowa Organizacja Zdrowia. Global status report on noncommunicable diseases 2010. Geneva: World Health Organization; 2011.

3. Katzmarzyk PT, Ardern CI. Overweight and obesity mortality trends in Canada, 1985-2000. Can J Public Heal Can Santee Publique. 2004;16–20.

4. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. [Internet]. Ottawa, Ont.: Canadian Agency for Drugs and Technologies in Health; 2006 [cited 2016 Jul 16]. Available from: http://www.deslibris.ca/ID/203719

5. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value Health. 2012 Sep;15(6):796–803.

6. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Can J Diabetes. 2013;37 (suppl 1):S1–212.

7. Pelletier C, Dai S, Roberts KC, Bienek A, others. Report summary Diabetes in Canada: facts and figures from a public health perspective. Chronic Dis Inj Can [Internet]. 2012 [cited 2016 Apr 30];33(1). Available from: http://search.proquest.com/openview/f3d1fef5a798f9e64c92a113a5d098bb/1?pq-origsite=gscholar

8. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. Diabetes Care. 2004;27(9):2262–5.

9. Public Health Agency of Canada. Diabetes in Canada: Facts and figures from a public health perspective. Ottawa ON: PHAC; 2011.

10. Dyck R, Osgood N, Lin TH, Gao A, Stang MR. Epidemiology of diabetes mellitus among First Nations and non-First Nations adults. Can Med Assoc J. 2010 Feb 23;182(3):249–56.

11. Chiu M, Austin PC, Manuel DG, Tu JV. Comparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. Can Med Assoc J. 2010 May 18;182(8):E301–10.

12. Gilmore J. Report on smoking prevalence in Canada, 1985 to 1999 [Internet]. Statistics Canada, Health Statistics Division; 2000 [cited 2016 Oct 13]. Available from: http://www.statcan.gc.ca/pub/82f0077x/4193722-eng.pdf

13. British Columbia Population Projections 2016-2041.pdf [Internet]. BC Stats: Population Projections 2016 - 2041. 2015. Available from: www.bcstats.gov.bc.ca/StatisticsBySubject/Demography/ PopulationProjections.aspx

14. Statistics Canada: Section 3 – Results at the provincial and territorial levels, 2013 to 2038 [Internet]. Statistics Canada. 2015. Available from: http://www.statcan.gc.ca/pub/91-520-x/2014001/section03-eng.htm#a12

15. Statistics Canada: NHS Focus on Geography Series: British Columbia [Internet]. Statistics Canada. 2016. Available from: https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/fogs-spg/Pages/FOG.cfm?lang=E&level=2&GeoCode=59

16. Diabetes: Canada at the tipping point: Charting a New Path [Internet]. Canadian Diabetes Association, Diabetes Quebec; Available from: https://www.diabetes.ca/CDA/media/documents/publications-and-newsletters/advocacy-reports/canada-at-the-tipping-point-english.pdf

17. Canadian Diabetes Association. An economic tsunami: the cost of diabetes in Canada [Internet]. 2009 Dec. Available from: https://www.diabetes.ca/publications-newsletters/advocacy-reports/economic-tsunami-the-cost-of-diabetes-in-canada

18. Canadian Institute for Health Information. National Health Expenditure Trends, 1975 to 2015. Ottawa, ON: CIHI; 2015.

19. Setting Priorities for the BC Health System - February 2014.pdf. British Columbia: BC Ministry of Health; 2014.

20. Canadian Agency for Drugs and Technologies in Health. Third-Line Pharmacotherapy for Type 2 Diabetes — Update. Ottawa: The Agency; (CADTH optimal use report; vol.3, no. 1b); 2013.

21. Rao DP, Dai S, Lagacé C, Krewski D. Metabolic syndrome and chronic disease. Chronic Dis Inj Can [Internet]. 2014 [cited 2016 Apr 30];34(1). Available from: http://search.proquest.com/openview/7f531531eff05244b0dc82580d0ee6c1/1?pq-origsite=gscholar

22. Palmera 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 Jan;20(s1):S27–40.

23. Rana A, de Souza RJ, Kandasamy S, Lear SA, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. CMAJ Open. 2014 Sep 22;2(3):E183–91.

24. Gupta M. South Asians and Cardiovascular Risk: What Clinicians Should Know. Circulation. 2006 Jun 27;113(25):e924–9.

25. Leiter LA, Fitchett DH, Gilbert RE, Gupta M, Mancini GBJ, McFarlane PA, et al. Cardiometabolic Risk in Canada: A Detailed Analysis and Position Paper by the Cardiometabolic Risk Working Group. Can J Cardiol. 2011 Mar;27(2):e1–33.

26. Physicians for a Smoke-Free Canada. Smoking in Canada 2005.pdf. Ottawa, Ontario; 2005.

27. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet. 1998 Sep 12;352(9131):837–53.

28. Gerstein, H_Lacet 2014, 384_ACCORD study_Effects of intensive glycemic control in diabets.pdf.

29. Patel A, MacMahon S, Chalmers J, Billot L, Woodward M, Marre M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. 2008 [cited 2015 Nov 5]; Available from: http://repository.ubn.ru.nl/handle/2066/69678

30. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). The Lancet. 2002;360(9334):711–715.

31. Walker A, Colagiuri S, others. Cost-Benefit Model System of Chronic Diseases in Australia to Assess and Rank Prevention and Treatment Options. In: International Microsimulation Association conference, Vienna, Austria [Internet]. 2007 [cited 2016 Jul 16]. p. 20–22. Available from: http://www.microsimulation.org/IJM/V4_3/Walker%20Paper%20v3.pdf

32. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006. 237 p. (Oxford handbooks in health economic evaluation).

33. 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 Sep;15(6):843–50.

34. Drummond MF, editor. Methods for the economic evaluation of health care programmes. 3. ed., reprint. Oxford: Oxford Univ. Press; 2007. 379 p. (Oxford medical publications).

35. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value Health. 2012 Sep;15(6):835–42.

36. McEwan P, Foos V, Palmer JL, Lamotte M, Lloyd A, Grant D. Validation of the IMS CORE Diabetes Model. Value Health. 2014 Sep;17(6):714–24.

37. Homer J, Jones A, Seville D, Essien J, Milstein B, Murphy D. The CDC's diabetes systems modeling project: developing a new tool for chronic disease prevention and control. In: 22nd International Conference of the System Dynamics Society [Internet]. 2004 [cited 2016 Jan 1]. p. 25–29. Available from: http://www.researchgate.net/profile/Andrew_Jones9/publication/228543463_The_CDC's_diabetes_systems_modeling_project_developing_a_new_tool_for_chronic_disease_prevention_and_control/links/0912f508aad061cc86000000.pdf

38. 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 Oct;17(10):701–11.

39. McEwan P, Ward T, Bennett H, Bergenheim K. Validation of the UKPDS 82 risk equations within the Cardiff Diabetes Model. Cost Eff Resour Alloc [Internet]. 2015 Dec [cited 2016 Jan 1];13(1). Available from: http://resource-allocation.biomedcentral.com/articles/10.1186/s12962-015-0038-8

40. 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 Oct;47(10):1747–59.

41. O'Reilly D, Hopkins R, Blackhouse G, Clarke P, Hux J, Guan J, et al. Development of an Ontario Diabetes Economic Model (ODEM) and application to a multidisciplinary primary care diabetes management program. Programs for Assessment of Technology in Health; 2007.

42. 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. Zhang H, editor. PLoS ONE. 2014 Oct 13;9(10):e110235.

43. Schlessinger L, Eddy DM. Archimedes: a new model for simulating health care systems—the mathematical formulation. J Biomed Inform. 2002;35(1):37–50.

44. Eddy DM, Schlessinger L. Validation of the Archimedes diabetes model. Diabetes Care. 2003;26(11):3102–3110.

45. Thurecht L, Brown L. Economic Modelling of Prevention of Diabetes in Australia. Int J MICROSIMULATION. 2011;4(3):71–80.

46. McEwan P, Williams JE, Griffiths JD, Bagust A, Peters JR, Hopkinson P, et al. Evaluating the performance of the Framingham risk equations in a population with diabetes. Diabet Med. 2004 Apr;21(4):318–23.

47. The Mount Hood 4 Modeling Group. Computer Modeling of Diabetes and Its Complications: A report on the Fourth Mount Hood Challenge Meeting. Diabetes Care. 2007 Mar 2;30(6):1638–46.

48. Zhou H, Isaman DJ, 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(12):2856–2863.

49. Hoerger T, Segel J, Zhang P, Sorensen S. Validation of the CDC-RTI Diabetes Cost-Effectiveness Model [Internet]. Research Triangle Park, NC: RTI Press; 2009 Sep [cited 2014 Oct 27]. Available from: http://www.rti.org/pubs/mr-0013-0909-hoerger.pdf

50. Palmer AJ. Computer Modeling of Diabetes and Its Complications: A Report on the Fifth Mount Hood Challenge Meeting. Value Health. 2013 Jun;16(4):670–85.

51. Dunstan DW, Zimmet PZ, Welborn TA, Cameron AJ, Shaw J, De Courten M, et al. The Australian diabetes, obesity and lifestyle study (AusDiab)—methods and response rates. Diabetes Res Clin Pract. 2002;57(2):119–129.

52. Leal J, Hayes AJ, Gray AM, Holman RR, Clarke PM. Temporal Validation of the UKPDS Outcomes Model Using 10-Year Posttrial Monitoring Data. Diabetes Care. 2013 Jun 1;36(6):1541–6.

53. 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 Sep;15(6):843–50.

54. Asche CV, Hippler SE, Eurich DT. Review of Models Used in Economic Analyses of New Oral Treatments for Type 2 Diabetes Mellitus. PharmacoEconomics. 2014 Jan;32(1):15–27.

55. Tarride J-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(4):255–277.

56. 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. Value Health. 2012 Sep;15(6):804–11.

57. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. J Health Serv Res Policy. 2004;9(2):110–118.

58. Caro JJ, Möller J, Getsios D. Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations? Value Health. 2010 Dec;13(8):1056–60.

59. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario Determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care. 2002;25(3):512–516.

60. Clarke PM, Gray AM, Briggs A, Stevens RJ, Matthews DR, Holman RR, et al. Cost-utility analyses of intensive blood glucose and tight blood pressure control in type 2 diabetes (UKPDS 72). Diabetologia. 2005 May;48(5):868–77.

61. Tao L, Wilson ECF, Griffin SJ, Simmons RK. Performance of the UKPDS Outcomes Model for Prediction of Myocardial Infarction and Stroke in the ADDITION-Europe Trial Cohort. Value Health. 2013 Sep;16(6):1074–80.

62. O'Reilly D, Hopkins R, Blackhouse G, Clarke P, Hux J, Tarride J-E, et al. Long-term cost-utility analysis of a multidisciplinary primary care diabetes management program in Ontario. Can J Diabetes. 2007;31(3):205–214.

63. 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 Aug;16(8):1007–21.

64. Jaime Caro J, 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 Health. 2014 Mar;17(2):174–82.

65. Gaebler J, Blickensderfer, Hoogwerf, Han, Gaebler J, Alperin, et al. Health and economic outcomes for exenatide once weekly, insulin, and pioglitazone therapies in the treatment of type 2 diabetes: a simulation analysis. Vasc Health Risk Manag. 2012 Apr;255.

66. Colagiuri S, Walker AE. Using An Economic Model Of Diabetes To Evaluate Prevention And Care Strategies In Australia. Health Aff (Millwood). 2008 Jan 1;27(1):256–68.

67. Najafzadeh M, Marra CA, Lynd LD, Sadatsafavi M, FitzGerald JM, McManus B, et al. Future Impact of Various Interventions on the Burden of COPD in Canada: A Dynamic Population Model. de Torres JP, editor. PLoS ONE. 2012 Oct 11;7(10):e46746.

68. Ransom T, Goldenberg R, Mikalachki A, Prebtani APH, Punthakee Z. Reducing the Risk of Developing Diabetes. Can J Diabetes. 2013 Apr;37:S16–9.

69. McCabe MC, Dixon S. Testing the validity of cost-effectiveness models. Pharmacoeconomics. 2000;17(5):501–513.

70. Karnon J, Vanni T. Calibrating models in economic evaluation. Pharmacoeconomics. 2011;29(1):51–62.

71. Haji Ali Afzali H, Gray J, Karnon J. Model Performance Evaluation (Validation and Calibration) in Model-based Studies of Therapeutic Interventions for Cardiovascular Diseases: A Review and Suggested Reporting Framework. Appl Health Econ Health Policy. 2013 Apr;11(2):85–93.

72. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology. 2010 Jan;21(1):128–38.

73. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics [Internet]. Boston, MA: Springer US; 2012 [cited 2016 Jul 17]. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-1-4614-1353-0

74. Clarke P, Gray A, Holman R. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). Med Decis Making. 2002;22(4):340–349.

75. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010 Dec 1;96(1):5–21.

76. Maddigan S, Feeny D, Majumdar, S, Farris, K, Johnson, J. Construct validity of the Health Utilities Index Mark 3 in Type 3 Diabetes: Evidence from a nationally representative sample of Canadians. Working Paper 05-02 [Internet]. 2005 Aug [cited 2016 Jan 1]. Available from: http://www.ihe.ca/download/construct_validity_of_the_health_utilities_index_mark_3_in_type_2_diabetes_evidence_from_a_nationally_representative_sample_of_canadians.pdf

77. Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med Care. 2004;1125–1131.

78. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med. 2005 Apr;60(7):1571–82.

79. McGovern A, Tippu Z, Hinton W, Munro N, Whyte M, de Lusignan S. Systematic review of adherence rates by medication class in type 2 diabetes: a study protocol. BMJ Open. 2016;6(2):e010469.

80. Karnon J. Model Validation: Has it's Time Come? PharmacoEconomics [Internet]. 2016 May 25 [cited 2016 Jul 16]; Available from: http://link.springer.com/10.1007/s40273-016-0415-y

81. Kothari V, Stevens RJ, Adler AI, Stratton IM, Manley SE, Neil HA, et al. UKPDS 60: Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine. Stroke. 2002 Jul 1;33(7):1776–81.

82. Khan NA, Wang H, Anand S, Jin Y, Campbell NRC, Pilote L, et al. Ethnicity and Sex Affect Diabetes Incidence and Outcomes. Diabetes Care. 2011 Jan 1;34(1):96–101.

83. Public Health Agency of Canada. Fast facts about diabetes 2011 data compiled from the 2011 survey on living with Chronic Diseases in Canada. [Internet]. [Ottawa]: Public Health Agency of Canada; 2011 [cited 2015 Nov 5]. Available from: http://publications.gc.ca/collections/collection_2012/aspc-phac/HP35-24-2011-eng.pdf

84. Tremblay M, Katzmarzyk P, Willms J. Temporal trends in overweight and obesity in Canada, 1981–1996. Int J Obes. 2002;26:538–43.

85. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague PA, et al. Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). The lancet. 2000;356(9226):279–284.

86. Chiu M, Austin PC, Manuel DG, Shah BR, Tu JV. Deriving Ethnic-Specific BMI Cutoff Points for Assessing Diabetes Risk. Diabetes Care. 2011 Aug;34(8):1741–8.

87. Braga MF, Casanova A, Teoh H, Dawson KG, Gerstein HC, Fitchett DH, et al. Treatment gaps in the management of cardiovascular risk factors in patients with type 2 diabetes in Canada. Can J Cardiol. 2010;26(6):297–302.

88. Harris SB, Ekoé J-M, Zdanowicz Y, Webster-Bogaert S. Glycemic control and morbidity in the Canadian primary care setting (results of the diabetes in Canada evaluation study). Diabetes Res Clin Pract. 2005 Oct;70(1):90–7.

89. Oster RT, Johnson JA, Hemmelgarn BR, King M, Balko SU, Svenson LW, et al. Recent epidemiologic trends of diabetes mellitus among status Aboriginal adults. Can Med Assoc J. 2011;183(12):E803–E808.

90. Jinyuan LIU, Wan T, Guanqin C, Yin LU, Changyong F, others. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry. 2016;28(2):115.