Projections of preventable risks for cardiovascular disease in Canada from 2001 to 2020: a microsimulation modelling approach

Douglas G. Manuel[1, 2, 3, 4]

Meltem Tuna[1, 2, 3]

Deirdre Hennessy[1, 2]

Carol Bennett[2, 3]

Anya Okhmatovskaia[5]

Philippe Finés[1]

Jack V. Tu[6]

William Flanagan[1]

for members of Simulation Technology for Applied Research (STAR) and the Canadian Cardiovascular Outcome Research Team (CCORT)

[1]Statistics Canada, Ottawa, Ontario, Canada

[2]Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

[3]Institute for Clinical Evaluative Sciences, Ottawa and Toronto, Ontario, Canada

[4]Department of Family Medicine and Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada

[5]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

[6]Department of Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada

Funding and acknowledgements:

The study was supported by funding from the Canadian Institutes for Health Research through the Canadian Cardiovascular Outcome Research Network (CCORT) and the New Emerging Team in Simulation Technology for Applied Research (STAR). The Population Health Model (POHEM) was developed by and receives support from Statistics Canada and partnering agencies. DGM holds a Chair in Applied Population Health from the Canadian Institutes for Health Research and the Public Health Agency of Canada. JVT is supported by a Canada Research Chair in Health Services Research and a Career Investigator Award from the Heart and Stroke Foundation of Ontario.

Abstract

Background: Improvements in preventable risks have contributed to a steady decrease of cardiovascular disease (CVD) incidence for the past 50 years in most developed countries. It is unclear whether the preventable risks and CVD incidence will continue to decrease in the future. Our objective was to inform future CVD trends by projecting preventable risk factors in Canada from 2001 to 2020.

Methods: The population-based microsimulation model used national data on births, deaths and migration; population socioeconomic position and CVD risk factors; and algorithms for CVD risk factor change, based on sociodemographic characteristics and previous CVD risks. The initial population—22.5 million individual "actors" representing Canada, 2001, age 20+ years old—had 18 characteristics including cardiovascular risk factors used for clinical risk prediction. There were 6.1 million potential exposure profiles for each actor each year. Outcome measures included: yearly risk factor prevalence (smoking, body mass index, diabetes, hypertension, and lipid levels), and prevalence of multiple risks.

Results: From 2001 to 2009, projected CVD risks closely approximated observed risks. Excepting obesity and diabetes, all CVD risk factors were projected to decrease through to 2020. The largest decreases projected were for smoking (from 25.7% in 2001 to 17.7% in 2020) and uncontrolled hypertension (from 13.3% to 6.0%). In 2017, obesity is projected to exceed smoking as the most prevalent risk factor. The prevalence of multiple risks is projected to remain low.

Interpretation: CVD risks are projected to improve modestly in Canada. The summary effect will likely be a continued decline in CVD incidence.

Word count: 250

Introduction

It is uncertain whether cardiovascular disease (CVD) incidence in Canada will increase, decrease or level out in the foreseeable future. This lack of insight is surprising given the steady and remarkable 70% decline in CVD incidence over the past 50 years as well as the impact of the disease, which accounts for over 70,000 Canadian deaths annually.[1]

In Canada, as in other Western countries, the uncertain future of CVD incidence is largely a consequence of the opposing trends of CVD preventable risks. Historically, CVD risk factors have steadily declined—most notably smoking, which has decreased in Canada by two-thirds since its peak. Similarly, uncontrolled hypertension has declined more than 70% since 1992.[2] Obesity, the exception to this trend, has steadily rising rates. It has been suggested that the increase in obesity will overshadow improvements in other preventable risks, resulting in a reversal of existing gains in life expectancy.[3]

We sought to project the prevalence of risk factors for CVD in Canada from 2001 to 2020. Because CVD is a product of multiple risk factors, we projected prevalence trends for both individual and multiple risks. We modelled social and demographic characteristics that are associated with CVD risks, including future change in the Canadian population structure. We also projected strata-specific risks by age, sex, socioeconomic position and region of residence.

Methods

Overview

The CVD risk factor prevalence projections (2001-2020) were generated using a Canadian microsimulation model (i.e., a simulation model that generates individual life histories): the Population Health Model for Cardiovascular Disease (POHEM: CVD). Projections were considered well-calibrated if the predicted estimates closely approximated observed estimates available from population health surveys (2001-2009). Following model

development and assessment, we projected future risk factors from 2010 to 2020. We examined common and important CVD risk factors, focusing on those used in a clinical setting for CVD risk stratification: hypertension, lipid levels, smoking, obesity and diabetes.

Study population

The study population represented all Canadian residents, aged 20 years and older, living in a community setting from 2001 to 2020.

Model characteristics

Model specification defined the purpose of the model, its overall structure and data sources, and was performed by the Simulation Technology for Advanced Research and the Canadian Cardiovascular Outcome Research teams. The main considerations for the CVD model included the following:

1) Population-based—reflecting the Canadian population, including important sub-populations such as age, sex, socioeconomic position and provincial regions.
2) Open population—allowing the population to change over time, reflecting births, deaths, immigration and emigration.
3) Relevant risk factors—inclusion of individual CVD risks that have an important population attribution and/or are used for risk stratification (estimation of future risk of CVD events). CVD risk should include categories or levels of exposure that are commonly used in risk stratification.
4) Multiple risk factors—ability to project the population prevalence of exposure to multiple risk factors.
5) Predictive accuracy—ability to generate accurate (well-calibrated) projections for the total Canadian population and also for specific sub-groups, such as by sex, age, and socioeconomic position.
6) Usefulness for population health planning—developed to estimate the health benefit of prevention scenarios, including the potential to estimate future CVD events.

Population Health Model (POHEM) framework

Statistics Canada's POHEM framework is an empirically-grounded, longitudinal microsimulation model of diseases and risk factors representing the lifecycle dynamics of

the Canadian population.[4] POHEM's basic unit of analysis is the individual person or "actor." Through dynamic simulation, POHEM recreates the Canadian population at a given point in time and ages it, one actor at a time, until death. We created a microsimulation model of CVD risk factors called POHEM: CVD. The life trajectory of each actor unfolds to create a life course of their CVD-related risks. Each actor's CVD risk is estimated yearly using individual, predictive risk equations that consider their characteristics in the previous year(s), including age, sex, other socio-demographic characteristics and health behaviours. These predictive equations are derived from a range of Canadian data sources (described within **Web Appendix 1)**.

Model development (POHEM:CVD)

Figure 1 shows the four steps that make up the process of microsimulation model development: initialization (creation of initial population), yearly updates, model validation and projection. Canadian population-based data sources were used for model initialization, yearly updates—including the generation of risk algorithms—and validation (see **Figure 2** and **Web Appendix 1**).

Initialization

POHEM: CVD was initialized using the 2000-2001 Canadian Community Health Survey (CCHS). Each CCHS respondent, aged 20+ (n=105,908), was replicated using the survey weight to generate a simulated population of approximately 22.5 million community-dwelling Canadians. See **Web Appendix 1** for details about the CCHS.

Each actor in the model had 18 characteristics. The following were used in modelling CVD risk: five demographic characteristics (age, sex, Canadian region of residence, ethnicity and immigrant status); two socioeconomic characteristics (income and education); one health behaviour (smoking); one intermediate risk (body mass index [BMI]); three proximal risks (blood pressure, lipid levels and diabetes). The CVD risk factors had levels of exposure that are typically used for clinical CVD risk prediction. Five major CVD risks (blood pressure, lipid levels, BMI, diabetes and smoking) had two to five levels (totaling 600 unique combinations of CVD exposure). Along with other characteristics, there were a total of 6.14

million potential actor states (i.e., exposure profiles). Variable definitions and levels of exposure are described in **Web Appendices 2 and 3**.

Data for CVD risks that were not available in the CCHS (measured systolic and diastolic blood pressure and lipid levels [total cholesterol and HDL]) were imputed from the 1990 Canadian Heart Health Survey (CHHS) using a validated "hot-deck" method based on age-group, sex, BMI category and diabetes status.[5-8] For details, see **Web Appendix 4**. Because medical treatment levels were low in 1990, the CCHS data allowed us to project the lifetime course of "native" blood pressure and lipid levels. Treated levels of blood pressure and lipids were then incorporated separately during model calibration to reflect uptake in treatment since 1990.

Yearly updates and risk transitions

Figure 2 shows five steps required to update the model each year from 2001 to 2020. The first four steps updated the model's population structure—age and total population size (births/deaths and immigration/emigration)—and socioeconomic characteristics (education and family income). See **Web Appendix 5** for details. The fifth step updated each actor's CVD risks by applying predictive algorithms for CVD risk change (described further below).

Separate predictive algorithms were developed for each CVD risk, except lipid levels and blood pressure, which shared a common algorithm. The outcome of each algorithm was a change from one category of health risk exposure to another. Examples of potential CVD risk change include: transition from a non-smoker to a current smoker, transition from current smoker to a former smoker, or transition from no diabetes to a person with diabetes. Each algorithm had different predictors that varied depending on CVD risk. All algorithms included age and sex, as well as up to seven additional predictors, depending on the risk factor modelled. See **Web Appendix 2** for a summary table of variables used to predict CVD risks.

All algorithms were generated using Canadian population data. Specifically, change in BMI (obesity) was modelled using the longitudinal National Population Health Survey (NPHS) from 1996/97 to 2004/06. Change in smoking status was modelled using the Canadian Health Survey (1979), the NPHS (1994) and the CCHS (2008), while change in blood pressure and lipid levels were modelled using the CHHS (1990). Diabetes risk was estimated using the Diabetes Population Risk Tool[9], which was developed using the 1996/7 NPHS and validated using the 2001 CCHS linked to a population-based diabetes registry. See **Web Appendices 1 and 4** for descriptions of the data sources and methods for updating CVD risks.

Validation and calibration

The predicted prevalence of the CVD risk factors obtained from POHEM: CVD were compared to Canadian observed estimates from the CCHS (2001-09) and the 2007 Canadian Health Measures Survey (CHMS). In total, five cycles of the CCHS collected between 2001 and 2009 allowed for validation of smoking, diabetes and obesity. The CHMS, which ascertained biophysical measures, was only collected once, allowing only one validation/calibration opportunity for blood pressure and lipid levels. Predicted births, deaths and immigration counts were compared to Statistics Canada's counts compiled from provincial vital statistics (see **Web Appendix 5** for details). Observed and projected estimates were compared for multiple risk factors for Canadians overall, as well as by age, sex, region and socioeconomic position.

The model was considered well-calibrated if there was <5% difference when the model was compared to observed estimates (CCHS or CHMS) or <10% difference for age 75+ due to small sampling size in the CCHS and CHMS. Our *a priori* calibration approach was to adjust model estimates for any CVD risk if over half of the strata-specific estimates for demographic or socioeconomic characteristics exceeded the calibration cut-off (5% or 10% differences). In this situation, the model estimates were adjusted by the corresponding strata difference between the model and observed estimates.

Projection

The model generated projections of CVD risks for each actor from 2001 to 2020. The projections were then aggregated by year for predefined groups including age, sex, Canadian region, and socioeconomic group. The projections reflect the "baseline" trend in preventive health care and societal factors that affect CVD risk.

Microsimulation model and statistical analyses

The POHEM: CVD risk model was generated using MODGEN, a microsimulation programming language developed and supported by Statistics Canada.[10] Statistical analysis of the NPHS, CCHS and CHHS were performed in SAS.

Results

Error! Reference source not found. describes the prevalence of CVD risk factors in the simulated Canadian population in 2001. Of the selected CVD risk factors (current smoking, obesity **[BMI ≥ 30]**, hypertension **[SBP ≥ 140 mm Hg]**, high cholesterol **[total cholesterol ≥ 6.22 mmol/L]** and diabetes), smoking was the most common (30.8% of men and 25.4% of women) with approximately twice the risk exposure as obesity (15.7 % of men and 14.4% of women), the next most common risk. For men, 12.0% had high cholesterol, 6.6% were hypertensive, and 5.0% had diabetes. The women's risk profile was similar (12.6% had high cholesterol, 6.7% were hypertensive, and 4.4% had diabetes).

The model was well-calibrated for both the initial 2001 population and projected estimates from 2001 to 2009 (see **Figures 3 and 4**). Hypertension and high cholesterol were exceptions, requiring calibration due to projected estimates that were approximately 20% (hypertension) and 7% (high cholesterol) higher than observed estimates (2007 CHMS).

Figure 3 shows projected risks from 2001 to 2020, along with the observed risk, from 2001 to 2009, for smoking, obesity, diabetes, high cholesterol (calibrated) and hypertension (calibrated). Obesity was projected to increase and, in 2017, overtake smoking as the most prevalent risk.

Figure 4 shows a projected, slight decrease in exposure to one of three risks: obesity, smoking and diabetes. The number of people with these three risk factors was low throughout the projected period (0.3% in 2001 declining to 0.2% in 2020).

Discussion

This study projects overall modest improvement regarding most CVD risks from 2009 through to 2020, excepting an increase in the proportion of Canadians with obesity and diabetes. Smoking, hypertension and high cholesterol have the largest projected decreases. The projected reduction in smoking is concordant with widely implemented smoking prevention strategies across Canada—for example, by 2007 most provinces prohibited smoking in public indoor spaces.[11] Additionally, controlled hypertension has increased four-fold in Canada (from 13% in 1992 to 65% in 2009)[2] and statin medications were the leading treatment prescribed in government-funded drug plans.[12]

The risk factor projections should be interpreted with care. We purposely used the term "baseline" to draw attention to estimates generated using data that was available at the "baseline" period (2001) and risk transition algorithms derived from other historic data. The actual future prevalence of risk factors beyond the validation period (2001-2009) will undoubtedly vary from our projection. Risk factor prevalence will be lower than our projections if healthy physical and social conditions are supported and further implemented, or if prevention of hypertension and dsylipidemia is increased. Conversely, Canada's physical and social environment may become conducive to an increase in obesity. The primary use of population predictive models, such as POHEM: CVD, is analogous to the use of predictive models at the clinical level—namely to inform decision-making through the examination of plausible future outcomes, including before and after implementation of new interventions or therapies.[13] In this way, the POHEM: CVD model can be used to examine the consequences of preventive strategies, such as increasing coverage of blood pressure therapy or implementing a new community-wide smoking policy, if there is known or postulated effectiveness of such strategies.

Future CVD incidence can be projected by combining POHEM: CVD with the same risk hazard estimates used in historic CVD models such the Wijeysundera et al. study.[14] Alternatively, we could incorporate a CVD incidence algorithm such as the Framingham Risk Tool.[15] Had we performed either analysis, we would expect CVD incidence—adjusted for Canada's larger, aging population—to decrease through to 2020. POHEM: CVD projects a marked decrease in smoking, uncontrolled hypertension and dyslipidemia, each of which has a two- to three-fold risk of CVD. While diabetes has a similar hazard and is projected to increase, influencing an upward trend in CVD incidence, the projected increase of diabetes is considerably smaller than the projected decrease in smoking, or other major CVD risks. Obesity will have a notable prevalence increase, but has a comparatively smaller CVD risk.[16]

Our study has two strengths. First, POHEM: CVD uses predictive algorithms for CVD risk changes that are developed in the same way as clinical algorithms, except that they are generated using population-based data. A microsimulation approach more readily allows for risk factor projection while considering multiple predictors and open populations (with births, deaths and migration). It also allows projections for multiple risks and varied strata (such as age, sex and socioeconomic position).

Second, we validated our projections against external data. This validation approach has become common for clinical risk prediction, but is not widely practiced in population settings. The projected CVD risk estimates compared closely to the observed estimates (2001 to 2009) except for blood pressure and cholesterol, where uncalibrated projections exceeded observed estimates. It was expected blood pressure and lipid levels would require calibration due to the strong trend of improved treatment in Canada.[17]

Accurate or well-calibrated projection of BMI, smoking and diabetes status during the validation time period (2001 to 2009) was possible because these CVD risks appear to be largely predictable over the short- or medium-term based on sociodemographic characteristics and related health behaviours.[18-20] For example, a person's present BMI

measurement is highly related to that of the previous year, after adjusting for the life course of BMI that generally results in a weight increase up to middle-age, followed by a slow decrease as a person ages.[18] The POHEM: CVD model captures these life-course patterns.

The study has several limitations, including the inherent challenge of projecting CVD risks that are influenced by constantly changing social environments and health care systems. Despite including many CVD risks, there were notable omissions in POHEM: CVD, including: physical activity, sedentary activity, diet, stress, and biophysical markers (such as Apolipoprotein B and C-reactive protein). However, compared to other studies, we included a larger range of risks and exposure levels. Of particular note, we considered CVD risks that are important for predicting CVD incidence. It has been debated whether additional risk factors would improve CVD incidence risk prediction/characterization, but generally calibration shows no or very minor incremental improvement with the addition of risks beyond those included in POHEM: CVD.[21]

In conclusion, our results suggest that, excepting obesity and diabetes, cardiovascular risk factors will continue to decline in Canada. The prevalence of people with multiple cardiovascular risk factors will remain low. Projected CVD risks can inform decision-making with the goal of continued gains in CVD risk and disease incidence. Although our study did not project future CVD incidence, there will likely be continued decline resulting from the combined effect of notable decreases in the prevalence of smoking and other risks, as well as the considerable, individual CVD risk associated with these risk factors. The projected increase in obesity and diabetes will likely attenuate but not reverse improvements in CVD incidence.

Reference List

(1) Public Health Agency of Canada. 2009 Tracking Heart Disease and Stroke in Canada. http://www.phac-aspc.gc.ca.

(2) McAlister FA, Wilkins K, Joffres M et al. Changes in the rates of awareness, treatment and control of hypertension in Canada over the past two decades. *CMAJ* 2011;183:1007-1013.

(3) Olshansky SJ, Passaro DJ, Hershow RC et al. A potential decline in life expectancy in the United States in the 21st century. *N Engl J Med* 2005;352:1138-1145.

(4) Wolfson MC. POHEM - a framework for understanding and modelling the health of human populations. *World Health Stat Q* 1994;47:157-76.

(5) Durrant GB. Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review. NCRM/002. 2005. ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute, University of Southampton. NCRM Methods Review Papers.

(6) Durrant GB, Skinner C. Using data augmentation to correct for non-ignorable non-response when surrogate data are available: an application to the distribution of hourly pay. *Journal of the Royal Statistical Society Series A, (Statistics in Society)* 2006;169:605-623.

(7) Durrant GB, Skinner C. Using Missing Data Methods to Correct for Measurement Error in a Distribution Function. *Surv Methodol* 2006;32:25-36.

(8) Hennessy D, Bennett C, Tuna M, Nadeau C, Flanagan W, Manuel D. An assessment of methods to impute risk exposure into model actor's risk profile for microsimulation. *International Methodology Symposium* 2011;Conference Proceedings.

(9) Rosella LC, Manuel DG, Burchill C, Stukel TA. A population-based risk algorithm for the development of diabetes: development and validation of the Diabetes Population Risk Tool (DPoRT). *J Epidemiol Community Health* 2010.

(10) Statistics Canada. Modgen (Model generator). *http://www statcan gc ca/microsimulation* [serial online] 2011; Accessed May 29, 2012.

(11) Health Canada. Public Smoking Bans by Province and Territory. *http://www hc-sc gc ca/hc-ps/tobac-tabac/res/news-nouvelles/ban-interdiction-public-eng php* [serial online] 2007; Accessed April 18, 2012.

(12) Canadian Institute for Health Information. Drug Use Among Seniors on Public Drug Programs in Canada, 2002 to 2008. 2010. Ottawa ON, CIHI.

(13) Barlow J, Bayer S. Raising the profile of simulation and modelling in health services planning and implementation. *Journal of Health Services Research Policy* 2011;16:129-130.

(14) Wijeysundera HC, Machado M, Farahati F et al. Association of Temporal Trends in Risk Factors and Treatment Uptake With Coronary Heart Disease Mortality, 1994-2005. *JAMA: The Journal of the American Medical Association* 2010;303:1841-1847.

(15) Sheridan S, Pignone M, Mulrow C. Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians. *J Gen Intern Med* 2003;18:1039-1052.

(16) Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. *Arch Intern Med* 2002;162:1867.

(17) McAlister FA, Wilkins K, Joffres M et al. Changes in the rates of awareness, treatment and control of hypertension in Canada over the past two decades. *Can Med Assoc J* 2011;183:1007-1013.

(18) Orpana H, Tremblay M, Fines P. Trends in weight change among Canadian adults. *Health Rep* 2007;18.

(19) Shields M. Smoking bans: Influence on smoking prevalence. *Health Rep* 2007;18.

(20) Ross N, Gilmour H, Dasgupta K. 14-year diabetes incidence: The role of socio-economic status. *Health Rep* 2010;21.

(21) Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol* 1992;45:85-89.

(22) Beland Y. Canadian Community Health Survey - Methodological Overview. *Health Rep* 2002;13:9-14.

(23) MacLean DR, Petrasovits A, Nargundkar M et al. Canadian heart health surveys: a profile of cardiovascular risk. Survey methods and data analysis. Canadian Heart Health Surveys Research Group. *Can Med Assoc J* 1992;146:1969-1974.

(24) Swain L, Catlin G, Beaudet MP. The National Population Health Survey--its longitudinal nature. *Health Rep* 1999;10:69-82.

(25) Catlin G, Will P. The National Population Health Survey: highlights of initial developments. *Health Rep* 1992;4:313-319.

(26) Wilkins K, Campbell N, Joffres M et al. Blood pressure in Canadian adults. *Health Rep* 2010;21.

(27) Statistics Canada. The LifePaths Microsimulation Model: An overview. *http://www statcan gc ca/microsimulation/* [serial online] 2009; Available from: Statistics Canada. Accessed May 29, 2012.

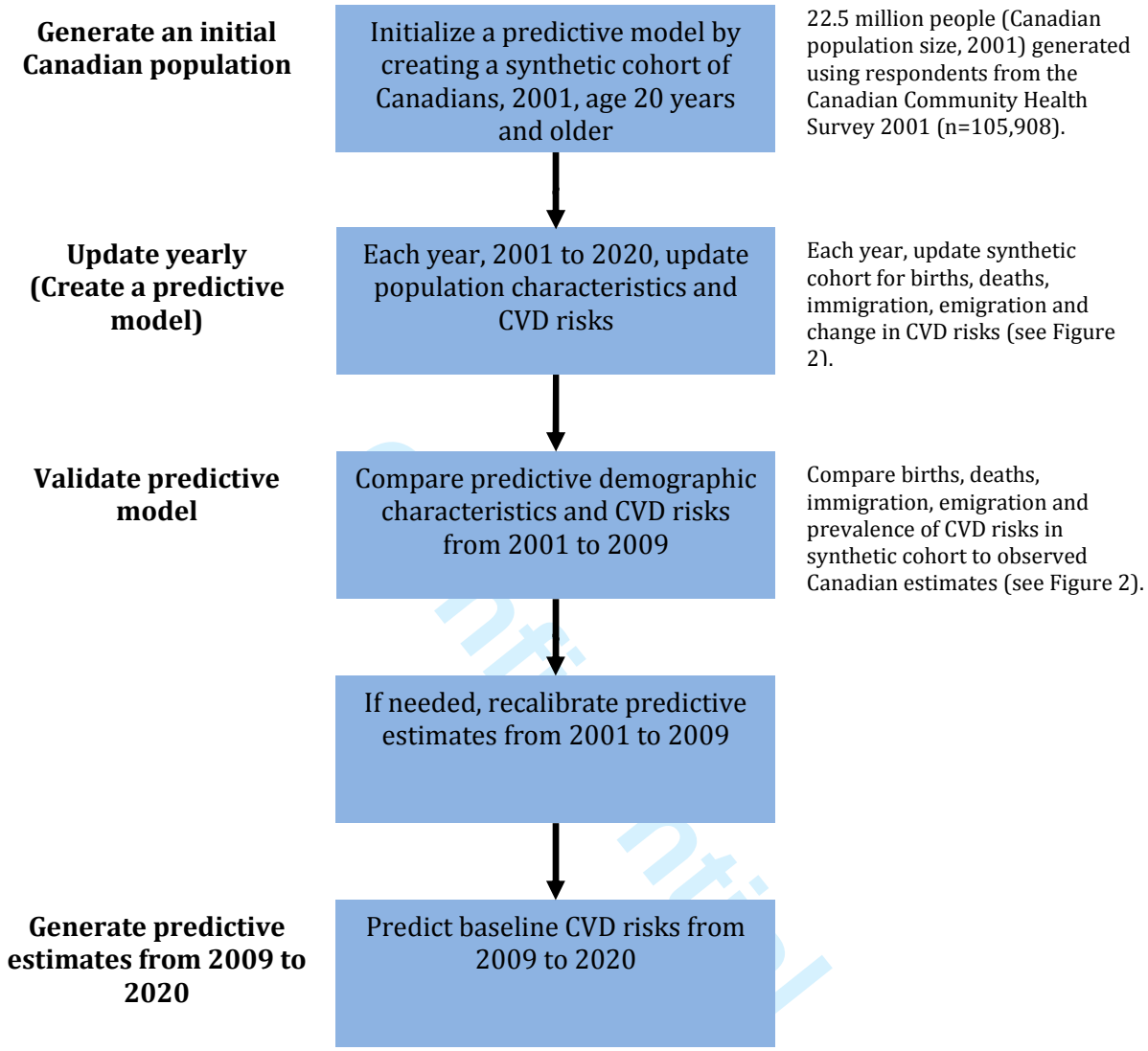

Figure 1 Process for predicting baseline cardiovascular disease (CVD) risks in Canada, 2001 to 2020.

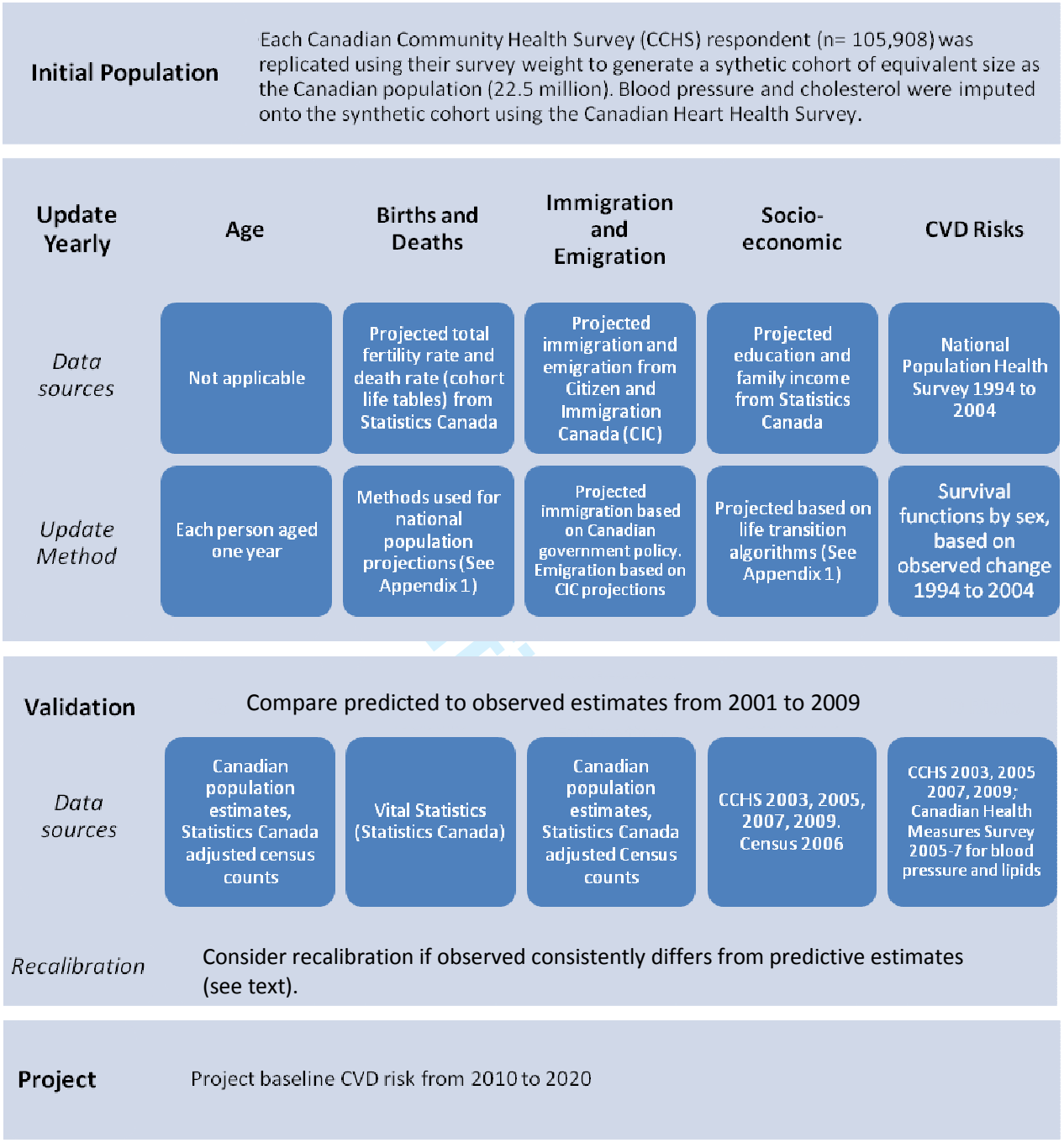

Figure 2 Data sources and methods to generate and validate cardiovascular disease (CVD) risk projections, Canada, 2001 to 2020.

Table 1. Cardiovascular disease risk factor prevalence, Canada 2001

	Males (n=11,000,000)*	Females (n=11,500,000)*
Age		
20-44	52.9%	50.2%
45-64	32.7%	32.0%
65-74	9.1%	10.0%
75+	5.3%	7.8%
Smoking Status		
Current	30.8%	25.4%
Former (quit ≥1 year)	42.4%	35.6%
Never	26.8%	39.1%
Body Mass Index (kg/m²)		
Underweight (<18.5)	1.3%	4.5%
Normal (18.5-24.9)	43.0%	54.0%
Overweight (25-29.9)	40.0%	27.1%
Obese (≥30)	15.7%	14.4%
Systolic Blood Pressure (mm Hg)		
Optimal (< 120)	60.5%	72.0%
Normal (120-129)	22.8%	13.4%
High-normal (130-139)	11.9%	7.9%
Hypertensive Stage I (140-159)	5.9%	4.8%
Hypertensive Stage II-IV (≥160)	0.7%	1.9%
Total Cholesterol (mmol/L)		
Low (<4.15)	20.2%	18.2%
Low-medium (4.15-5.17)	34.8%	38.8%
Medium (5.18-6.21)	33.0%	30.4%
Medium-high (6.22-7.24)	10.0%	9.8%
High (≥7.25)	2.0%	2.8%
Diabetes		
No	95.0%	95.6%
Yes	5.0%	4.4%

*Number rounded to three significant digits.

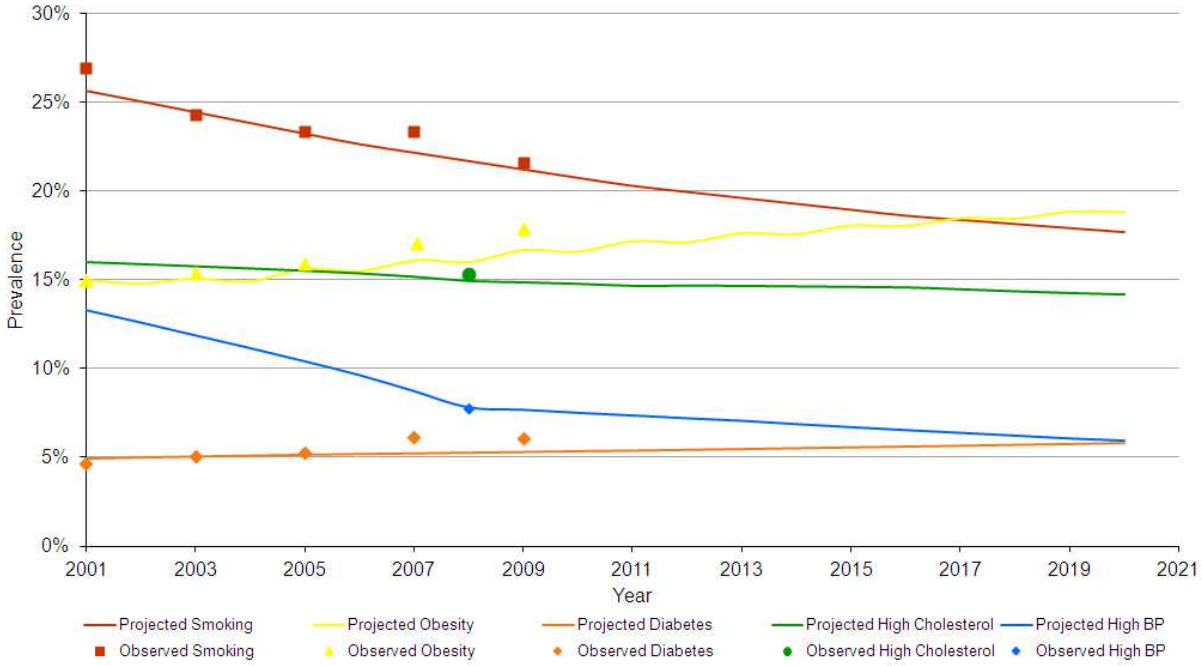

Figure 3. Projected (2001 to 2020) and observed (2001 to 2009) prevalence of cardiovascular disease risks, Canada. Obesity [BMI ≥ 30], hypertension [SBP ≥ 140 mm Hg], high cholesterol [total cholesterol ≥ 6.22 mmol/L].

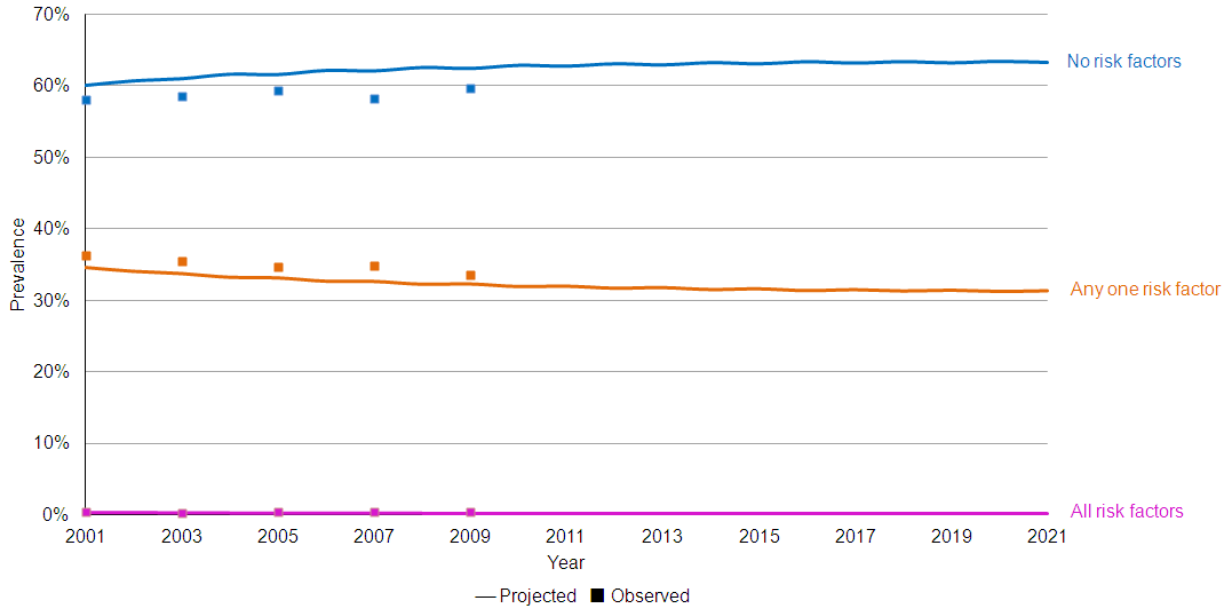

Figure 4. Projected (2001 to 2020) and observed (2001 to 2009) prevalence of combinations of cardiovascular disease risks (smoking, obesity, and diabetes), Canada.

Appendix 1

Data sources

Statistics Canada Census projections (CANSIM)—Information related to the population structure over time was obtained from CANSIM, a tool developed by the Demography Division at Statistics Canada. This tool is used to project future numbers of new births and immigrants. In addition, death rate (number of deaths) was projected by age, sex and year of birth using Vital Statistics data available at Statistics Canada. In POHEM: CVD, CANSIM and Vital Statistics data were used to initialize and update the population structure over time.

Canadian Community Health Survey (CCHS)—The CCHS is a cross-sectional survey started in 2000/01 with a sample size of 131,535. It was initially repeated every two years but as of 2009, data is collected on an ongoing basis to provide annual estimates. The CCHS was designed to be representative of the Canadian household population aged 12 years and older and elicited a wide range of self-reported information related to health status, health care utilization and health determinants. The CCHS survey and sampling strategies have been described in detail elsewhere.[22] The CCHS 2000/01 survey was used to define the initial subject population of the POHEM: CVD model that was projected forward in time by the simulation.

Canadian Heart Health Survey (CHHS)—The CHHS are cross-sectional surveys conducted separately in each of the ten provinces between 1986 and 1992. These surveys were designed to gather information on risk factors associated with cardiovascular disease for individuals aged 18-74 years and had a combined sample size of 23,129. Unlike the CCHS, the CHHS also collected physical measures such as blood pressure and cholesterol. The CHHS survey and sampling strategies have been described in detail elsewhere.[23] In POHEM: CVD the CHHS dataset was used to initialize blood pressure and cholesterol variables and to impute values of blood pressure and cholesterol into the CCHS to allow the risk transitions for these risk factors to be updated yearly.

National Population Health Survey (NPHS)—The NPHS is a longitudinal survey started in 1994/95 with a sample of 17,276 individuals aged 12 years and older. The survey is conducted every two years and currently has 18 years of follow-up. Like the CCHS, the NPHS elicited a wide range of self-reported information related to health status, health care utilization and health determinants. The NPHS survey and sampling strategies have been described in detail elsewhere.[24;25] In POHEM: CVD, the NPHS data set was used to model risk factors related to health behaviour and diabetes prevalence.

Canadian Health Measures Survey—The CHMS is a cross-sectional survey started in 2007. In cycle 1 of the CHMS, data were collected at 15 sites across Canada from March 2007 through February 2009.[26] The survey covered the population aged 6 to 79 years living in private households. It was designed to provide sex-specific, statistically reliable, national estimates of conditions for which prevalence was at least 10% for five age-groups: 6 to 11, 12 to 19, 20 to 39, 40 to 59, and 60 to 79 years.[26] The CHMS does not include residents of Indian Reserves or Crown lands, institutions and certain remote regions, nor full-time members of the regular Canadian Forces. Of the households selected for inclusion in the CHMS, the response rate was 69.6%—meaning that in 69.6% of selected households, the sex and date of birth of all household members were provided by a household resident. In each responding household, one or two members were selected to participate in the survey; for the age group 20 to 79 years, 87.9% of selected household members completed the household questionnaire, and 83.6% of the responding household members participated in the subsequent examination component of the survey. The household questionnaire collected information about a wide range of self-reported information related to health status, chronic disease status and health behaviours. The examination portion of the survey collected physical measures of health including blood and urine samples, as well as measures of blood pressure, lipid levels, physical fitness and BMI.[26]

Appendix 2: Covariates that exist in POHEM: CVD model

CVD risk factors (# categories)	Data sources	POHEM Covariates (# categories)																
		Demographics					SES		Chronic disease profile				Biophysical measures		HS	Health behaviours		
	Initial data source	5 years age groups (16)	Sex (2)	Region (5)	Ethnicity (2)	Immigrant (2)	Income (4)	Education (4)	Diabetes (2)	Heart disease (2)	Arthritis (2)	Osteoarthritis (2)	Blood pressure (5)	Total cholesterol and HDL (5)	HUI (continuous)	BMI/previous BMI (4)	Smoking (3)	Alcohol consumption (4)	Nutrition (2)
Blood pressure*(5)	Imputed using CHHS (1990)	√	√						√					√		√			
Total cholesterol and HDL* (5)	Imputed using CHHS (1990)	√	√						√				√			√			
Obesity (4)	NPHS (1996/97 - 2004/05)	√	√	√			√	√								√†			
Diabetes (2)	NPHS (1996-97)	√†	√‡		√	√		√		√			√			√†	√		
Smoking (3)	CHS (1979), NPHS (1994) and CCHS (2008)	√	√‡	√															

Notes: SES= socioeconomic status, HS= health status, HUI= health utilities index, CHHS= Canadian Heart Health Study, NPHS=Nation Population Health Survey, CHS= Canadian Health Study, CHHS= Canadian Community Health Survey, HDL= high density lipoprotein, * indicates coevolving risk factors, † indicates characteristics that are strongly predictive of the outcome (CVD risk factor), ‡ indicates that models were constructed separately for males and females. Columns shaded grey are not involved in modelling the CVD risk factors.

Appendix 3: Definition of categorical variables used in modelling

POHEM: CVD

Variables	Category	Definition		
Total and HDL cholesterol (mmol/L)		**Total**		**HDL**
	Low	<4.15	or	<0.90
	Low-medium	4.15-5.17	or	0.90-1.16
	Medium	5.18-6.21	or	1.17-1.29
	Medium-high	6.22-7.24	or	1.30-1.54
	High	>7.25	or	>1.55
Blood pressure (mm Hg)		**Systolic**		**Diastolic**
	Optimal	<120	and	<80
	Normal	120-130	or	80-85
	High-normal	130-140	or	86-90
	Hypertensive stage I	140-160	or	90-100
	Hypertensive stage II-IV	>160	or	>100
BMI (kg/m^2)	Underweight	<18.5		
	Normal	18.5-25		
	Overweight	25.1-30		
	Obese	>30		
Diabetes	Yes	Physician diagnosed diabetes		
	No			
Smoking status	Light smoker	Less than 20 cigarettes a day		
	Heavy smoker	At least 20 cigarettes a day		
	Non-smoker	Never/former smoker		
Age	Multiple age groups	Defined as needed		
Sex	Male			
	Female			
Region	Atlantic	New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador		
	Quebec	Quebec		
	Ontario	Ontario		
	Prairies	Manitoba, Saskatchewan, Alberta		
	British Columbia	British Columbia		
Ethnicity	White	Self-reported ethnicity		
	Non-white			
Immigration	Non-immigrant	Self-reported immigrant status		
	Immigrant			
Income	Low	Self-reported income		
	Medium-low			
	Medium			
	High			
Education	Low	Self-reported education		
	Medium-low			
	Medium			
	High			
Heart disease	Yes	Self-reported heart disease		
	No			

Notes: CVD= Cardiovascular disease, HDL= high density lipoprotein, mg/dl= milligrams per decilitre, mm Hg= millimetres of mercury, BMI= body mass index, kg/m^2 = kilograms per metre squared.

Appendix 4

Updating CVD risks

CVD risks were updated yearly through prediction algorithms that were generated using the longitudinal NPHS, the CCHS and the CHHS. The CVD risk algorithms were applied separately to each actor on a yearly basis for each CVD risk. As actors were added to the model over time (through births or deaths), they also had CVD risk algorithms applied each year. Similarly, actors who died or emigrated were no longer part of the model population and were no longer updated. The predicted change for the following year for each CVD risk was estimated using the predictor variables contained in the algorithms that were specific to the particular CVD risk. Each CVD risk algorithm is described in detail below. Risk algorithm development used different statistical approaches depending on the specific risk and data source; however, there was a common approach. The approach included: assessment of predictive risks based on established relationships and empiric findings from Canadian national development data; and the balancing of discrimination and calibration and/or risk factor distribution.

APPENDIX 4.1: DIABETES	
Name	DPoRT (Diabetes Population Risk Tool)
Objective/Purpose	To create and validate a population-based risk prediction tool for incident diabetes using commonly collected national survey data.
Process of Development	
- Data source/development data	1996/7 **National Population Health Survey (NPHS)** for 23,403 Ontario residents collected by Statistics Canada. Survey data was linked to administrative data (**Ontario Diabetes Database**) to ascertain physician diagnosed diabetes status.
- Outcome definition	Outcome was physician diagnosed diabetes defined in administrative data as a hospital admission with a diabetes diagnosis code 250 (ICD9-CM) before 2002 or ICD-10 code E10-E14 after 2002; OR a physician services claim with a diabetes diagnosis followed within 2 years by either a physician services claim or a hospital admission with a diabetes diagnosis. Cases of gestational diabetes were excluded.
- Predictive/stratifying/causal variables	The variables used to model diabetes incidence were based on established evidence, easily captured using population surveys and in a consistent manner across surveys and populations and included age, height, weight, chronic conditions diagnosed by a health professional, ethnicity, immigration status, smoking status, educational achievement, household income, alcohol consumption and physical activity. Important predictor variables for men were hypertension, non-white ethnicity, heart disease, current smoking status, education and age/BMI category. The general form of the male model is outlined below: **Log(Diabetes incidence time)= $\alpha + ß_1HTN + ß_2NW + ß_3HD + ß_4CS + ß_5ED + ß_6age/BMI + \sigma\varepsilon$** **with ε following the extreme value distribution** Important predictor variables for women were hypertension, non-white ethnicity, immigration, education and age/BMI category. The general form of the female model is outlined below: **Log(Diabetes incidence time) = $\alpha + ß_1HTN + ß_2NW + ß_3IMM + ß_4ED + ß_5age/BMI + \sigma\varepsilon$**
- Statistical Method	Weibull accelerated failure time model allows user to predict diabetes probability for range of follow-up periods. Diabetes functions were derived separately for men and women.
Performance Characteristics and Assessment Process	
- Validation/external validation	The model was assessed for discrimination and calibration. Discrimination was measured using a C-statistic modified for survival data. C- statistics ranged between 0.77 and 0.79 in the development and validation datasets, indicating good discrimination. Accuracy or calibration was measured using a Hosmer-Lemeshow chi-squared statistic modified for survival data.

	Calibrated models in validation cohorts fell below the acceptable cut off of 20, indicating good calibration. The model was externally validated on 2 external datasets from Ontario and Manitoba and performed well both in terms of discrimination and calibration.
- Estimates produced/valid for which subpopulations?	Estimates of diabetes incidence were produced for age, BMI, ethnicity and educational level subgroups
Plain Language Summary	DPoRT (Diabetes Population Risk Tool) is a means to determine diabetes risk estimates, using national survey data to calculate the number of Canadians at risk of developing diabetes and to determine how this disease risk is distributed among the population. The purpose of the study was not only to provide risk estimates but also to inform health policy. Accuracy and discrimination of the model was described by comparing observed diabetes rates with predicted estimates. Results of the study provided predictive risk factors, including BMI, age, ethnicity and smoking. Through the two external cohorts, DPoRT showed good discrimination and calibration. Models such as DPoRT can inform healthcare planning and disease prevention strategies.
References	L. Rosella, D. Manuel, C. Burchill, T. Stukel and the PHIAT-DM team. A population-based risk algorithm for the development of diabetes: development and validation of the Diabetes Population Risk Tool (DPoRT). J Epidemiol Community Health. June 2010.

<table>
<tr><td colspan="2">APPENDIX 4.2: CHOLESTEROL AND BLOOD PRESSURE</td></tr>
<tr><td>Name</td><td>Cholesterol, high density lipoprotein (HDL) and hypertension (POHEM:CVD)</td></tr>
<tr><td>Objective/Purpose</td><td>To derive the joint probability of changing cholesterol and blood pressure states from one age group to the next.</td></tr>
<tr><td>Process of Development</td><td></td></tr>
<tr><td>- Data source/development data</td><td>1986 to 1992 Canadian Heart Health Surveys (CHHS)</td></tr>
<tr><td>- Outcome definition</td><td>Cholesterol and HDL were categorized in 5 groups, tabulated below for simplicity:
<table>
<tr><th>Cholesterol mmol/L</th><th>Total Cholesterol</th><th>HDL</th></tr>
<tr><td>Low</td><td><4.15</td><td><0.90</td></tr>
<tr><td>Low-medium</td><td>4.15-5.17</td><td>0.90-1.16</td></tr>
<tr><td>Medium</td><td>5.18-6.21</td><td>1.17-1.29</td></tr>
<tr><td>Medium-high</td><td>6.22-7.24</td><td>1.30-1.54</td></tr>
<tr><td>High</td><td>7.25+</td><td>1.55+</td></tr>
</table>
Blood pressure (systolic and diastolic) was also categorized into 5 groups. The definition for each group, except the optimal group, depends on reaching a cut off value for either systolic OR diastolic blood pressure tabulated below for simplicity:
<table>
<tr><th>Blood pressure (mm Hg)</th><th>Systolic</th><th>Diastolic</th></tr>
<tr><td>Optimal</td><td><120</td><td>AND <80</td></tr>
<tr><td>Normal</td><td>120-130</td><td>OR 80-85</td></tr>
<tr><td>High normal</td><td>130-140</td><td>OR 85-90</td></tr>
<tr><td>Hypertensive stage I</td><td>140-160</td><td>OR 90-100</td></tr>
<tr><td>Hypertensive stage II-IV</td><td>≥160</td><td>OR ≥100</td></tr>
</table></td></tr>
<tr><td>- Predictive/stratifying/causal variables</td><td>5 year age group, sex, BMI, diabetic status</td></tr>
<tr><td>- Statistical methods for imputing cholesterol, HDL and blood pressure from CHHS to the POHEM startup population (CCHS 2001)</td><td>Specifically, using variables common to the CCHS and CHHS, individual's blood pressure and cholesterol categories were imputed using "hot-deck" methods. In other words individuals, in the CCHS were matched to those in the CHHS based on 5-year age-group, sex, BMI category and diabetes status and were assigned the corresponding categories total cholesterol/blood pressure available in the CHHS. HDL was subsequently imputed based on the total cholesterol level also using CHHS records, for persons having the same 5 year age-group, sex, total cholesterol, BMI and diabetes status.</td></tr>
<tr><td>- Statistical method for determining the transitional</td><td>Joint distributions of total cholesterol, blood pressure, BMI, and diabetic status, as measured on the CHHS, were</td></tr>
</table>

probabilities from one cholesterol/blood pressure group to another to another	used to derive the joint probability of changing cholesterol and blood pressure states from one age group to the next. Transport flow methodology, with the SAS NETFLOW procedure, was used to estimate these transition probabilities from one state to the next. Essentially this method generates a probability matrix that allows for change in individuals actors' states (of total cholesterol, HDL and blood pressure) while minimizing the flow needed to achieve the actual distribution. Diabetes and BMI were included as covariates to smooth the transitions of cholesterol and blood pressure.
Performance Characteristics and Assessment Process	
- Validation/external validation	During development and subsequently the accuracy of this prediction method was assessed against observed population-based survey data, however as this was not a classical prediction model so no formal statistical assessment of calibration was made. This study will be the first published external validation of the cholesterol, HDL and hypertension module.
- Estimates produced/valid for which subpopulations?	Not reported
Plain Language Summary	Using the CHHS blood pressure, total cholesterol and HDL were imputed into the POHEM start-up dataset (CCHS 1.1). In addition, transition probability matrices, that allowed co-evolution of blood pressure, total cholesterol and HDL were calculated from the CHHS data and implemented into POHEM to allow projection of these physiological measures over time.
References	None published

APPENDIX 4.3: SMOKING	
Name	Smoking (Canadian Cancer Risk Management Model (CRMM)) integrated into POHEM:CVD
Objective/Purpose	To estimate the distribution of smokers in the Canadian population and to generate a probability of moving from one smoking group to another, given a simulated individual's 5-year age group, sex, province and former smoking category.
Process of Development	
- Data source/development data	1979 **Canadian Health Survey** (CHS), 1994 **National Population Health Survey** (NPHS) and 2008 **Canadian Community Health Survey** (CCHS)
- Outcome definition	Smokers were categorized into 3 exposure groups, non-smoker, light smoker (less than 20 cigarettes per day) and heavy smoker (20 or more cigarettes a day). These categories were derived from variables in the CHS which had 4 categories: heavy smoker, light smoker, former smoker, never smoked; and from variables in the CCHS and NPHS which had 4 categories: never smoked, daily smoker, occasional smoker and former smoker. For daily smoker, if the number of cigarettes smoked per day is >=20 then the individual is categorized as heavy, else light smoker. For occasional smokers, if the number of cigarettes per months is >=600 then the individual is categorized as heavy, else light smoker. For never smoked or former smoker the individual is categorized as non-smoker.
- Predictive/stratifying/causal variables	Age, sex and region (province)
- Statistical method for determining the distribution of smokers	Starting with data of persons aged 15-19 years in 1979 from CHS (and 30-34 in 1994 from NPHS), the percentage of heavy smokers, light smokers and non smokers (pH, pL, pN) in 1979 and in 1994 was computed. For any year (y) between 1979 and 1994 (including these bounds), the following **linear interpolation** was applied: pH(y)=pH(1979) + (pH(1994)-pH(1979))/15*(y-1979), pL(y)=pL(1979) + (pL(1994)-pL(1979))/15*(y-1979), pN(y)=pN(1979) + (pN(1994)-pN(1979))/15*(y-1979) [because 15 is the number of years between 1979 and 1994]. The same process was repeated for each province, each age group, sex and for the time period between 1994 (NPHS data) and 2008 (CCHS data).
- Statistical method for determining the transitional probabilities from one smoking	For the time period 1979-1994, it was assumed that the persons could change their exposure group only during the last year of a 5-year period, i.e. in 1984, 1989 and

group to another

1994. Therefore, 3 transition periods were considered: 1979-1984, 1984-1989, and 1989-1994. For each of these periods, a transition matrix was built, where the row marginals are the prevalence rates for the beginning year, and the columns marginals are the prevalence observed for the end year (for the age group 5 years older), see example and tables below.
The transition matrix was filled in using the North-West method, where the values of the cells are determined in such a way that (i) the marginals are satisfied, (ii) a constraint is to minimize the proportion of persons changing their exposure group, thus reproducing the likely tendency to reduce the probability of changing one's exposure group, (as a consequence, most values will lie in the diagonal) (iii) filling in starts at the cell at the North-West corner of the table. As an additional constraint, we "forbid" any person to transit directly from Non-smoker to Heavy smoker and vice versa. Once the values in the transition matrix have been determined, they are transformed into probabilities of falling in a given group, given their (initial) group.
For example, if, for a given sex, province, age group the row marginals are 0.1, 0.6 and 0.3 (representing the proportion of heavy, light and non-smokers) and the columns marginals (for the age group 5 years older 5 years later) are: 0.2, 0.7, 0.1, then the transitions satisfying the method are indicated in the "Change required" table below. Then these values are transformed into probabilities of falling in a given final group, in the following way (see "Probability matrix" table). For persons who were Heavy smokers at the beginning of the period, the probability of remaining smoker is 1 (100%); for the persons who were Light smokers at the beginning of the period, the probability is .5/.6 of remaining light and .1/.6 of becoming a heavy smoker; for the ones who were non-smokers initially, the probability of becoming light and remaining non-smoker is respectively .2/.3, .1/.3.
Change required Probability matrix

	To	**H**	**L**	**N**			**H**	**L**	**N**
From		**0.2**	**0.7**	**0.1**			**0.2**	**0.7**	**0.1**
H	**0.1**	0.1				**0.1**	1		
L	**0.6**	0.1	0.5			**0.6**	1/6	5/6	
N	**0.3**		0.2	0.1		**0.3**		2/3	1/3

For the period 1994-2008, 3 transitions were also considered: 1994-1999, 1999-2003, 2003-2008 (that is, periods of respectively 4 and 5 years). During this period more data were available and were used in this

	paper to externally validate the linear interpolation. Adjustments of the method were made for the cases where the age group was the youngest or the oldest; also, to extend the projection before 1979 and after 2008, some assumptions were made. In particular, for projection after 2008, it was assumed that the marginals for a given age group followed the trend already observed in period 2003-2008.
Performance Characteristics and Assessment Process	
- Validation/external validation	During development and subsequently the accuracy of this prediction method was assessed against observed population-based survey data, in order to ensure that the distribution of the smoking groups was maintained over time and between data sources. This study will be the first published external validation of the smoking module.
- Estimates produced/valid for which subpopulations?	Estimates of smoking prevalence were produced for 5 year age groups, sex and province.
- Assumptions	For the CHS 1979, only 5 regions were available (instead of provinces): Atlantic, Québec, Ontario, Prairies and Vancouver. In designing the linear interpolation it was assumed that all 4 Atlantic provinces shared the same prevalence and that all Prairie provinces shared the same prevalence and that BC was similar to Vancouver.
Plain Language Summary	The smoking module of the CRMM provided a means to estimate the prevalence of smoking exposure by different age groups, sex and province, using observed population-based data. These prevalences provided the data necessary to calculate the probability of transitioning from one smoking group to another, given 5-year age group, sex, province and former smoking category.
References	None published

APPENDIX 4.4: OBESITY	
Name	Body weight (POHEM:CVD)
Objective/Purpose	To estimate equations to model change in weight over time among Canadian adults.
Process of Development	
- Data source/development data	**Longitudinal National Population Health Survey (NPHS)**
- Outcome definition	The outcome was change in BMI over time. Weight change was assessed by calculating an individual's body mass index (BMI) from self-reported weight and height, at each of five consecutive cycles of the NPHS, conducted in two year intervals from 1996–97 to 2004–05. A Box-Cox transformation was applied to normalize the distribution of the outcome variable.
- Predictive/stratifying/causal variables	Age, sex, income quartile, education, region of residence and previous values of BMI.
- Statistical Method	Linear regression models, stratified by age, sex and previous BMI category, were used to estimate change in individual's self-reported BMI using NPHS data from 1996–97 to 2004–05. In addition, because previous BMI may be limited to 1, 2, 3 or 4 occasions in the course of the simulation, 4 versions of the regressions were constructed. Briefly, each of the 4 sets of regression were subdivided into 28 groups according to age, group, sex and previous BMI category, totalling 112 separate regressions. The general form of the regression equation is outlined below: $\mathbf{(30+\Delta BMI)^{\lambda} = \alpha + ß_1 BMI + ß_2 I + ß_3 E + ß_4 R + \sigma\varepsilon}$ Each of these models predicts a 2 year change in BMI. In this analysis outliers were excluded based on the Student's residuals and cases with missing values for the important variables were removed from the analysis. Because income was frequently missing, imputation was used rather than letting the affected observations go to waste.
Performance Characteristics and Assessment Process	
- Validation/ external validation	Regressions were validated by performing within-sample 5-fold cross-validation. The average error between predicted and observed BMI was 1.60 kg/m^2 for females and 1.25 kg/m^2 for males. Regressions with smaller number of cases (underweight category) had errors as high as 3.62 kg/m^2. There appeared to be little bias in the error, so that the overall distribution of BMI in the entire sample was well represented. The 112 R-squares values derived from the regression models ranged from 35% to 85%.

	After implementing the regression in POHEM:CVD further external validation was completed by comparing the projected distribution of BMI to that observed in the CCHS in 2003, 2005 and 2007. In addition, further validation was undertaken to evaluate the projections over a longer period.
- Estimates produced/Valid for which subpopulations?	Not reported.
Plain Language Summary	Using the longitudinal NPHS, a set of regression models to estimate weight change over time were estimated for Canadian adults. The outcome modeled was change in BMI and explanatory variables were age, sex, income quartile, education and region of residence in Canada and previous BMI value(s). The models were internally validated using 5-fold cross validation and externally validated by comparing projections of BMI categories to observed data collected in later years (CCHS).
References	None published

Appendix 5

Updating age, population structure and sociodemographic characteristics

Canada, like many developed countries, maintains population projections based on programs that project each component of population change (i.e., projected births based on historic trends and projections of total fertility rates and number of women). The population projections and related program components are used for a wide range of national planning, both within and outside of government, such as planning the Canadian Pension Plan (within government). Our study's model population structure was updated using the same national standard population projections and components, which for the most part are developed and maintained at Statistics Canada.

When an immigrant actor was newly introduced into the model between 2002 and 2010, their sociodemographic and CVD was assigned based on a randomly chosen respondent of the same age, sex, province of immigration, and recent immigrant status from the CCHS 2001. After initialization, immigrants are governed by the same equations as the general population.

Mortality hazards were estimated from vital statistics and other sources.[27] In the model, the risk of dying is evaluated every year of the person's simulated life and is conditional on their year of birth, sex, and age. When a simulated individual develops AMI, their risk of death changes from that of the general population (captured by analyzing Ontario survival data from time of index AMI to death from all causes).