Article details: 2012-0015	
Title	Projections of preventable risks for cardiovascular disease in Canada from 2001 to 2020: a microsimulation modelling approach
Authors	Douglas Manuel, Meltem Tuna, Deirdre Hennessy, Carol Bennett, Anya Okhmatovskaia, Philippe Finès, Jack Tu
Reviewer 1	**Robert Tate**
Institution	University of Manitoba, Community Health Sciences
General comments	Review of manuscript by Douglas Manuel, et.al. … CMAJOpen-2012-0015 The overall objective is to project trends in "preventable CVD risk factors" in the Canadian population over the next 10 years. An often employed approach to project risk factors, disease events, hospital usage, etc. into the future is to combine a projection of the population size, stratified by age, sex, region, etc. with a projection of the number of events of interest within each stratum. The accepted methodology to project future events is most frequently based on a regression model of already observed factors/events/rates over previous times. Credible results hinge on the credibility of each of the two components. Further the credibility of the results depend on the credibility, or believed accuracy of one's ability to predict each component within each strata of interest. Of course the accuracy, and usefulness of any projection will not be known during the short term. The team of researchers assembled to attack the current question has much expertise with epidemiology, cardiovascular disease, administrative data, Statistics Canada files, and health services research. The POHEM microsimulation model is a great tool to be used for this project in Canada. The authors talk about 18 characteristics, when in fact they really only use 12. I have little concern over the population projections, age, sex, region, etc. but I have some concern about the trends in, and projections of the risk factors, themselves. Specifically, I'm not convinced that the projections for hypertension and high cholesterol will have any utility. Both are based on a single value from the CHHS conducted twenty years ago. There was some discussion of the impact of improved treatment over time, and the implication that might have on future rates of both, but the data presented in Figure 3 is not at all convincing, or reassuring. I suggest that there is merit in the publication of this paper, but ask that the authors add some clarification to their methodology, specifically describe 12 characteristics not 18, i.e. maybe just delete heart disease, arthritis, osteoarthritis, etc. from Appendix 2. The biggest question to address comes from a careful look at Figure 3, where historical trends and future projections for hypertension and high cholesterol appear to based on a single point. Indeed, trends and projections should never even be based on two points. I ask that authors address this in their discussion section.
Reviewer 2	**Jeffrey Bakal**
Institution	University of Alberta, Canadian VIGOUR Centre
General comments	The authors present the results of a micro simulation model for the prediction of CVD risk factors in a national-level population. The study is an interesting concept, but could use some additional clarity on a few points. While the simulated population appears to be validated against the national data (which presumably drove the parameter estimates), its not clear to me why the 5% level of agreement is acceptable especially for forecasting. Would be interesting to see if the distribution of the "actors" was the same as the general population on other multivariate predictors (e.g., Charlson). Since the rates of comorbidities etc are forecast there should be some sensitivity analysis around the estimates and an idea of the impact on the risk models. A few minor points

	1)Ref 2 and 17 are the same 2) There is an error from the automated reference manager on page 27. 3) In the table I would remove the line for No Diabetes
Author response	Thank you for your review and comments. We have addressed each of the raised points below. Editor's comments Editor, Comment 1: The model appears to be derived from and validated on the same data. The Editor is correct that CCHS 2001 was used for model initiation and we also use the same data to validate the baseline estimates in 2001. All subsequent validation data (CCHS 2003 to 2009 and the two CHMS (2007 and 2011) were not used as derivation data. The use of the CCHS 2001 for validation confirms that the model was correctly specified at baseline – which should be a relatively straightforward technical exercise. To clarify, we have removed reference to the 2001 validation and have clarified in the text (page 9, results section, paragraph 2) that we expect the estimates in 2001 to be in line with those produced by the model. We have also clarified throughout the text that validation of the estimates take place from 2003 onwards. Editor, Comment 2: The projections for hypertension and hypercholesterolemia appear to be based on a single value, imputed from CHHS. It is unclear how a trend can be predicted. Review 1, Comment 1: Specifically, I'm not convinced that the projections for hypertension and high cholesterol will have any utility. Both are based on a single value from the CHHS conducted twenty years ago. There was some discussion of the impact of improved treatment over time, and the implication that might have on future rates of both, but the data presented in Figure 3 is not at all convincing, or reassuring. Reviewer 1, Comment 3: The biggest question to address comes from a careful look at Figure 3, where historical trends and future projections for hypertension and high cholesterol appear to based on a single point. Indeed, trends and projections should never even be based on two points. I ask that authors address this in their discussion section. The Editor's and 1st Reviewer's comments in relation to blood pressure and cholesterol projections are very similar so we have addressed them here in one section. The Editor's comment highlights a lack of clarity around how we describe the imputation and subsequent transition of blood pressure and lipid values in the model. We have clarified the description of the approach in the text and appendices and have expanded the description of the assumptions and limitations. We have observed that most previous CVD models perform projections on risk factors based on two points. Furthermore, many /most CVD models do not use individual-based longitudinal cohort data that was available to us for most risk factors we examined. However, the method used to project blood pressure and lipid levels was necessarily different due to data availability. The approach to project blood pressure and cholesterol values follows a period, actuarial or life table approach where age-specific cross-sectional data is assumed to follow a life course perspective. For example, the change in blood pressure in a 40 year old male from 2001 to 2006 (male now aged 45 years) is assumed to be the same as the observed difference between a 40 and 45 year old male in the cross-sectional data (CHHS, 1986-1992)- meaning that there is a life course of blood pressure change that is independent of a cohort effect (calendar time). This approach is the same as the period life expectancy (as commonly reported for all countries internationally, including Canada and the provinces). The assumptions are notable but, as Keyfitz stated, inform our basic understanding of life. The approach is the most robust method available when population-based individual longitudinal data for physical measures is lacking – a very common situation worldwide. In many regards, the assumption is appropriate for our model because the predictive

	approach reflects a baseline life course of cholesterol and blood pressure that is unaffected by the institution of therapeutics or other influences. In this way, we can first model the baseline life course of untreated blood pressure and cholesterol and then add a separate factor on level of treatment – our calibration factor (see manuscript text page 8, paragraph 4). This allows model users to vary the future predictive treatment coverage level to assess future trends in population blood pressure and lipid levels. The inflection in Figure 3 reflects the base assumption that coverage does not change further in the future. Blood pressure and cholesterol levels will decline faster than we project if coverage levels increase further than was reported in the CHMS 2007 and 2011. Text has been added in the main manuscript (page 7, paragraph 3) and Appendix 4.2 to clarify how we modelled blood pressure and cholesterol values. In addition, we have added a paragraph to the limitations section to reiterate that the blood pressure and lipid level projections are based on data available at a single time point. To further summarize: - the categories of lipid and blood pressure values were imputed from the Canadian Heart Health Survey (CHHS), into the initial population (CCHS 2001). - the joint transition probabilities, which govern how individuals move between lipid and blood pressure categories in the subsequent simulation were derived from cross-sectional data. The method, transport flow methodology, maximized the information gained from the data by taking the joint distributions of total cholesterol, blood pressure, BMI, and diabetic status and deriving the joint probability of changing cholesterol and blood pressure states from one age group to the next. Since age is the most important risk factor for high blood pressure and high cholesterol it was assumed that the transition probabilities between age groups estimated from the cross-sectional CHHS would also apply over time as individuals in the population aged. Editor, Comment 3: Some of the risk factors overlap (e.g. obesity/hypertension/lipids). We'd like more information in the methods describing how the effects of individual factors can be teased out and a discussion of the implications of this in the limitations section of the discussion. This may result in a toning down of the conclusion. We weren't quite sure how to interpret your comment on "the effects of individual risk factors". Where you speaking to the effects of individual risk factors on the predictive risk of developing cardiovascular disease? We referenced two ways that this has been examined in previous modelling studies: Wijeysundera used an "epidemiology" approach that assumes the risk from individual risk factors follows what is reported in epidemiology literature, including meta-analyses. The risks, if the underlying studies have been well performed, are adjusted risks, meaning the effect has been teased out from other risks. The second approach assumes the risk hazards follow the beta-coefficients of predictive algorithms such Framingham. The beta-coefficents are examined in the presence of multiple risks we have examined in POHEM:CVD. It is clear to us that if we incorporated the hazards for risks from either approach, the CVD incidence would decline. That said, we appreciate that we did not perform these analyses. Our plan, as stated, is to use predictive risk algorithms such as Framingham. We appreciate that the there is considerable interaction and overlap between the risk factors. Capturing the correlation of risk was an explicit purpose of the microsimulation modelling approach and, in our view, an important advance over other modelling approaches. That said, we appreciate the overlap contributes considerably to the complexity and that we could do a better job of explaining the methods – particularly in the main text. We have made modifications in the appendices and text, as mentioned in the previous response, to more fully explain the way hypertension and cholesterol were modelled. We have also added a section to the discussion (page 11, paragraph 2) and toned down the conclusions (page 13, paragraph 1) in response to the Editor's concerns. Editor, Comment 4: How is the variability of the estimates accounted for? The Editor's comments reflect the fact that we did not produce variance estimates around the projections. Creating variance estimates for microsimulation projections is very complex and computationally intensive and methods currently under development

	by the Simulation Technology for Advanced Research team to produce variance estimates, however they are not available at the time of this resubmission. We note that previous CVD modelling has not included variance estimates of risk factor prevalence and usually incorporates only rough estimates of overall CVD predictions. To provide additional guidance to the reader we we have produced and graphed the confidence intervals for the validation estimates, see revised Figures 3 and 4. Reviewer 1- Major comments Reviewer1, Comment 2: (the model describes) 12 characteristics not 18, i.e. maybe just delete heart disease, arthritis, osteoarthritis, etc. from Appendix 2. The overall POHEM:CVD model includes arthritis and osteoarthritis, but these were not reported or used in the current manuscript, as the Reviewer points out. We have deleted reference to these extra characteristics in the text and in Appendix 2. Reviewer 2- Major comments Reviewer 2, Comment 1: While the simulated population appears to be validated against the national data (which presumably drove the parameter estimates), its not clear to me why the 5% level of agreement is acceptable especially for forecasting. The 5% level of agreement was predetermined at the time of model specification in consultation with policy actors and modellers. We have referred to this process in the text (page 8, paragraph 4). We were reassured that most estimates fell within the 5% level of agreement between predicted and observed estimates. The observed estimates from 2003 onward were not used to generate model parameters, so meet the definition of external data and validation. We were not surprised that blood pressure and cholesterol did not meet the 5% threshold for reasons described in the manuscript. To reiterate, the projected, uncalibrated estimates did not consider the change in treatment coverage, which was considerable for the time period leading up to 2007. Reviewer 2, Comment 2: Would be interesting to see if the distribution of the "actors" was the same as the general population on other multivariate predictors (e.g., Charlson). We appreciate the Reviewer's comment about assessing the calibration of the model across other multivariate predictors (not in the model), such as is done for classical prediction models. We hope to perform these additions in the future. We would like to point out that, to our knowledge, ours is the first CVD modelling study to incorporate and assess the distribution of risks using a multivariable or co-related approach. This includes the recently reported burden of global burden of disease estimates reported in Lancet Dec 2012. Reviewer 2, Comment 3: Since the rates of comorbidities etc are forecast, there should be some sensitivity analysis around the estimates and an idea of the impact on the risk models. Please see our response to the Editor's comment #4 above Reviewer 2- Minor comments Ref 2 and 17 are the same We have changed the reference section to address the Reviewer's comment. There is an error from the automated reference manager on page 27. We have changed the reference section to address the Reviewer's comment. In the table I would remove the line for No Diabetes We have changed Table 1 to address the Reviewer's comment.