Identification of predictors

Twenty-one predictor variables were identified and selected based on previous stroke and diabetes risk algorithms derived using the same data source.(1, 2) All 21 predictor variables were included after a formal check of multicollinearity. An indicator variable for immigration status together with fraction of life lived in Canada was used to account for both recent and non-recent immigrants. Indicator variables for smoking status were created to allow inclusion of smoking pack-years as a continuous predictor. Detailed definitions and measurement of these variables are presented in **Table 1**. The specific wording of questions are available as online appendices and within the survey documentation.(3, 4)

Data cleaning and coding of predictors

Data cleaning and coding was completed without examining outcome-risk factor associations. After inspection of histograms and boxplots, continuous variables were truncated to the 99.5th percentile.

Missing data

We used multiple imputation to impute missing values on predictor variables.(5) The imputation model consisted of the full list of predictor variables, time to event and censoring variables, and auxiliary variables—that is, variables that are not predictors but may nevertheless be useful in generating imputed values (for example, income and self-perceived health). We generated five multiple imputation datasets. The imputation procedure employed predictive mean matching to generate imputed values and used the bootstrap to approximate the process of drawing predicted values from a full Bayesian predictive distribution. Continuous variables in the imputation model were transformed using restricted cubic splines. The final prediction model was estimated separately for each imputation-completed dataset and the results combined using the rules developed by Rubin(6) to account for imputation uncertainty.

Model specification

We fit a preliminary main effects model that included an initial degree of freedom allocation for each predictor.(5, 7) Continuous predictors were flexibly modelled using restricted cubic splines, i.e., piecewise cubic functions that are smooth at the knots and restricted to be linear in the tails. The knots were placed at fixed quantiles of the distribution: in particular, at the 5th, 27.5th, 50th, 72.5th and 95th percentiles for 5 knot splines and at the 10th, 50th and 90th percentiles for 3 knot splines. Ordinal variables with few categories were specified as either linear terms, or as categorical if the expected association was more complex than linear (i.e., level of area deprivation).

The initial model specification, presented in **Table 2**, included a total of 61 degrees of freedom (48 main, 13 interaction), compared to a possible maximum (based on number of events) of 110. We allocated the final degrees of freedom to individual predictors based on a partial test of

association with the outcome. Partial association chi-squared statistics for each predictor variable minus their degrees of freedom were plotted in descending order. We retained the initial degrees of freedom for variables with higher predictive potential, but predictors with lower predictive potential were modelled with reduced degrees of freedom, i.e., as simple linear terms or after combining infrequent categories. Interaction terms, specified above, were added to the final model and were restricted to linear terms.

The full models were specified in accordance with the pre-specified plan of maintaining all predictors in the model but with reduced DFs, reflecting the importance in the partial correlation plots. After applying the step-down procedure, the final application (reduced) model had 36 and 37 degrees of freedom with 12 predictors (6 and 7 continuous) and 11 interaction terms for female and male models, respectively.

Model estimation

Models were estimated using a proportional hazards model for the subdistribution of a competing risk with death from a non-CVD cause considered as a competing risk.(8) All predictors were centred about their means prior to analysis. The proportionality assumption was assessed using plots of raw and smoothed scaled Schoenfeld residuals versus time for each predictor. Influence of data points was assessed by plotting scaled df beta residuals for each covariate. The degree of over-fitting (shrinkage) in the model was estimated using the heuristic shrinkage estimator (based on the log likelihood ratio chi-square statistic for the full model).(9) Survey weights were not used for model development. We recommend survey weight for population application when available or used in population application data.(10) Analyses were conducted using cmprsk and Harrell's HMisc(5) and rms packages of functions in R(11) as well as SAS v9.3.

Assessment of model performance

Nagelkerke's R^2 and the Brier score were calculated as overall measures of accuracy. Discrimination was assessed using Harrell's overall concordance statistic, with 95% confidence intervals estimated using bootstrap samples. Internally validated performance measures were obtained from 200 bootstrap samples, using the procedure described by Steyerberg.(12)

Calibration plots were created by comparing mean 5-year predicted probabilities with cumulative incidence function estimates of observed rates stratified by deciles of predicted risk. The calibration slope was estimated by including the linear predictor as a single term in the model fitted to the validation cohort. Deviation from a slope of 1 was tested using a Wald test. The calibration slope reflects the combined effect of over-fitting to the derivation data as well as true differences in effects of predictors. All model performance measures were calculated using the first of the multiply imputed datasets.

No major violation of the proportional hazards assumption was observed. Inspection of influence diagnostics revealed a small number of influential observations; these values were inspected for measurement error. The degree of over-fitting in the model was minimal (heuristic shrinkage estimator 0.983 for females and 0.978 for males) indicating that the risk of over-fitting was low.

Risk groups

Subgroup validation was implemented as a conceptually easy check of calibration by comparing observed and predicted risks within predefined subgroups of importance to clinicians and policy makers.(13) We examined subgroups using predefined criteria for clinically or policy relevant standards of calibration (<20% difference between observed and predicted estimates for categories with prevalence higher than 5%). In total, there were 94 subgroups for females and 111 for males, which were defined based on age, behavioural risk exposure categories, health regions, socio-demographic groups, hypertension status, and diabetes status.

Estimation of the final parsimonious model

The parsimonious model was generated by deleting variables to a desired degree of accuracy based on contribution to model R^2.(9) To maximize the amount of data for the final model, the final regression coefficients were estimated using the combined data from both the derivation and validation cohorts with outcome events updated to reflect the most recent years available.

To maximize the amount of data for the final model, the final regression coefficients were estimated using the combined data from both the derivation and validation cohorts with outcome events updated to reflect the most recent years available.

Sensitivity testing

We performed the following sensitivity testing and exploratory analyses that were not specified in our initial protocol. None of these analyses were incorporated into the final algorithms. Algorithms were generated using a Cox proportional hazards model without competing risks. There was a non-significant improved calibration in several subgroups, particularly those at higher risk deciles and for people older than age 70 years (see **Additional file 9**). The models had very similar discrimination (C-stat: 0.84 versus 0.82 for males; 0.87 versus 0.86 for females – reduced model). The observation of small improvement or no difference was unexpected based on previous studies of cardiovascular risk.(14, 15) Further examination of the role is warranted.

In exploratory analyses, we examined models with more specific exposure ascertainment (more degrees of freedom) where differences in calibration were greater than 20%. These models included specifying age using greater knots in the restrictive cubic splines and other continuous forms(16), physical activity considering exposure of zero METs, as well adding measures of fragility that may be associated with low physical activity. In addition, we examined expanded categories of ethnicity, including Chinese, South Asian, East Asian and Black, and further interaction terms. None of these models improved predictive performance, including both discrimination and calibration assessment within the relevant subgroups.

Calculation of heart age

Heart age is calculated by comparing an individual's CVDPoRT risk to the average age- and sex-specific CVDPoRT risk in the Ontario population using the full development data, Ontario

sample of the CCHS 1.1 to 4.1 (2001-2006), mean risk by age, 5-year moving average (see table below).

Age	5-year CVDPoRT risk of incident CVD, Ontario 2001 - 2006	
	Mean male	Mean female
20	0.000177	0.000246
21	0.000237	0.000284
22	0.000304	0.000322
23	0.000388	0.000371
24	0.000498	0.000413
25	0.000624	0.000464
26	0.000758	0.000518
27	0.000912	0.000617
28	0.001076	0.000710
29	0.001325	0.000823
30	0.001610	0.000897
31	0.001983	0.000955
32	0.002377	0.001010
33	0.002823	0.001139
34	0.003307	0.001315
35	0.003846	0.001539
36	0.004612	0.001746
37	0.005521	0.001920
38	0.006572	0.002248
39	0.007496	0.002507
40	0.008570	0.002951
41	0.009705	0.003166
42	0.011064	0.003600
43	0.012206	0.003970
44	0.013832	0.004448
45	0.015622	0.004920
46	0.017283	0.005342
47	0.018687	0.005732
48	0.019924	0.006189
49	0.021033	0.006735
50	0.022331	0.007338
51	0.023514	0.007916
52	0.025441	0.008633
53	0.027009	0.009538
54	0.027961	0.010259
55	0.028794	0.010961

56	0.029511	0.011721
57	0.031275	0.012507
58	0.032953	0.013484
59	0.034962	0.014491
60	0.036757	0.016147
61	0.038411	0.017212
62	0.039427	0.018274
63	0.040573	0.019654
64	0.042365	0.021702
65	0.044648	0.023903
66	0.046977	0.025815
67	0.048806	0.027361
68	0.051398	0.029237
69	0.054247	0.030656
70	0.056716	0.033289
71	0.059167	0.036023
72	0.061266	0.039495
73	0.064792	0.042323
74	0.067395	0.044990
75	0.070273	0.048040
76	0.070809	0.051702
77	0.074041	0.055665
78	0.078739	0.059357
79	0.083745	0.063752
80	0.086496	0.068913
81	0.088092	0.073081
82	0.091158	0.077341
83	0.096163	0.081097
84	0.100321	0.088140
85	0.108525	0.094434
86	0.114049	0.100503
87	0.120773	0.104461
88	0.120888	0.109521
89	0.128259	0.124560
90+	0.130378	0.132655

1. Rosella LC, Manuel DG, Burchill C, Stukel TA, 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. 2011;65(7):613-20.
2. Manuel DG, Tuna M, Perez R, Tanuseputro P, Hennessy D, Bennett C, et al. Predicting Stroke Risk Based on Health Behaviours: Development of the Stroke Population Risk Tool (SPoRT). PloS one. 2015;10(12):e0143342.

3. Beland Y. Canadian Community Health Survey - Methodological Overview. Health Reports. 2002;13(2):9-14.
4. Canada H. Chronic Diseases in Canada (Health Canada Report). PMID: 14733756; 2003. p. 81.
5. Harrell FE. HMisc Package. http://biostat.mc.vanderbilt.edu/wiki/Main/Hmisc ed: GitHub; 2015.
6. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med. 1991;10(4):585-98.
7. Harrell FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis. Harrell FE, Jr., editor. New York, NY: Springer-Verlag New York, Inc.; 2001. 1-568 p.
8. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association. 1999;94(446):496-509.
9. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. Stat Med. 2002;21(24):3803-22.
10. Manuel DG, Perez R, Sanmartin C, Taljaard M, Hennessy D, Wilson K, et al. Measuring Burden of Unhealthy Behaviours Using a Multivariable Predictive Approach: Life Expectancy Lost in Canada Attributable to Smoking, Alcohol, Physical Inactivity, and Diet. PLoS Med. 2016;13(8):e1002082.
11. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0; 2014.
12. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. London: Springer; 2009.
13. Taljaard M, Tuna M, Bennett C, Perez R, Rosella L, Tu JV, et al. Cardiovascular Disease Population Risk Tool (CVDPoRT): predictive algorithm for assessing CVD risk in the community setting. A study protocol. BMJ open. 2014;4(10):e006701.
14. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks. Stat Med. 2016;35(22):4056-72.
15. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data: John Wiley & Sons; 2011.
16. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. Methods Inf Med. 2005;44(4):561-71.