# DEMENTIA IN THE POPULATION: INVESTIGATION USING POPULATION-LEVEL ADMINISTRATIVE HEALTH DATA IN ONTARIO, CANADA

Stacey Lynn Fisher

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for a
PhD in Epidemiology

School of Epidemiology and Public Health,
Faculty of Medicine, University of Ottawa

**ACKNOWLEDGEMENTS**

I would first like to thank my supervisor, Dr. Doug Manuel, whose enthusiasm and tireless dedication to his work continually inspires me. Thank you for your mentorship and guidance, and for providing me with countless opportunities to learn and become a better epidemiologist and population health scientist. To Dr. Peter Tanuseputro for his mentorship, clinical expertise and for encouraging me to take time to have a little fun. To my other committee members, Dr. Monica Taljaard and Dr. Geoff Anderson for providing valuable feedback and insight. Special thanks also to Dr. Meltem Tuna, Anan Bader Eddeen, Sarah Spruin and Yulric Sequeira for their methodological expertise and support. I would also like to thank Carol Bennett, Dr. Amy Hsu, Dr. Mahsa Jessri, Sarah Beach and Karen Pacheco for their support and encouragement. To Dr. Laura Rosella for her support and for generously providing me with space to complete this work from Toronto, and to her team for their support during the last year. To Katie and Mark, Marie and Gauruv, and Julia for their friendship and for taking me in as I travelled between Toronto and Ottawa. Finally, thank you to my husband Kirles for his unending patience, encouragement and love, and to my family for always supporting me no matter what path I choose, or how far away it takes me.

**TABLE OF CONTENTS**

**ABSTRACT**

As life expectancy increases and populations age, the burden of disease from dementia is becoming an increasingly significant concern worldwide. The goal of this thesis was to support population health decision-making for dementia, including health resource planning and the development of preventative strategies.

A population risk tool for predicting dementia risk among community-dwelling older adults was pre-specified, developed and validated using survival methodology and population-based health surveys linked to administrative data in Ontario. The resulting Dementia Population Risk Tool (DemPoRT) is discriminating and predicts dementia risk across a range of health profiles. This algorithm was then applied to a recent national population health survey and used to describe the population risk and future burden of dementia in Canada, and the impact of unhealthy lifestyle behaviours. In five years, there were approximately 450,000 new dementia cases among individuals 55+ years of age in Canada, of which 40% were associated with unhealthy lifestyle behaviours. Lastly, life expectancy and health care utilization of individuals with dementia was investigated. Administrative data was used to identify all individuals with physician-diagnosed dementia in Ontario from 2014 to 2017, and complete period life table methodology was used to estimate life expectancy and the expected health care utilization of individuals from diagnosis to death. Life expectancy at dementia diagnosis was 5.8 years and was longer for women and those diagnosed before the age of 75. Approximately half of this time was spent in long-term care or receiving home care, inpatient care or outpatient care.

Development of policy to prepare for and reduce the burden of dementia requires population health planners to have reliable projections of future disease prevalence and resource requirements, and understanding of what subgroups of the population are at the highest risk and have the highest absolute burden. The results of this thesis will support implementation of Canada's recently-released dementia strategy by providing detailed information about the current and future burden of dementia, and by informing the development of population-based primary prevention strategies.

## LIST OF TABLES

## LIST OF FIGURES

**LIST OF APPENDICES**

**LIST OF ABBREVIATIONS**

AIC: Akaike information criterion

APOE: Apolipoprotein E

C-statistic: Calibration statistic

CCHS: Canadian Community Health Survey

CCRS: Continuing Care Reporting System

CI: Confidence interval

COPD: Chronic obstructive pulmonary disease

CPS: Cognitive Performance Score

DemPoRT: Dementia Population Risk Tool

Df: Degrees of freedom

DRS: Dementia Risk Score

EMR: Emergency medical record

FCSRT: Free and Cued Selective Reminding Test

HC: Home care

sHR: Subdistribution hazard ratio

IQR: Interquartile range

LTC: Long-term care

MET: Metabolic equivalent unit of task

OHIP: Ontario Health Insurance Plan

PAF: Population attributable fraction

PMML: Predictive Modelling Markup Language

POHEM: Population Health Models

RCS: Restricted cubic spline

RR: Relative risk

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

RAI- HC: Resident Assessment Instrument-Home Care database

UI: Uncertainty interval

# CHATPER 1: Thesis Introduction

## 1.1 BACKGROUND

Dementia is a growing global concern as incidence increases dramatically with age, and average life expectancy has been increasing around the world[1,2]. An estimated 600,000 Canadians and 47 million individuals worldwide have dementia, which has been projected to grow to 1.1 million Canadians by 2038 and 140 million worldwide by 2050[3–5] when global costs of dementia are expected to reach $2 trillion[6]. The goal of this thesis was to support population health decision-making for dementia, including informing health resource planning and the development of prevention strategies.

### *1.1.1 Epidemiology of dementia*

Dementia is a general term for cognitive deterioration severe enough to interfere with a person's ability to perform daily activities[4]. It is characterized by cognitive and behavioural symptoms including changes in personality and mood, and problems with memory, reasoning, language and communication. Dementia is not itself a disease, despite often being referred to as such, but rather a group of symptoms; any pathology that causes significant damage to the brain can cause dementia[7,8]. There are therefore many types of dementia, including some which are reversible, however most dementias are irreversible and the result of a long underlying disease process. These include Alzheimer's disease, vascular dementia and mixed dementia. Alzheimer's disease is characterized by brain degeneration including neuronal loss and accumulation of amyloid-β peptide and hyper-phosphorylated tau. Vascular dementia is associated with impaired blood flow to the brain from stroke or other complications of large and small vessel diseases. Most elderly

dementia patients present with both degenerative and vascular pathology, known as mixed dementia[9–11].

Dementia has been called the "greatest global challenge for health and social care in the 21st century"[5]. Estimated global costs in 2018 were US$1 trillion and are projected to double by 2030[12]. Incidence rates are strongly associated with age, approximately doubling every 10 years after the age of 60[13]. Recent population-based studies suggest that incidence rates of dementia may be declining in high income countries, however, and that rises in dementia prevalence are driven by a growing number of older adults and longer survival with the disease[14]. Canadian age-standardized incidence rates in 2015 among those 65+ years of age were 13.8 and 15.4 per 1,000 in 2010, falling by 4.7% and 7.1% compared to 2010 rates, among males and females, respectively[15]. An estimated 43% of individuals who live to age 65 will develop dementia[16].

In Canada, age-standardized all-cause mortality among those with dementia is over 4 times higher than in those without dementia[15]. Life expectancy estimates vary greatly due to differences in patient selection and study design. Many studies of dementia survival are performed in highly selected groups of patients, such as those who have been referred to specialistic clinics or volunteered to participate in research studies, resulting in selection bias. Differences in age distribution can also have strong effects. For example, a study using population-based data from the Canadian Study of Health and Aging reported a 3.3 year median survival for individuals with Alzheimer's dementia, with a mean age of 84[17]. In contrast, a study of patients referred to a neurology clinic had a mean age of 70 years and a median survival of 11.8 years[18].

*1.1.2 Risk factors for dementia*

The strongest risk factor for dementia is age; an estimated 80% of people with dementia are 75 years of age and older[16,19]. Other non-modifiable risk factors include family history[20,21], and genetic susceptibility genes[22,23]. The apolipoprotein (APOE) ε4/ε4 genotype, for example, has been associated with over 10 times the odds of Alzheimer's disease compared to the most common ε3/ε3 genotype[24].

Many of the risk factors for both vascular and Alzheimer's dementia are associated with vascular health, including diabetes, mid-life obesity, hypertension and hyperlipidemia[25–27]. Diabetes, for example, has been associated with a 46% increased risk of Alzheimer's disease and 148% increased risk of vascular dementia[28]. Midlife obesity has been associated with a 41% increased risk of dementia[29]. Years of formal education, traumatic brain injury, history of depression, sleep disturbances, level of social engagement and poor oral health have also been associated with development of dementia[26,27,30,31].

Behavioural risk factors of dementia include smoking, poor diet, physical inactivity, and heavy alcohol consumption[25–27]. Smoking is a risk factor of vascular disease and is harmful to the blood-brain barrier and to cerebral blood flow. Current smoking, compared to never smoking, has been associated with an almost 80% increased risk of both Alzheimer's and vascular dementia[32]. Poor diet is hypothesized to cause dementia and cognitive decline through obesity and/or oxidative stress[33,34]. Dietary improvements have been found to improve cardiovascular risk factors and decrease inflammation that may be associated with dementia development[35,36]. Physical activity is also associated with reducing vascular risks, while inactive lifestyles may

cause dementia through association with obesity, higher blood pressure, chronic disease and depression[37,38].

There is evidence that heavy alcohol consumption increases dementia risk, and that mild to moderate consumption, compared to abstinence, may be protective against cognitive impairment and dementia[39–42]. Chronic heavy drinkers are often malnourished and have nutritional deficiencies that can damage the brain, leading to alcohol-induced dementia. Ethanol is also known to be neurotoxic and can lead to irreversible structural and functional damage[43]. Heavy alcohol use can also lead to other conditions that damage the brain including hepatic encephalopathy from cirrhotic liver disease. Although many studies have reported a protective effect for mild to moderate alcohol consumption, methodological concerns and residual confounding prevent recommendations for mild alcohol consumption among abstainers [39,44–46].

*1.1.3 Population health planning for dementia*

The goal of population health is to improve the health of the entire population, while reducing inequities. It focuses on all of the determinants of health including the social, economic and physical environment, individual characteristics and lifestyles, health care services, and human biology[47]. Population health has many tasks, including surveillance of disease and disease risk factors, estimation of the future burden of disease, informing health policy and resource allocation, and the development, implementation and evaluation of population-level strategies for disease prevention and health improvement. Given the increasing burden, population health planning for dementia has become a major priority of governments and organizations around the world[1,4,48–50].

Canada's first national dementia strategy was released in June 2019, backed by a 50 million dollar investment[48]. It has three main objectives: 1) prevent dementia; 2) advance therapies and find a cure, and; 3) improve the quality of life of people living with dementia and their caregivers. The strategy recognizes the importance of disease surveillance and data to understanding the scope of dementia and focusing research efforts and health care resources to where they will have the largest impact. It focuses largely on population health approaches to dementia prevention, with particular emphasis on the promotion of healthy lifestyle by all Canadians, not just those currently living with or affected by dementia, and recognizes that healthy living from an earlier age may prevent or delay dementia onset.

Strategies for disease prevention from the population perspective differ from individual-level strategies. While preventing disease in an individual is best accomplished by reducing or eliminating risk factors associated with the largest relative risks, prevention of disease in a population is best accomplished by reducing or eliminating the causes of disease incidence[51]. Population approaches therefore often target upstream determinants of disease risk and promote changes to behavioural norms that make it easier for individuals to make healthy choices[52]. As there is no cure for dementia and no disease-modifying therapies, population-level prevention of dementia is an important goal of population health planning and policy development.

### *1.1.4 Prediction algorithms for population health planning*

A useful tool for population health planning and disease prevention are predictive risk algorithms. Predictive risk algorithms can be used for diagnostic or screening purposes, where the algorithm is designed to model the probability of disease presence, or for prognostic

purposes, where risk of future disease development is modelled. More generally, population risk algorithms can be used to describe population risk of disease, identify populations at high risk, project future disease risk and burden, estimate the impact of risk factors on disease burden, and evaluate the potential population health benefit of risk factor reduction[53].

Risk algorithms are evaluated using measures of discrimination and calibration. Discrimination is a measure of how well the model is able to separate those who experience the outcome from those who do not. It is most often quantified using the concordance statistic (c-statistic), or area under the curve[54]. C-statistics range from 0 to 1, where 0.5 indicates that the model performs no better than chance and 1 indicates that the model perfectly predicts who will and will not experience the outcome. Calibration is a measure of how well the predicted probabilities agree with the observed proportions, i.e. how well predicted risk agrees with the observed risk. Calibration within deciles of predicted risk are often plotted and calibration summarized using the slope and intercept of the plot. A perfectly calibrated algorithm would have a slope of 1 and an intercept of 0. Other measures of overall model performance include the Brier score[55], a measure of overall agreement between the observed and predicted risk, and the Nagelkerke $R^2$, a measure of how well the model explains the risk variation in the data[56]. Practically speaking, in a regression model, discrimination is associated with the beta coefficients, while calibration is associated with the baseline risk.

Clinical algorithms are primarily concerned with discrimination, as they are designed to inform decisions about disease prevention and treatment within individuals. Calibration is therefore often not assessed during clinical algorithm development. Predictive algorithms to be used in the

population setting and for population health planning purposes, however, should generally prioritize calibration over discrimination, as accurate prediction is essential when predicting the number of future cases of disease[53]. For example, a model that accurately predicts that the risk of dementia among those who smoke is three times that of those who do not smoke is discriminating. However, if the predicted risks were 1% and 3%, while the observed risks were 10% and 30%, the model is poorly calibrated, and not suited for population health purposes. Discrimination is also important in the development of prediction algorithms for population health purposes, however, as the identification of social inequities requires adequate discrimination[53]. As discrimination and calibration mathematically 'compete' with each other[57], the development of population health risk algorithms must take care to strike the right balance.

*1.1.5 Existing dementia prediction algorithms*

A recent systematic review found 61 studies describing dementia risk prediction models, most of which modelled risk of developing dementia in non-demented individuals in late life (N=39)[58]. Four studies focused on those in mid-life and three studies developed models for predicting dementia risk among those with diabetes. Another 15 studies predicted risk of transitioning from mild cognitive impairment to Alzheimer's disease. Most models were very simplistic, with 24 studies including less than 5 variables, and 11 studies including only 1 variable. Of the late life models, most (62%; N=24) require some form of neuropsychological assessment such as the Mini-Mental State Examination or the related Modified Mini-Mental State, the Free and Cued Selective Reminding Test (FCSRT), or a word list delayed recall test. Several models also include APOE genotype, MRI findings, serum cholesterol, blood pressure, and hemoglobin

levels. Half (N=19) include some measure of education, and only 28% (N=11) include any modifiable health behaviours (smoking, alcohol consumption, diet, or physical activity).

Two studies report models with excellent discrimination (c-statistic >0.90). Jungwirth *et al*. 2009 evaluated short neuropsychological instruments for predicting five-year incidence of Alzheimer's dementia in a cohort of 585 75-year old individuals[59]. The best performing model included reduced verbal memory, reduced visual motor speed, subjective memory complaints and the APOE ε4 carriage, with an c-statistic of 0.91. The other study, by Mura *et al*. 2017, investigated whether dementia prediction using the FCSRT was improved by the inclusion of socio-demographic variables[60]. In a cohort of 2558 community-dwelling individuals 65 years of age and older, three-year dementia risk was best predicted by supplementing the FCSRT score with age, sex, education level and the interaction of these with the FCSRT score.

The only study with a reasonable c statistic (>0.80) that did not include neuropsychological or genetic testing and could therefore potentially be used at the population-level is the Dementia Risk Score (DRS), which predicts five-year dementia risk among individuals 60 to 79 years of age using primary care data (c-statistic: 0.84)[61]. It includes age and BMI, and measures for smoking, heavy alcohol use, depression, social deprivation and aspirin use in addition to various disease states. It was developed using over 800,000 patients from a nationally representative primary care database in the United Kingdom. Potential usefulness for population health planning is limited, however by the lack of many sociodemographic and health behaviour variables, and the use of clinical data.

## 1.2 THESIS OBJECTIVES AND FORMAT

The goal of this thesis is to support population health decision-making for dementia through three main objectives:

1) Develop and validate a predictive risk algorithm for dementia designed specifically for population use;
2) Use this tool to describe the population risk and future burden of dementia in Canada, and estimate the impact of unhealthy lifestyle behaviours, and;
3) Examine the life expectancy and health care experience of those with dementia in Ontario.

The first objective of this thesis is to develop and validate the Dementia Population Risk Tool (DemPoRT), a predictive risk algorithm uniquely designed to predict risk of dementia in community-dwelling individuals at the population level. Sociodemographic, general health, behavioural, functional and health condition variables from the Canadian Community Health Survey (CCHS) linked to routinely-collected health administration data in Ontario, Canada were used. Dementia was ascertained by a validated case ascertainment definition. Fine and Gray subdistribution hazard models, considering death as a competing risk, were used to develop sex-specific models for five-year dementia incidence among those at least 55 years of age. Predictor variables and the analytic plan were fully pre-specified, registered and published prior to model development to improve transparency, and limit overfitting and bias. This is the first predictive risk tool for dementia designed specifically for population uses. As DemPoRT was developed using population-level data representative of Canadians and includes health behaviour and socio-demographic variables of interest to policy makers, results of this study can be used to inform

population health planning and resource allocation decisions, and to inform prevention strategies. The study protocol is in Chapter 2, while Chapter 3 presents the development and validation of the DemPoRT model.

The second objective, presented in Chapter 4, is to use the newly developed DemPoRT algorithm to describe the five-year population risk and burden of dementia in Canada, and to estimate impact of unhealthy lifestyle behaviours. DemPoRT was applied to 2013/14 CCHS to estimate the five-year risk and number of incident dementia cases among Canadians at least 55 years of age. To evaluate the impact of smoking, alcohol consumption, poor diet and physical activity, a cause-deleted risk factor approach and external risk factor estimates from five increasingly adjusted Cox proportional hazard regression models were used. It is important to understand the population risk of dementia and the distribution of risk across population subgroups to prepare for the growing burden of dementia. The impact of modifiable risk factors on the burden of dementia informs the development of preventative strategies.

The third and last objective of this thesis is to examine the life expectancy and formal care experience of those with dementia from both the individual and the health care system perspectives. Administrative health data of all individuals with dementia from 2014 to 2017 in Ontario and period life table methodology was used. This work informs physician conversations with patients, their families and caregivers around what to expect after diagnosis with dementia and provides important information to health planners about the burden of dementia on the health care system throughout the disease trajectory. This objective is presented in Chapter 5.

## 1.3 REFERENCES


1. Wortmann M. Dementia: a global health priority - highlights from an ADI and World Health Organization report. *Alzheimers Res Ther*. 2012;4(5):40. doi:10.1186/alzrt143
2. Alzheimer, Association, Alzheimer's Association. 2014 Alzheimer's Disease Facts and Figures Includes a Special Report on Women and Alzheimer's Disease. *Alzheimer's Dement*. 2014;10(2).
3. The Alzheimer Society of Canada. *Rising Tide: The Impact of Dementia on Canadian Society*.; 2010.
4. World Health Organization. *Dementia: A Public Health Priority*.; 2012. doi:978 92 4 156445 8
5. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390:2673-2734. doi:10.1016/S0140-6736(17)31363-6
6. Prince M, Comas-Herrera MA, Knapp M, Guerchet M, Karagiannidou MM. World Alzheimer Report 2016 Improving healthcare for people living with dementia coverage, QualIty and costs now and In the future. 2016.
7. Peng FCC. Is Dementia a Disease? *Gerontology*. 2003;49(6):384-391. doi:10.1159/000073767
8. Jellinger KA. Should the word "dementia" be forgotten? *J Cell Mol Med*. 2010;14(10):2415-2416. doi:10.1111/j.1582-4934.2010.01159.x
9. Aguero-Torres H, Kivipelto M, von Strauss E. Rethinking the dementia diagnoses in a population-based study: what is Alzheimer's disease and what is vascular dementia?. A study from the kungsholmen project. *Dement Geriatr Cogn Disord*. 2006;22(3):244-249. doi:10.1159/000094973
10. Schneider JA, Bennett DA. Where vascular meets neurodegenerative disease. *Stroke*. 2010;41(10 Suppl):S144-6. doi:10.1161/STROKEAHA.110.598326
11. Savva GM, Wharton SB, Ince PG, et al. Age, Neuropathology, and Dementia. *N Engl J Med*. 2009;360(22):2302-2309. doi:10.1056/NEJMoa0806142
12. Alzheimer's Disease International. *World Alzheimer Report 2018 - The State of the Art of Dementia Research: New Frontiers*.
13. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's Dement*. 2013;9(1).

doi:10.1016/j.jalz.2012.11.007

14. Prince M, Ali G-C, Guerchet M, et al. Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimers Res Ther*. 2016;8(1):23. doi:10.1186/s13195-016-0188-8
15. Public Health Infobase. Canadian Chronic Disease Surveillance System Data on Dementia from 2015-16, excluding Saskatchewan and Yukon. Government of Canada, 2018. https://health-infobase.canada.ca/ccdss/data-tool/. Published 2018.
16. Carone M, Asgharian M, Jewell NP. Estimating the lifetime risk of dementia in the Canadian elderly population using cross-sectional cohort survival data. *J Am Stat Assoc*. 2014;109(505):24-35. doi:10.1080/01621459.2013.859076
17. Wolfson C, Wolfson DB, Asgharian M, et al. A reevaluation of the duration of survival after the onset of dementia. *N Engl J Med*. 2001;344(15):1111-1116. doi:10.1056/NEJM200104123441501
18. Roberson ED, Hesse JH, Rose KD, et al. Frontotemporal dementia progresses to death faster than Alzheimer disease. *Neurology*. 2005;65(5):719-725. doi:10.1212/01.wnl.0000173837.82820.9f
19. Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurologia*. 2017;32(8):523-532. doi:10.1016/j.nrl.2016.02.016
20. Lautenschlager NT, Cupples LA, Rao VS, et al. Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old? *Neurology*. 1996;46(3):641-650.
21. Green RC, Cupples LA, Go R, et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. *JAMA*. 2002;287(3):329-336.
22. Stocker H, Möllers T, Perna L, Brenner H. The genetic risk of Alzheimer's disease beyond APOE ε4: systematic review of Alzheimer's genetic risk scores. *Transl Psychiatry*. 2018;8(1):166. doi:10.1038/s41398-018-0221-8
23. Liu C-C, Kanekiyo T, Xu H, Bu G, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9(2):106-118. doi:10.1038/nrneurol.2012.263
24. Farrer LA, Cupples LA, Haines JL, et al. Effects of Age, Sex, and Ethnicity on the

Association Between Apolipoprotein E Genotype and Alzheimer Disease. *JAMA*. 1997;278(16):1349. doi:10.1001/jama.1997.03550160069041

25. Institute of Medicine. *Cognitive Aging: Progress in Understanding and Opportunities for Action*. Washington, D.C.: National Academies Press; 2015. doi:10.17226/21693
26. Williams JW, Plassman BL, Burke J, Benjamin S. Preventing Alzheimer's disease and cognitive decline. *Evid Rep Technol Assess (Full Rep)*. 2010;(193):1-727.
27. Baumgart M, Snyder HM, Carrillo MC, Fazio S, Kim H, Johns H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's Dement*. 2015;11(6):718-726. doi:10.1016/j.jalz.2015.05.016
28. Cheng G, Huang C, Deng H, Wang H. Diabetes as a risk factor for dementia and mild cognitive impairment: a meta-analysis of longitudinal studies. *Intern Med J*. 2012;42(5):484-491. doi:10.1111/j.1445-5994.2012.02758.x
29. Pedditizi E, Peters R, Beckett N. The risk of overweight/obesity in mid-life and late life for the development of dementia: a systematic review and meta-analysis of longitudinal studies. *Age Ageing*. 2016;45(1):14-21. doi:10.1093/ageing/afv151
30. Kamer AR, Janal MN, de Leon M. Letter to the editor regarding: Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's Dement (Amsterdam, Netherlands)*. 2015;1(4):385-386. doi:10.1016/j.dadm.2015.08.003
31. Wu B, Fillenbaum GG, Plassman BL, Guo L. Association Between Oral Health and Cognitive Status: A Systematic Review. *J Am Geriatr Soc*. 2016;64(4):739-751. doi:10.1111/jgs.14036
32. Anstey KJ, von Sanden C, Salim A, O'Kearney R. Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. *Am J Epidemiol*. 2007;166(4):367-378. doi:10.1093/aje/kwm116
33. Napoli N, Shah K, Waters DL, Sinacore DR, Qualls C, Villareal DT. Effect of weight loss, exercise, or both on cognition and quality of life in obese older adults. *Am J Clin Nutr*. 2014;100(1):189-198. doi:10.3945/ajcn.113.082883
34. Lourida I, Soni M, Thompson-Coon J, et al. Mediterranean Diet, Cognitive Function, and Dementia. *Epidemiology*. 2013;24(4):479-489. doi:10.1097/EDE.0b013e3182944410

35. Zhao C, Noble JM, Marder K, Hartman JS, Gu Y, Scarmeas N. Dietary Patterns, Physical Activity, Sleep, and Risk for Dementia and Cognitive Decline. doi:10.1007/s13668-018-0247-9
36. Estruch R, Ros E, Salas-Salvadó J, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. *N Engl J Med*. 2018;378(25):e34. doi:10.1056/NEJMoa1800389
37. Muscari A, Giannoni C, Pierpaoli L, et al. Chronic endurance exercise training prevents aging-related cognitive decline in healthy older adults: a randomized controlled trial. *Int J Geriatr Psychiatry*. 2010;25(10):1055-1064. doi:10.1002/gps.2462
38. Sink KM, Espeland MA, Castro CM, et al. Effect of a 24-Month Physical Activity Intervention vs Health Education on Cognitive Outcomes in Sedentary Older Adults. *JAMA*. 2015;314(8):781. doi:10.1001/jama.2015.9617
39. Piazza-Gardner AK, Gaffud TJB, Barry AE. The impact of alcohol on Alzheimer's disease: A systematic review. *Aging Ment Health*. 2013;17(2):133-146. doi:10.1080/13607863.2012.742488
40. Sabia S, Fayosse A, Dumurgier J, et al. Alcohol consumption and risk of dementia: 23 year follow-up of Whitehall II cohort study. *BMJ*. 2018;362:k2927. doi:10.1136/bmj.k2927
41. Xu W, Wang H, Wan Y, et al. Alcohol consumption and dementia risk: a dose–response meta-analysis of prospective studies. *Eur J Epidemiol*. 2017;32(1):31-42. doi:10.1007/s10654-017-0225-3
42. Anstey KJ, Mack HA, Cherbuin N. Alcohol Consumption as a Risk Factor for Dementia and Cognitive Decline: Meta-Analysis of Prospective Studies. *Am J Geriatr Psychiatry*. 2009;17(7):542-555. doi:10.1097/JGP.0b013e3181a2fd07
43. Kruman II, Henderson GI, Bergeson SE. DNA damage and neurotoxicity of chronic alcohol abuse. *Exp Biol Med*. 2012;237(7):740-747. doi:10.1258/ebm.2012.011421
44. Rehm J, Hasan OSM, Black SE, Shield KD, Schwarzinger M. Alcohol use and dementia: a systematic scoping review. *Alzheimers Res Ther*. 2019;11(1):1. doi:10.1186/s13195-018-0453-0
45. Ilomaki J, Jokanovic N, Tan ECK, Lonnroos E. Alcohol Consumption, Dementia and Cognitive Decline: An Overview of Systematic Reviews. *Curr Clin Pharmacol*.

2015;10(3):204-212.

46. Zhao J, Stockwell T, Macdonald S. Non-response bias in alcohol and drug population surveys. *Drug Alcohol Rev*. 2009;28(6):648-657. doi:10.1111/j.1465-3362.2009.00077.x
47. Public Health Agency of Canada. What is the Population Health Approach? https://www.canada.ca/en/public-health/services/health-promotion/population-health/population-health-approach.html. Accessed August 28, 2019.
48. Public Health Agency of Canada. *A Dementia Strategy for Canada: Together We Aspire*.
49. Australian Health Ministers Advisory Council. *National Framework for Action on Dementia 2015-2019*. Canberra, Australia: Deparment of Health; 2015.
50. Assistant Secretary for Planning and Evaluation U.S. Department of Health and Human Services. *National Plans to Address Alzheimer's Disease*. Washington, D.C.; 2012.
51. Rose G. Sick Individuals and Sick Populations. *Int J Epidemiol*. 1985;14(1):32-38. doi:10.1093/ije/14.1.32
52. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427-432. doi:10.1093/ije/30.3.427
53. Manuel DG, Rosella LC, Hennessy D, Sanmartin C, Wilson K. Predictive risk algorithms in a population setting: an overview. *J Epidemiol Community Health*. 2012;66(10):859-865. doi:10.1136/jech-2012-200971
54. Harrell FE. *Regression Modeling Strategies with Applicaitons to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer; 2001.
55. Hu B, Palta M, Shao J. Properties of R(2) statistics for logistic regression. *Stat Med*. 2006;25(8):1383-1395. doi:10.1002/sim.2300
56. Nagelkerke NJD. A Note on a General Definition of the Coefficient of Determination. *Biometrika*. 1991;78(3):691-692.
57. Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol*. 1992;45(1):85-89.
58. Hou X, Feng L, Zhang can, cao X, Tan L, Yu J-T. Models for predicting risk of dementia: a systematic review. *J Neurol Neurosurg Psychiatry*. 2019;90:373-379. doi:10.1136/jnnp-2018-318212
59. Jungwirth S, Sonja AE, Ae Z, et al. Prediction of Alzheimer dementia with short neuropsychological instruments. doi:10.1007/s00702-009-0318-6

60. Mura T, Baramova M, Gabelle A, et al. Predicting dementia using socio-demographic characteristics and the Free and Cued Selective Reminding Test in the general population. *Alzheimers Res Ther*. 2017;9(1):21. doi:10.1186/s13195-016-0230-x
61. Walters K, Hardoon S, Petersen I, et al. Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Med*. 2016;14(1):1-12. doi:10.1186/s12916-016-0549-y

## CHATPER 2: Dementia Population Risk Tool (DemPoRT): Study protocol for a predictive algorithm assessing dementia risk in the community

### 2.0 PREFACE

This chapter presents the study protocol for the development of the DemPoRT prediction algorithm which was published in *BMJ Open* in 2017 as:

> Stacey Fisher, Amy Hsu, Nassim Mojaverian, Monica Taljaard, Gregory Huyer, Douglas G Manuel, Peter Tanuseputro. Dementia Population Risk Tool (DemPoRT): Study protocol for a predictive algorithm assessing dementia risk in the community. *BMJ Open* 2017; 7(10):e018018

All study information is in the primary text and in the table. The use of data in this project was authorized under section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a Research Ethics Board.

SF was responsible for the study design and protocol development, and drafted and revised the manuscript. NM contributed to the study design, protocol development and provided data/statistical support. AH, MT, and GH contributed to the design of the study and protocol development. PT and DGM was responsible for the conception of the project, grant application and contributed to the study design and protocol development. All authors provided critical reviews of the manuscript.

## 2.1 ABSTRACT

**Introduction:** The burden of disease from dementia is a growing global concern as incidence increases dramatically with age, and average life expectancy has been increasing around the world. Planning for an aging population requires reliable projections of dementia prevalence; however, existing population projections are simple and have poor predictive accuracy. The Dementia Population Risk Tool (DemPoRT) will predict incidence of dementia in the population setting using multivariable modeling techniques and will be used to project dementia prevalence.

**Methods and Analysis:** The derivation cohort will consist of elderly Ontario respondents of the Canadian Community Health Survey (CCHS) (2001, 2003, 2005, 2007; 18 764 males and 25 288 females). Pre-specified predictors include socio-demographic, general health, behavioral, functional and health condition variables. Incident dementia will be identified through individual linkage of survey respondents to population-level administrative health care databases (1 797 and 3 281 events, and 117 795 and 166 573 person-years of follow-up, for males and females, respectively until March 31, 2014). Using time of first dementia capture as the primary outcome and death as a competing risk, sex-specific proportional hazards regression models will be estimated. The 2008/2009 CCHS survey will be used for validation (approximately 4 600 males and 6 300 females). Overall calibration and discrimination will be assessed as well as calibration within predefined subgroups of importance to clinicians and policy-makers.

**Ethics and Dissemination:** Research ethics approval has been granted by the Ottawa Health Science Network Research Ethics Board. DemPoRT results will be submitted for publication in peer-review journals and presented at scientific meetings. The algorithm will be assessable online for both population and individual uses.

**Trial Registration Number:** ClinicalTrials.gov NCT03155815.

## 2.2 INTRODUCTION

The burden of disease from dementia is a growing global concern as incidence increases exponentially with age and average life expectancy has been increasing around the world[1,2]. Planning for an aging population requires reliable projections of future dementia prevalence and its implications on resource requirements. Existing population projections for dementia, however, are overly simplistic and likely inaccurate[3].

### 2.2.1 Limitations of current dementia projection methodology

Almost all existing dementia projections have used extrapolation and macrosimulation methods, which are simplistic and make assumptions that may not hold true into the future[3]. Most extrapolations simply apply current age- and sex-specific prevalence estimates of dementia to future population projections. Macrosimulations typically use estimates of dementia incidence and mortality, stratified by age and sex, to simulate disease prevalence as the population ages[1,4–6]. Projections from extrapolations incorrectly assume that the risk of mortality among those with and without dementia are equivalent[7,8], and both methods assume that the age and sex-specific prevalence of dementia risk factors will not change with time. The assumption of stable risk factor prevalence is widely thought to be the major source of error in existing dementia projections[3,9–11].

Changing trends of dementia risk factors has the potential to have a dramatic impact on dementia prevalence estimates, as up to 50% of dementia cases have been attributed to modifiable risk factors[9,12], and the prevalence of several factors has been projected to change significantly in the near future. For example, the population prevalence of diabetes and obesity in Canada has been projected to increase, while smoking, hypertension and dyslipidemia have been projected to

decline[13]. Consideration of risk factor prevalence is therefore important to improve the accuracy of dementia projections, and simple extrapolations and macrosimulations are often inadequate.

### 2.2.2 Predictive multivariable modeling of dementia incidence

Population-based predictive risk algorithms examine the effect of risk factors on dementia incidence and can be used for dementia burden projection. Population-based data that contain detailed risk factor information, such as health surveys, are linked at the individual-level to administrative data that capture dementia development. A multivariable model of dementia incidence is derived, validated against external data, and predictive performance is assessed. Once developed, the algorithm can be used to project disease incidence and prevalence. To obtain prevalence projections, the algorithm can be integrated into a microsimulation model such as Statistics Canada's Population Health Models (POHEM). POHEM dynamically models individual life trajectories of a population representative of Canada including births, deaths and migration, disease incidence and progression, and exposure to risk factors, facilitating detailed examination of the influence of changing risk factor prevalence on future dementia prevalence.

Predictive risk algorithms can also be used to describe the risk of dementia in the population, assess the contribution of specific risk factors to the population risk, identify high-risk groups, and evaluate risk reduction strategies.

### 2.2.3 Existing dementia prediction models

Many models have been developed to predict risk of dementia[14–26], most with the primary goal of identifying individuals in the clinical setting at high risk. They have varying discriminative ability (c-statistics ranging from 0.49[16] to 0.89[17]) and have generally been derived from small samples, rarely including more than a few thousand individuals. Existing models are therefore simplistic, including few predictors and rarely including interaction or non-linear terms, facilitating understanding and use by physicians in clinical practice, but limiting discriminatory ability and predictive accuracy. Walters et al[26] developed an algorithm for predicting 5-year dementia risk among individuals 60-79 years of age in the United Kingdom using an enormous derivation dataset of 800 000 individuals, and a simple model. The derivation model had a c-statistic of 0.84 (95% CI: 0.81, 0.87), but a low positive predictive value at most risk thresholds, and therefore is poor at identifying those at high risk of dementia. Additionally, as most dementia risk models are intended for use in the clinical setting, many include results from neuropsychological tests[17–23], MRI findings[18] and APOE genotype[18,24,25]. The inclusion of these variables, however, limits the application of these models as these variables are not available at the population level.

The objective of this study is to develop and validate the Dementia Population Risk Tool (DemPoRT) algorithm to predict dementia incidence in the population setting. This will be done using multivariable modelling techniques, linking self-reported risk factors captured by a population-based health survey in Canada with administrative databases across healthcare sectors that capture healthcare diagnosed dementia.

DemPoRT will be developed with a large population-based dataset and using only variables that are available at the population level, allowing for population-level application. DemPoRT will also utilize many methodological improvements over existing models. This protocol pre-specifies the predictor variables and analytic plan for model development, reducing the potential for overfitting and bias, and improving transparency. Interaction terms and flexible functions for continuous predictors will be investigated, increasing potential discriminative ability. The pre-specified analytic plan avoids data-driven variable selection procedures, further reducing the potential for bias.

To our knowledge, the DemPoRT predictive model will be the first algorithm designed to predict and project dementia incidence at the population level. It will be used to estimate the future burden of dementia using techniques that consider changes in risk factor prevalence and will identify modifiable risk factors that can be targeted by individuals, clinicians and policy makers to reduce the burden of dementia.

## 2.3 METHODS AND ANALYSIS

### 2.3.1 Study design

Two DemPoRT models, one for males and females, will be derived and validated using population-based data in Ontario, Canada, a multicultural province with 13.6 million residents. Predictors will be obtained from the Canadian Community Health Surveys (CCHS), and outcomes (i.e., diagnosis of dementia) will be obtained from routinely-collected health care data.

The derivation cohorts will consist of eligible respondents of the 2001, 2003, 2005 and 2007 CCHS (Cycles 1.1, 2.1, 3.1 and 4.1), while validation cohorts will consist of respondents to the 2008/2009 cycle. The CCHS is a national, cross-sectional survey developed by Statistics Canada to collect information related to health and health care utilization of the Canadian population. The survey has a multistage stratified cluster design that represents approximately 98% of the Canadian population aged 12 years and over and attained an average response rate of 79% over the study period. The CCHS is conducted through telephone and in-person interviews, and all responses are self-reported. The details of survey methodology have been published elsewhere[27]. Survey respondents will be excluded if they are less than 55 years of age at survey administration, self-reported a history of dementia, or are not eligible for Ontario's universal health insurance. If a respondent was included in more than one CCHS cycle, only their earliest survey response will be used.

### 2.3.2 Outcome

Survey respondents diagnosed with dementia will be identified through individual linkage to several population-based administrative databases at the Institute for Clinical Evaluative Sciences (ICES). Dementia case ascertainment is based on a validated definition: 1 hospital record OR 3 physician claim records at least 30 days apart within a 2-year period OR a dispensing record for a cholinesterase inhibitor from Ontario Drug Benefit (ODB). This definition has a sensitivity of 79.3% and a specificity of 99.1% when validated against emergency medical record (EMR) data[28]. Due to known underdiagnosis of dementia[29,30], we will supplement this definition by adding survey respondents with dementia codes captured on home care and long-term care assessments (dementia flag AND Cognitive Performance Scale [CPS]

score ≥2) using the Resident Assessment Instrument-Home Care (RAI-HC) database and the Continuing Care Reporting System (CCRS), respectively. We have found this addition adds substantially (approximately 18%) to the number of dementia cases captured.

Survey respondents with dementia will be excluded if they meet the criteria for dementia within 2 years of survey administration (to remove potentially prevalent cases) or are younger than 65 years of age at the time of dementia diagnosis (to exclude early onset dementia which likely has a different set of risk factors). Eligible survey respondents will be followed from the date of survey administration or age 65, whichever came later, until the earliest date of: dementia ascertainment, death (defined as competing risk), loss to follow-up (defined as loss of healthcare eligibility) or end of study (March 31, 2014).

### 2.3.3 Sample Size

The male and female derivation cohorts consist of 18 764 and 25 288 respondents, and 117 795 and 166 573 person-years of follow-up, respectively. For predictive models with time to event outcomes the number of participants experiencing the event should exceed 10 times the number of degrees of freedom to ensure adequate sample size[31]. The number of dementia events in the derivation cohort is 1 797 for men and 3 281 for women; therefore, the maximum number of total degrees of freedom for each of the DemPoRT models is 179 and 328, respectively, which we do not anticipate surpassing.

The validation cohorts will consist of approximately 4 600 males and 6 300 females, and 15 000 and 21 000 person-years of follow-up, respectively. Vergouwe *et al*[32] recommend a minimum of

100 events and 100 non-events for external validation studies. We expect approximately 225 events for men and 400 for women in our validation cohort.

### 2.3.4 Analysis Plan

The analysis plan was developed following guidelines by Harrell[31] and Steyerberg[33] after accessing the derivation data set, but prior to model fitting or descriptive analyses involving exposure-outcome associations. This was done to avoid Type 1 error introduced by data-driven variable selection or model specification. Key considerations of our analysis approach include full pre-specification of the predictor variables, use of flexible functions for continuous predictors, and preserving statistical properties by avoiding data-driven variable selection procedures. Analysis will be conducted using Harrell's Hmisc[34] package of functions in R[35] as well as SAS v9.4.

This study protocol and the reporting of our model estimation results will be guided by the TRIPOD statement for multivariable predictive models[36].

*Identification of Predictors*

Predictor variables were identified through review of existing predictive algorithms for dementia[9,16,18–22,24–26,37,38] and comparison to available data collected in the CCHS. Variable inclusion was informed by consultation with subject-matter experts and the project's advisory team, and informed by our previous work developing predictive models for cardiovascular disease and life expectancy[39,40].

Variables with narrow distributions or insufficient variation were excluded. Obvious cases of redundancy (e.g. alternate definitions of the same underlying behaviour) were not included. A total of 32 predictor variables were identified: 7 sociodemographic, 3 general health, 9 behavioural, 7 functional, 5 health conditions and 1 design variable (CCHS survey cycle). As the effect of dementia risk factors varies by sex, separate models will be derived for men and women. Education, rather than individual income, was selected as a predictor due to several concerns with income including lack of generalizability, measurement error, stability over time and substantial missing values. Indicator variables for smoking status were created to allow the inclusion of smoking pack-years as a continuous predictor. The models will additionally include age interactions with the behavioural, functional and health condition variables as the effect of these risk factors on dementia are expected to vary with age. Detailed definitions and measurement of the predictor variables are presented in **Table 2-1**.

*Data Cleaning and Coding of Predictors*

Continuous variables will be inspected using boxplots and descriptive statistics to determine values outside a plausible range. Values that are clearly erroneous will be corrected, where possible, or set to missing. Continuous predictors with highly skewed distributions will be truncated to the $99.5^{th}$ percentile. Categorization of continuous variables will be avoided to minimize the loss of predictive information. All data cleaning and coding will occur prior to examining exposure-outcome associations.

**Table 2-1.** Pre-specification of predictor variables for DemPoRT with initial degrees of freedom (df) allocation

| Variable | Scale | Initial Variable Specification | df |
|---|---|---|---|
| **Socio-demographic Factors** | | | |
| Age | Continuous | **5 knot spline:** Valid range: 55-102 (male), 55-101 (female) | 4 |
| Sex | Categorical | **Stratified:** Male; Female | NA |
| Ethnicity | Categorical | **7 categories:** Caucasian; African-American; Chinese; Aboriginal; Japanese/Korean/South East Asian/Filipino; Other/Multiple origin/Unknown/Latin American; South Asian/Arab/West Asian | 6 |
| Immigrant | Dichotomous | Yes; No | 1 |
| Education | Categorical | **4 categories:** Less than secondary school; Secondary school graduation; Some postsecondary; Postsecondary graduation | 3 |
| Marital Status | Categorical | **4 categories:** Now married/Common-law; Separated/Divorced; Widowed; Single | 3 |
| Neighborhood Social and Material Deprivation (Pampalon et al. 2009) | Ordinal | **3 categories:** Low (1st or 2nd quintile); High 4th or 5th quintile; Moderate (3rd quintile) | 2 |
| **General Health** | | | |
| Sense of belonging to local community | Ordinal | **4 categories:** Very strong; Somewhat strong; Somewhat weak; Very weak | 3 |
| Self-perceived stress | Ordinal | **5 categories:** Not at all stressful; Not very stressful; A bit stressful; Quite a bit stressful; Extremely stressful | 4 |
| Self-rated health | Ordinal | **5 categories:** Poor; Fair; Good; Very Good; Excellent | 4 |
| **Health Behaviors** | | | |
| Pack years of smoking | Continuous | **3 knot spline:** Valid range: 0-112 (male), 0-78 (female) | 2 |
| Smoking status | Categorical | **4 categories:** Non-smoker; Current smoker; Former smoker quit <5 years ago; Former smoker quit >5 years ago | 3 |
| Alcohol consumption (number of drinks last week) | Continuous | **3 knot spline:** Valid range: 0-50 (male), 0-24 (female) | 2 |
| Former drinker | Dichotomous | Yes; No | 1 |
| Consumption of fruit, salad, carrot and other vegetables (average daily frequency) | Continuous | **3 knot spline:** Valid range: 0-48 (male), 0-31 (female) | 2 |
| Potato consumption (average daily frequency) | Continuous | **3 knot spline:** Valid range: 0-2 | 2 |
| Juice consumption (average daily consumption | Continuous | **3 knot spline:** Valid range: 0-6 (male), 0-5 (female) | 2 |
| Leisure physical activity (average daily METs (kcal/kg/day)) | Continuous | **3 knot spline:** Valid range: 0-16 (male), 0-12 (female) | 2 |
| **Functional Measures** | | | |
| Personal hygiene and care | Dichotomous | Does not need help; Needs help | 1 |
| Locomotion in the home | Dichotomous | Does not need help; Needs help | 1 |
| Meal preparation | Dichotomous | Does not need help; Needs help | 1 |
| Running errands | Dichotomous | Does not need help; Needs help | 1 |
| Ordinary housework | Dichotomous | Does not need help; Needs help | 1 |
| Heavy housework | Dichotomous | Does not need help; Needs help | 1 |
| Finances | Dichotomous | Does not need help; Needs help | 1 |
| **Health Conditions** | | | |
| Heart disease | Dichotomous | Yes; No | 1 |
| Stroke | Dichotomous | Yes; No | 1 |
| Diabetes | Dichotomous | Yes; No | 1 |
| Mood disorder | Dichotomous | Yes; No | 1 |
| High blood pressure | Dichotomous | Yes; No | 1 |
| Body mass index | Continuous | **3 knot spline:** Valid range: 10-44 (male), 10-47 (female) | 2 |
| **Design** | | | |
| Survey year | Ordinal | **4 categories:** 2000/01, 2002/03, 2004/05, 2006/07 | 3 |

DemPoRT, Dementia Population Risk Tool; df, degrees of freedom; MET, metabolic equivalent task

*Missing Data*

As traditional complete cases analyses suffer from inefficiency, selection bias, and other limitations[33], multiple imputation methods will be used to impute missing values using the 'aregImpute' function in the HMisc library[34]. This function simultaneously imputes missing values using predictive mean matching and uses bootstrapping to take all aspects of uncertainty into account. The imputation model will consist of the full list of predictor variables, time to event and censoring variables, as well as auxiliary variables that are not predictors, but may nevertheless be useful in generating imputed values (e.g., income). The final model will be estimated in each of five multiple imputation data sets and the results combined using the rules developed by Rubin and Schenker[41] to account for imputation uncertainty.

*Model Specification*

Initial sex-specific main effects models will be fit using the pre-specified predictors and an initial degree of freedom allocation for each predictor (Table 1). Decisions on initial degree of freedom allocations were informed by the anticipated importance of each predictor and known dose-response relationships with dementia. Continuous predictors will be flexibly modelled using restricted cubic splines, with the knots placed at fixed quantiles of the distribution (e.g., $5^{th}$, $27.5^{th}$, $50^{th}$, $72.5^{th}$, and $95^{th}$ centiles). Frequency distributions for categorical predictors will be examined and categories with small numbers of respondents will be combined, with analysts blinded to the number of events per category, to avoid instability in the regression analyses. Ordinal variables will be specified as either linear terms or as categorical if the expected association is more complex. Interactions will be restricted to linear terms. The initial model

specification, presented in Table 1, includes a total of 86 degrees of freedom (63 main, 23 interaction).

Partial association chi-square statistics for each predictor minus their degrees of freedom (to level the playing field among predictors with varying degrees of freedom) will be plotted in descending order. Variables with higher predictive potential will retain their initial degrees of freedom, while predictors with lower predictive potential will be modeled as simple linear terms or recoded by combining infrequent categories. This process of model specification does not increase the Type I error rate because all predictors will be retained in the full model regardless of their strength of association[31].

*Model Estimation*

The initial models will be estimated using competing risk Cox proportional hazards regression with time to dementia ascertainment as the outcome and death as a competing risk. Alternative model specifications, including subdistribution hazard and flexible parametric models, will be considered. All predictors will be centered about their means. A formal check of multicollinearity will be carried out using a variable clustering algorithm[31].

Proportional hazards models assume that the relative risk of the outcome between strata of predictors and the baseline risk must be constant over time. Violation of this assumption has been shown to produce biased results[42] although it has also been argued that the estimated coefficients of time-varying variables can simply be interpreted as an average rather than instantaneous hazard[43]. Plots of raw and smoothed scaled Schoenfeld residuals versus time for

each predictor will be assessed to test this assumption and identify non-proportionality. If a violation of this assumption is identified, we will consider addition of interaction terms between the predictor and log-transformed time.

Although the risk of overfitting will be minimal due to pre-specification of the models and a large sample size, the need for overfitting adjustment will be assessed. The degree of overfitting will be estimated using the heuristic shrinkage estimator, based on the log likelihood ratio chi-square statistic for the full model[44]. If shrinkage is <0.90, models will be adjusted for overfitting.

*Estimation of the Reduced Models*

Model pre-specification has advantages in limiting overfitting and spurious statistical significance but can result in a final model that is overly complex, difficult to interpret, and difficult to apply. Unnecessary predictor variables also distort the estimated effects of other predictors making the model more computationally intensive. It is suggested that a more parsimonious model that retains most of the prognostic information and performs as well as or better than the full model can be derived without increasing the Type 1 error rate[31,45]. We will identify a more parsimonious model using a stepdown procedure described by Ambler[45], which involves deleting the variable that results in the smallest decrease in model $R^2$ until removal leads to an $R^2$ that is less than a desired level. The reduced model will be evaluated against the full model using Akaike's Information Criterion, and by examining the effect on discrimination and calibration.

DemPoRT will be developed and validated using temporal split samples, however the final regression coefficients will use the full data set to maximize follow-up duration. A cohort-specific intercept and/or interaction term may be included in the final model if the derivation and validation cohorts differ; otherwise, the final combined model will maintain the same predictors and form as the derivation model.

*Assessment of Predictive Performance*

Predictive performance in the derivation and validation cohorts will be assessed and reported using overall measures of predictive accuracy, discrimination and calibration. Accuracy will be assessed with Nagelkerke's $R^2$ [46] and the Brier score[47]. Discrimination will be assessed using the concordance statistic. Model calibration is especially important in the development of prognostic models, as probabilities of future risk are of primary interest[33,48,49]. Calibration will be assessed by comparing the observed and predicted risk of dementia within vigintiles (20 groups of equal frequency) of predicted risk with emphasis on visual inspection of plots rather than formal statistical significance testing, which can be influenced by large sample sizes[32]. Calibration slopes will be generated by regressing the outcome in the validation cohort on the predicted dementia risk, reflecting the combined effect of overfitting to the derivation data as well as true differences in effects of predictors. Deviation of the slope from 1 (perfect calibration) will be tested using a Wald or likelihood ratio test. Calibration within predefined subgroups of importance to clinicians and policy makers (e.g., age group, health behaviour, socio-demographic groups and health conditions) will additionally be evaluated. The clinically relevant standard of calibration was defined as less than 20% difference between observed and predicted

estimates within subgroups with a dementia prevalence of at least 5%. All model performance measures will be calculated using the first of the multiply imputed data sets.

*Model Presentation*

The final regression model, derived from the combined sample of the derivation and validation cohorts, will be presented using estimated hazard ratios and 95% confidence intervals, along with results for the derivation and validation cohorts separately. We have found, however, this usual presentation less meaningful when presenting complex models[39]. To allow interpretation of the estimated effect of each predictor, model behavior will additionally be described using interactive visual tools to display the shape of the effect of each predictor[50]. The regression formula will also be published and used as the basis for web-based implementation.

### 2.3.5 Analyses Beyond Initial Model Development

We will conduct further analyses exploring the added predictive ability of novel risk factors that were ascertained in single CCHS cycles (e.g. sedentary activity, cognitive stimulation, sleep quality and duration, deafness), as well as risk factors that can be ascertained through linkage of additional data sources and similar cohorts (e.g. air pollution, detailed dietary consumption, lipid levels, blood pressure). In addition, sensitivity analysis of the age at survey administration cutoff used for cohort creation will be performed.

Once developed, DemPoRT will be used to project dementia incidence under different assumptions by entering counterfactual risk factor levels into the algorithm at the population

level, or at individual level and summed, and will be integrated in to POHEM for microsimulation modelling of prevalence projections.

A second, causal model (DemPoRT-C) will also be created to assess the relative contribution of lifestyle, socio-demographic and health factors to dementia incidence. Development will exclude variables believed to be in the causal pathway of dementia occurrence (e.g., self-rated health and functional measures) to reduce the attenuation of hazards from upstream risk factors but will otherwise be the same as in the predictive model. DemPoRT-C will be applied to the most recent unlinked national CCHS survey.

### 2.4 LIMITATIONS

One of the limitations of this study will be the potential for misclassification error resulting from the use of self-reported predictors captured at one point in time and administrative data for outcome ascertainment. However, discriminating and well-calibrated algorithms have been developed using self-report information and although detailed cognitive testing to ascertain dementia diagnoses is preferable over the use of administrative data, it is not available or feasible at the population level. Another concern common to the development of highly complex risk algorithms, such as DemPoRT, is the potential for statistical overfitting and increased type 1 error, which can occur when the relationship between a predictor and the outcome influences whether it is used and how it is fit. This risk is reduced by pre-specification of the predictors and analytic plan, as we have done in this protocol. The model will also be adjusted for overfitting if necessary, as specified previously. Lastly, although a rigorous approach to model development

will be used, further validation will be needed to assess generalizability, and calibration will be required for application in other jurisdictions.

## 2.5 ETHICS AND MODEL DISSEMINATION

The DemPoRT project advisory committee has been created to ensure that the models meet the needs of knowledge users. This committee has worked with the study team to identify predictors of dementia based on scientific and policy importance and will aid in the identification of important target populations and the establishment of policy-relevant differences for calibration studies.

DemPoRT results will be submitted for publication in peer-review journals and presented at scientific meetings. A web-based individual-level calculator will be created if the models are appropriate for individual use. Although DemPoRT emphasizes risk prediction at the population level, we have found that individual-level calculators are an effective engagement and translation tool for both the general public and knowledge users.

## 2.6 CONCLUSIONS

To the best of our knowledge, DemPoRT will be the first population-based algorithm designed to predict and project dementia incidence at the population level. The DemPoRT models will produce estimates of future dementia burden that we believe will be more accurate than existing estimates, will assess the contribution of specific risk factors to the population risk, and will identify groups at high risk of developing dementia.

## 2.7 FUNDING

This work is supported by the Canadian Institutes of Health Research operating grant MOP 14223, and the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). SF is supported by a doctoral award from the Canadian Institutes of Health Research. The sponsors have no role in the design or conduct of the study; in the collection, analysis or interpretation of data; or in the preparation, review or approval of the manuscript. The opinions, results and conclusions reported are those of the authors and are independent from funding sources. No endorsement by CIHR, ICES or the Ontario MOHLTC is intended or should be inferred.

## 2.8 REFERNCES


1. Brookmeyer, R., Gray, S. & Kawas, C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am. J. Public Health* **88,** 1337–1342 (1998).

2. Brayne, C. The elephant in the room - healthy brains in later life, epidemiology and public health. *Nat. Rev. Neurosci.* **8,** 233–9 (2007).

3. Norton, S., Matthews, F. E. & Brayne, C. A commentary on studies presenting projections of the future prevalence of dementia. *BMC Public Health* **13,** 1 (2013).

4. Sloane, P. D. *et al.* The public health impact of Alzheimer's disease, 2000-2050: potential implication of treatment advances. *Annu. Rev. Public Health* **23,** 213–31 (2002).

5. Mura, T., Dartigues, J. F. & Berr, C. How many dementia cases in France and Europe? Alternative projections and scenarios 2010-2050. *Eur. J. Neurol.* **17,** 252–259 (2010).

6. Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A. & Evans, D. A. Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch. Neurol.* **60,** 1119–22 (2003).

7. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement.* **3,** 186–191 (2007).

8. Dewey, M. E. & Chen, C.-M. Neurosis and mortality in persons aged 65 and over living in the community: a systematic review of the literature. *Int. J. Geriatr. Psychiatry* **19,** 554–7 (2004).

9. Norton, S., Matthews, F. E., Barnes, D. E., Yaffe, K. & Brayne, C. Potential for primary prevention of Alzheimer's disease: An analysis of population-based data. *Lancet Neurol.* **13,** 788–794 (2014).

10. Joly, P. *et al.* Prevalence projections of chronic diseases and impact of public health intervention. *Biometrics* **69,** 109–17 (2013).

11. Lee, Y. The recent decline in prevalence of dementia in developed countries: implications for prevention in the Republic of Korea. *J. Korean Med. Sci.* **29,** 913–8 (2014).

12. Barnes, D. E. & Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet. Neurol.* **10,** 819–28 (2011).

13. Manuel, D. G. *et al.* Projections of preventable risks for cardiovascular disease in Canada to 2021: a microsimulation modelling approach. *C. Open* **2,** E94–E101 (2014).

14. Tang, E. Y. H. *et al.* Current developments in dementia risk prediction modelling: An updated systematic review. *PLoS One* **10,** 1–31 (2015).

15. Stephan, B. C. M., Kurth, T., Matthews, F. E., Brayne, C. & Dufouil, C. Dementia risk prediction in the population: are screening models accurate? *Nat. Rev. Neurol.* **6,** 318–326 (2010).

16. Anstey, K. J. *et al.* A self-report risk index to predict occurrence of dementia in three independent cohorts of older adults: The ANU-ADRI. *PLoS One* **9,** (2014).

17. Wolfsgruber, S. *et al.* The CERAD neuropsychological assessment battery total score detects and predicts alzheimer disease dementia with high diagnostic accuracy. *Am. J. Geriatr. Psychiatry* **22,** 1017–1028 (2014).

18. Barnes, D. E. *et al.* Predicting risk of dementia in older adults: The late-life dementia risk index. *Neurology* **73,** 173–179 (2009).

19. Chary, E. *et al.* Short- versus long-term prediction of dementia among subjects with low and high educational levels. *Alzheimers. Dement.* **9,** 562–71 (2013).

20. Jessen, F. *et al.* Prediction of Dementia in Primary Care Patients. *PLoS One* **6,** e16852 (2011).

21. Song, X., Mitnitski, A. & Rockwood, K. Nontraditional risk factors combine to predict Alzheimer disease and dementia. *Neurology* **77,** 227–234 (2011).

22. Tierney, M. C., Moineddin, R. & McDowell, I. Prediction of all-cause dementia using neuropsychological tests within 10 and 5 years of diagnosis in a community-based sample. *J. Alzheimer's Dis.* **22,** 1231–1240 (2010).

23. Meng, X., D'Arcy, C., Morgan, D. & Mousseau, D. D. Predicting the risk of dementia among Canadian seniors: a useable practice-friendly diagnostic algorithm. *Alzheimer Dis. Assoc. Disord.* **27,** 23–9 (2013).

24. Kivipelto, M. *et al.* Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol.* **5,** 735–741 (2006).

25. Reitz, C. *et al.* A Summary Risk Score for the Prediction of Alzheimer Disease in Elderly Persons. *Arch. Neurol.* **67,** 835–41 (2010).

26. Walters, K. *et al.* Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Med.* **14,** 1–12 (2016).

27. Béland, Y. Canadian community health survey--methodological overview. *Heal. reports / Stat. Canada, Can. Cent. Heal. Inf. = Rapp. sur la sant?? / Stat. Canada, Cent. Can. d'information sur la sant??* **13,** 9–14 (2002).

28. Jaakkimainen, R. L. *et al.* Identification of Physician-Diagnosed Alzheimer's Disease and

Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J. Alzheimer's Dis.* **54,** 337–349 (2016).

29. Connolly, A., Gaehl, E., Martin, H., Morris, J. & Purandare, N. Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence. *Aging Ment. Health* **15,** 978–984 (2011).

30. Kosteniuk, J. G. *et al.* Incidence and prevalence of dementia in linked administrative health data in Saskatchewan, Canada: a retrospective cohort study. *BMC Geriatr.* **15,** 73 (2015).

31. Harrell, F. E. *Regression Modeling Strategies with applicaitons to linear models, logistic regression and survival analysis*. (Springer, 2001).

32. Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J. C. & Habbema, J. D. F. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J. Clin. Epidemiol.* **58,** 475–483 (2005).

33. Steyerberg, E. W. *Clinical Prediction Models*. (Springer, 2009).

34. Harrell, F. E. *Hmisc: Harrell Miscellaneous. R package version 4.0-2*. (2016).

35. R Core Team. R: A language and environment for statistical computing. (2016).

36. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur. Urol.* **67,** 1142–1151 (2015).

37. Exalto, L. G. *et al.* Midlife risk score for the prediction of dementia four decades later. *Alzheimer's Dement.* **10,** 562–570 (2014).

38. Exalto, L. G. *et al.* Risk score for prediction of 10 year dementia risk in individuals with type 2 diabetes: A cohort study. *Lancet Diabetes Endocrinol.* **1,** 183–190 (2013).

39. Taljaard, M. *et al.* Cardiovascular Disease Population Risk Tool (CVDPoRT): predictive algorithm for assessing CVD risk in the community setting. A study protocol. *BMJ Open* **4,** e006701 (2014).

40. Manuel, D. G. *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.* **13,** 1–27 (2016).

41. Rubin, D. B. & Schenker, N. Multiple imputation in health-care databases: an overview and some applications. *Stat. Med.* **10,** 585–98 (1991).

42. Therneau, T. M., Grambsch, P. M. & Fleming, T. R. Martingale-based residuals for survival models. *Biometrika* **77,** 147–160 (1990).

43. Allison, P. D. *Survival analysis using SAS : A practical guide*. (SAS Institute, 2010).

44. Van Houwelingen, J. C. & Le, C. S. Predictive value of statistical models. *Stat Med* **9,** 1303–1325 (1990).

45. Ambler, G., Brady, A. R. & Royston, P. Simplifying a prognostic model: A simulation study based on clinical data. *Stat. Med.* **21,** 3803–3822 (2002).

46. Nagelkerke, N. J. D. A Note on a General Definition of the Coefficient of Determination. *Biometrika* **78,** 691–692 (1991).

47. Arkes, H. R. *et al.* The covariance decomposition of the probability score and its use in evaluating prognostic estimates. SUPPORT Investigators. *Med. Decis. Making* **15,** 120–31 (1995).

48. Cook, N. R. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin. Chem.* **54,** 17–23 (2008).

49. Cook, N. R. Comment: Measures to summarize and compare the predictive capacity of markers. *Int. J. Biostat.* **6,** Article 22; discussion Article 25 (2010).

50. Krause, J., Perer, A. & Bertini, E. Using Visual Analytics to Interpret Predictive Machine Learning Models. *arXiv Prepr. arXiv1606.05685.* (2016).

51. Pampalon, R., Hamel, D., Gamache, P. & Raymond, G. A deprivation index for health planning in Canada. *Chronic Dis. Can.* **29,** 178–91 (2009).

# CHAPTER 3: Development and validation of a predictive algorithm for risk of dementia in the community setting: The Dementia Population Risk Tool (DemPoRT)

## 3.0 PREFACE

This chapter presents the development and validation of the DemPoRT prediction algorithm, and is currently under review for publication with the following author list:

> Stacey Fisher, Douglas G Manuel, Carol Bennett, Meltem Tuna, Anan Bader Eddeen, Mahsa Jessri, Amy Hsu, Yulric Sequeira, Monica Taljaard, Geoffrey Anderson, Peter Tanuseputro

Supporting information is provided in appendices at the end of the thesis (Section 7.1) and on GitHub at https://github.com/Big-Life-Lab/predictive-algorithms. The use of data in this project was authorized under section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a Research Ethics Board.

SF was responsible for the study design, protocol development, data analysis, interpretation of the results, and drafting and revision of the manuscript. DGM and PT were responsible for conception of the project, grant application, and contributed to the study design, protocol development and result interpretation. MTuna, ABE and MTaljaard contributed to the study design, protocol development and result interpretation, and provided data/statistical support. CB, MJ and AH contributed to the design of the study, protocol development, and result interpretation. YS provided statistical support and is primarily responsible for the online web calculator, visualization tool and application programming interface. GA contributed to result interpretation. All authors provided critical reviews of the manuscript.

## 3.1 ABSTRACT

**Objectives:** Most existing dementia algorithms require neuropsychological testing, limiting their use to within the clinical setting. A predictive algorithm for estimate five-year dementia risk in the community setting was developed using only self-reported risk factors.

**Methods:** The Dementia Population Risk Tool (DemPoRT) was developed using Ontario respondents to the Canadian Community Health Survey (survey years 2001 to 2012). Five-year incidence of physician-diagnosed dementia was ascertained by individual linkage to administrative health care databases and using a validated case ascertainment definition with follow-up to March 2017. Sex-specific proportional hazards regression models considering competing risk of death were developed using self-reported risk factors including information on socio-demographic characteristics, general and chronic health conditions, health behaviors and physical function.

**Findings:** Among 75,460 respondents included in the combined derivation and validation cohorts, there were 8,448 cases of incident dementia in 348,677 person-years of follow-up (5-year cumulative incidence, men: 0.044, 95% CI: 0.042, 0.047; women: 0.057, 95% CI: 0.055, 0.060). The final full models each include 90 degrees of freedom (65 main effects, 25 interaction) and 28 predictors (8 continuous). The DemPoRT algorithm is discriminating (C-statistic in validation data: males 0.83 (95% CI: 0.81, 0.85); females 0.82 (95% CI: 0.81, 0.84)) and well-calibrated in a wide range of subgroups including behavioral risk exposure categories, sociodemographic groups, and chronic conditions.

**Conclusions:** This algorithm will support decision-making for population health including the development and evaluation of population-level dementia prevention strategies and can be used

by individuals or their clinicians for individual risk assessment. **Study registration:** ClinicalTrials.gov, no.NCT03155815.

## 3.2 INTRODUCTION

An estimated 50 million people worldwide have dementia, which is expected to grow to over 152 million by 2050[1], putting a tremendous strain on our health care systems, caregivers and families. As there is no cure and no disease-modifying therapies, primary prevention of dementia has become an important component of health care planning and policy development, and has been identified as a primary objective of many international dementia strategies[2–5]. The development of population prevention strategies requires tools that can predict who in a population will develop dementia in the future and evaluate how projection estimates may change with risk factor modification.

Most existing dementia algorithms are designed for use in the clinical setting and require neuropsychological assessment[6–13], genetic testing[10–14], neuroimaging[10] or other clinical variables (e.g. blood pressure or cholesterol values)[14–17], making them unsuitable for population-based dementia assessment due to the unavailability of this information at the population-level to policy and decision makers. Other algorithms have been developed using small samples, limiting their generalizability, or did not include modifiable risk factors of public health interest. Therefore, few existing dementia algorithms are suited for preventive population health planning.

In this study we developed and validated the Dementia Population Risk Tool (DemPoRT) for predicting dementia incidence in the community setting using self-reported risk factors and a population-based cohort. DemPoRT can be used at the population-level to inform the development of dementia prevention strategies and support decision-making for population

health. In addition to this population health planning purpose, DemPoRT can also be used by patients or their clinicians to assess individual dementia risk.

## 3.3 METHODS

### 3.3.1 Study design and participants

This prospective cohort study used linked population health survey data and health administrative data to develop and validate a population risk algorithm, DemPoRT, for predicting 5-year dementia incidence in the community setting. There were 4 pre-specified steps:

1. Model derivation – creation of male and female DemPoRT risk algorithms using respondents to the 2001, 2003, 2005 and 2007/08 Canadian Community Health Surveys (CCHS);
2. Model validation – validation of the DemPoRT algorithms using the 2009/10 and 2011/12 CCHS;
3. Final model generation – estimation of final, full DemPoRT models using combined derivation and validation data and the same model specification as the derivation models;
4. Derivation of the application model – creation of a parsimonious model with fewer predictors that attempts to maintain discriminatory ability, calibration and overall model performance.

The protocol for development and validation of DemPoRT was registered and published (ClinicalTrials.gov, no. NCT03155815)[18]. We adhered to the protocol with the following exceptions: the validated dementia definition was not supplemented with dementia information from home care and long-term care data; individuals had to be age 55 or older at the time of

survey administration; follow-up was extended to March 2017 with the availability of new data; ethnicity and functional measures were recategorized due to small sample sizes; variables for multilingualism, chronic obstructive pulmonary disorder (COPD) and epilepsy were added; and only the first of the multiply imputed datasets was used for model analyses, informed by previous work with this data[19]. This paper adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) checklist for prediction model development[20] (**Appendix 3-1**).

Survey respondents who agreed to share and link their survey interview information were eligible for study inclusion. Respondents were excluded if they were not eligible for Ontario's universal health insurance program, indicated a diagnosis of dementia in the CCHS, or were younger than age 55 at the time of survey administration. For individuals with multiple CCHS interviews, only the earliest interview was included.

### 3.3.2 Data sources

The CCHS is a national, cross-sectional survey developed by Statistics Canada to collect data related to health determinants, health status and health care use[21]. It employs a complex multistage sampling strategy to randomly select households in each region, with a target population of individuals aged 12 years and older. Over the study period, the surveys attained an average response rate of 79%. Individuals living on First Nation Reserves, institutionalized residents, full-time members of the Canadian Forces and residents of certain remote areas are excluded. Details of the survey methodology have been previously published[21]. Model predictors were ascertained from self-reported responses to CCHS. The derivation cohort was comprised of

Ontario respondents to the CCHS conducted in 2001, 2003, 2005 and 2007/08. The validation cohort consisted of Ontario CCHS respondents from 2009/10 and 2011/12. Temporal validation was used as it is a stronger validation approach than random creation of development and validation data sets, and because this tool will be used for prediction of future dementia risk[22].

Dementia incidence was ascertained using population-based datasets housed at ICES (formerly known as the Institute for Clinical Evaluative Sciences), which have been linked at the individual level to CCHS respondents. These datasets include hospital-admission records from the Canadian Institute for Health Information Discharge Abstract Database, physician billing and diagnoses from the Ontario Health Insurance Plan physician claims database, hospital and community-based ambulatory care from the National Ambulatory Care Reporting System, and drug dispensing data from the Ontario Drug Benefit program. Death was ascertained using Ontario Vital Statistics and the Registered Persons Database.

### 3.3.3 Outcome

The primary outcome of interest was 5-year incidence of physician-diagnosed dementia, ascertained using a validated case ascertainment definition: one hospital record OR three physician claims records at least 30 days apart within a 2-year period OR a dispensing record for a cholinesterase inhibitor[23]. This definition has a 79.3% sensitivity and a 99.1% specificity when validated against emergency medical record data. Survey respondents were followed from survey administration date until the earliest of dementia ascertainment, death, loss to follow-up (defined as loss of health care eligibility), or end of study (March 31, 2017).

### 3.3.4 Statistical Methods and Analyses

The analysis plan was developed following guidelines by Harrell[24] and Steyerberg[25] and informed by the development of other algorithms by the team[19,26].

Predictor identification, data cleaning, missing data and model specification methods have previously been described in the study protocol[18]. Predictor identification and specification as well as all data cleaning and coding occurred prior to examining exposure-outcome associations. Preliminary sex-specific main effects models were fit using the pre-specified predictors and degree of freedom allocation. Partial association chi-square statistics for each predictor minus their degrees of freedom were plotted to inform degree of freedom reduction, however, all initial degrees of freedom were retained. As all multiply imputed datasets showed the same rank-order associations, only the first dataset was used for all further analyses. **Table 3-1** presents the 33 pre-specified predictor variables and the final 29 variables in the full model, after recategorization of ethnicity and functional measures. The models include interaction between age and all variables except socio-demographic, general health and survey year variables, with continuous variable interactions restricted to linear terms. Survey questions used to ascertain the model variables are available at https://github.com/Big-Life-Lab/predictive-algorithms.

Models were estimated using Fine and Gray subdistribution hazard models[27], considering death as a competing risk. All predictors were centered on their means for ease of re-calibration in new populations and to allow for application in individuals or settings with missing values. Proportional hazard assumptions were tested using plots of raw and smoothed scaled Schoenfeld

**Table 3-1.** Predictor variables for the Dementia Population Risk Tool (DemPoRT) models

| Variable | Scale | Initial Variable Specification | Full Model | Reduced Model |
|---|---|---|---|---|
| **Degrees of Freedom** | | | | |
| Male model | - | 88 (65 main, 23 interaction) | 90 (65 main, 25 interaction) | 74 (51 main, 23 interaction) |
| Female model | - | 88 (65 main, 23 interaction) | 90 (65 main, 25 interaction) | 74 (50 main, 24 interaction) |
| **Socio-demographic Factors** | | | | |
| Age* | Continuous | **5 knot spline:** Valid range: 55-102 (male), 55-101 (female) | Unchanged | Unchanged |
| Sex | Categorical | **Stratified:** Male; Female | Unchanged | Unchanged |
| Ethnicity | Categorical | **7 categories:** White; Black; Chinese; Aboriginal; Japanese/Korean/South East Asian/Filipino; Other/Multiple origin/Unknown/Latin American; South Asian/Arab/West Asian | **4 categories:** White; South Asian/Arab/West Asian; South East Asian/Chinese/Japanese/ Korean/Filipino; Other/Multiple origin/Unknown | Excluded (female) |
| Immigrant | Dichotomous | Yes; No | Unchanged | Excluded |
| Education | Categorical | **4 categories:** Less than secondary school; Secondary school graduation; Some postsecondary; Postsecondary graduation | Unchanged | Excluded (male) |
| Marital status | Categorical | **4 categories:** Now married/Common-law; Separated/Divorced; Widowed; Single | Unchanged | Excluded (male) |
| Neighborhood deprivation | Ordinal | **3 categories:** Low (1st or 2nd quintile); High 4th or 5th quintile; Moderate (3rd quintile) | Unchanged | Excluded |
| Multilingualism | Dichotomous | - | Multilingual; Not multilingual | Excluded (female) |
| **General Health** | | | | |
| Sense of belonging | Ordinal | **4 categories:** Very strong; Somewhat strong; Somewhat weak; Very weak | Unchanged | Excluded |
| Stress | Ordinal | **5 categories:** Not at all stressful; Not very stressful; A bit stressful; Quite a bit stressful; Extremely stressful | Unchanged | Unchanged |
| Self-rated health | Ordinal | **5 categories:** Poor; Fair; Good; Very Good; Excellent | Unchanged | Excluded (female) |
| **Health Behaviors** | | | | |
| Smoking status | Categorical | **4 categories:** Non-smoker; Current smoker; Former smoker quit <5 years ago; Former smoker quit >5 years ago | Unchanged | Unchanged |
| Pack years of smoking | Continuous | **3 knot spline:** Valid range: 0-112 (male), 0-78 (female) | Unchanged | Unchanged |
| Number of drinks last week | Continuous | **3 knot spline:** Valid range: 0-50 (male), 0-24 (female) | Unchanged | Unchanged |
| Former drinker | Dichotomous | Yes; No | Unchanged | Unchanged |
| Fruit and vegetable consumption | Continuous | **3 knot spline:** Valid range: 0-48 (male), 0-31 (female) | Unchanged | Unchanged |
| Potato consumption | Continuous | **3 knot spline:** Valid range: 0-2 | Unchanged | Unchanged |

| | | | | |
|---|---|---|---|---|
| Juice consumption | Continuous | **3 knot spline:** Valid range: 0-6 (male), 0-5 (female) | Unchanged | Unchanged |
| Leisure physical activity (average kcal/kg/day) | Continuous | **3 knot spline:** Valid range: 0-16 (male), 0-12 (female) | Unchanged | Unchanged |
| **Functional Measures** | | | | |
| Personal hygiene and care | Dichotomous | Does not need help; Needs help | Excluded | - |
| Locomotion in the home | Dichotomous | Does not need help; Needs help | Excluded | - |
| Meal preparation | Dichotomous | Does not need help; Needs help | Excluded | - |
| Running errands | Dichotomous | Does not need help; Needs help | Excluded | - |
| Ordinary housework | Dichotomous | Does not need help; Needs help | Excluded | - |
| Heavy housework | Dichotomous | Does not need help; Needs help | Excluded | - |
| Finances | Dichotomous | Does not need help; Needs help | Excluded | - |
| Number of activities needing help | Categorical | - | **7 categories:** None; 1; 2; 3; 4; 5; 6 (sum of above measures, except heavy housework) | Unchanged |
| **Health Conditions** | | | | |
| Heart disease | Dichotomous | Yes; No | Unchanged | Unchanged |
| Stroke | Dichotomous | Yes; No | Unchanged | Unchanged |
| Diabetes | Dichotomous | Yes; No | Unchanged | Unchanged |
| Mood disorder | Dichotomous | Yes; No | Unchanged | Unchanged |
| High blood pressure | Dichotomous | Yes; No | Unchanged | Unchanged |
| COPD | Dichotomous | - | Yes; No | Unchanged |
| Epilepsy | Dichotomous | - | Yes; No | Excluded (male) |
| Body mass index | Continuous | **3 knot spline:** Valid range: 10-44 (male), 10-47 (female) | Unchanged | Unchanged |
| **Design** | | | | |
| Survey year | Ordinal | **6 categories:** 2001, 2003, 2005, 2007/08, 2009/10, 2011/12 | Unchanged | Unchanged |

* Age interaction included for all except socio-demographic, general health and survey year variables
Abbreviations: COPD, chronic obstructive pulmonary disease

residuals versus time for each predictor. Over-fitting was assessed in the full models using the heuristic shrinkage estimator[28], which indicated that shrinkage was unnecessary (male = 0.97; female = 0.98). Due to known challenges using survey weights in regression modelling[29], survey weights were not used for model derivation, however we recommend their use for population application and reporting[26]. To facilitate practical application, we used a step-down procedure described by Ambler[30] to create smaller, reduced models from the full models. This procedure involves removing variables that result in the smallest decrease in model $R^2$, one variable at a time, until AIC is minimized.

Model performance was assessed using overall measures of predictive accuracy, discrimination (how well the model is able to separate those who experience the outcome from those who do not) and calibration (agreement between predicted and observed risk). Predictive accuracy was assessed with Nagelkerke's $R^2$ [31] and the scaled Brier score[32,33]. Discrimination was assessed using Harrell's concordance statistic (c-statistic)[24,34]. Overall calibration, calibration within deciles of predicted risk and within population subgroups of importance to clinicians and policy-makers were assessed. Calibration within subgroups was evaluated using a predefined standard, defined as less than a 20% difference between the observed and predicted risk estimates in subgroups where at least 5% of individuals developed dementia[19,26,35].

Analyses were conducted in R v3.1[36] using the riskRegression[37] and Hmisc[38] packages.

## 3.4 RESULTS

### 3.4.1 Participants

The 2001 to 2011/12 CCHS include 253,189 Ontario respondents, of which 200,320 agreed to share their file, were successfully linked to administrative data at ICES, and were eligible for Ontario's health insurance plan. Of these, 78,097 were at least 55 years of age at survey administration.

After exclusion of those with prevalent dementia (N = 2,637), the derivation and validation cohorts included 47,739 and 27,721 respondents with 472,399 (males: 10.0 median years, interquartile range (IQR): 7.5-13.4; females: 10.2 median years, IQR 8.2-13.5) and 157,929 (males: 5.9 median years, IQR 4.7-7.0; females: 5.9 median years, IQR 4.8-7.0) total person-years of follow-up, respectively. In the derivation data, during the 5-year predicted time horizon of interest (220,972 person-years of follow-up), there were 6,734 dementia events and 2,521 deaths without dementia; in the validation data, there were 1,714 incident cases of dementia and 1,354 deaths without dementia over 127,705 person-years. Cumulative incidence curves for 5-year dementia ascertainment and death are in **Appendix 3-2**. The crude 5-year incidence rate of dementia was 2.0 per 1,000 person-years among males and 2.7 per 1,000 person-years among females, in the combined derivation and validation cohorts. Mean age at dementia ascertainment among males was 80.4 years (IQR: 75.6, 85.6) and among females was 82.6 years (IQR 77.9, 88.1).

Characteristics of the study populations are presented in **Table 3-2**. Mean age in the derivation cohorts was 66.0 (IQR 60.0-74.0) among males and 68.0 (IQR 61.0-76.0) among females and was similar in the validation cohorts. There was less than 1% missing data for most predictors. With the exception of variables that were not collected in all study years (multilingualism, former versus non-drinker, needing help with finances, mood disorder, epilepsy and COPD), smoking status had the most missing data (males: 11.2; females: 11.4%), due to missing information about time since quit among former smokers.

### 3.4.2 Model specification, development and validation

Predictor variables and degrees of freedom for the full and reduced models are presented in **Table 3-1**. Partial correlation plots are available in **Appendix 3-2**. The final full models for males and females each include a total of 90 degrees of freedom (65 main effects, 25 interaction), with 28 predictors (8 continuous) and 24 interaction terms (**Table 3-1**). The reduced models both have 74 degrees of freedom (male: 51 main, 23 interaction; female: 50 main, 24 interaction). **Appendix 3-3 and 3-4** present subdistribution hazard ratios from the full and reduced models for males and females, respectively; an interactive online visualization tool is in development to facilitate understanding of how the risk factors contribute to dementia risk. Model formulas and beta coefficients are available in **Appendix 3-5** on GitHub at https://github.com/Big-Life-Lab/predictive-algorithms, respectively.

### 3.4.3 Model performance

**Table 3-3** presents summary indicators of model performance. Both the male and female models are discriminating, indicating good ability to separate those who develop dementia from those

**Table 3-2.** Baseline study characteristics of male and female derivation and validation cohorts

| | Male Cohort | | Female Cohort | |
|---|---|---|---|---|
| **Characteristic*** | **Derivation†** | **Validation‡** | **Derivation†** | **Validation‡** |
| N | 20,506 | 11,791 | 27,233 | 15,930 |
| Person-years of follow-up | 94,237 | 53,979 | 126,735 | 73,726 |
| Dementia events (5 years) | 2,367 | 647 | 4,367 | 1,067 |
| Deaths (5 years) | 905 | 518 | 1,616 | 836 |
| **Socio-demographic Factors** | | | | |
| Age, median [IQR] | 66.0 (60.0, 74.0) | 66.0 [60.0, 74.0] | 68.0 [61.0, 76.0] | 67.0 (61.0, 76.0) |
| Ethnicity | | | | |
| White | 19,145 (93.4) | 10,785 (91.5) | 25,700 (94.4) | 14,765 (92.7) |
| South Asian/Arab/West Asian | 347 (1.7) | 208 (1.8) | 360 (1.3) | 184 (1.2) |
| Japanese/Korean/South East Asian/Filipino | 274 (1.3) | 224 (1.9) | 243 (0.9) | 276 (1.7) |
| Other/Multiple Origin/Unknown | 637 (3.1) | 479 (4.1) | 831 (3.1) | 597 (3.7) |
| Missing | 103 (0.5) | 95 (0.8) | 99 (0.4) | 108 (0.7) |
| Immigrant | | | | |
| Yes | 5,500 (26.8) | 2,926 (24.8) | 6,804 (25.0) | 3,769 (23.7) |
| No | 14,981 (73.1) | 8,818 (74.8) | 20,391 (74.9) | 12,109 (76.0) |
| Missing | 103 (0.5) | 47 (0.4) | 38 (0.1) | 52 (0.3) |
| Education | | | | |
| Less than secondary school | 6,380 (31.1) | 2,792 (23.7) | 9,402 (34.5) | 3,833 (24.1) |
| Secondary school graduate | 2,902 (14.2) | 1,840 (15.6) | 5,494 (20.2) | 3,429 (21.5) |
| Some post-secondary | 1,148 (5.6) | 499 (4.2) | 1,454 (5.3) | 665 (4.2) |
| Post-secondary graduate | 9,847 (48.0) | 6,564 (55.7) | 10,636 (39.1) | 7,888 (49.5) |
| Missing | 229 (1.1) | 96 (0.8) | 247 (0.9) | 115 (0.7) |
| Marital status | | | | |
| Now married/Common-law | 14,791 (72.1) | 8,316 (70.5) | 13,070 (48.0) | 7,873 (49.4) |
| Separated/Divorced | 2,122 (10.3) | 1,358 (11.5) | 3,180 (11.7) | 2,130 (13.4) |
| Widowed | 2,224 (10.8) | 1,175 (10.0) | 9,642 (35.4) | 4,928 (30.9) |
| Single | 1,364 (6.7) | 934 (7.9) | 1,325 (4.9) | 981 (6.2) |
| Missing | 5 (0.0) | 8 (0.1) | 16 (0.1) | 18 (0.1) |
| Neighbourhood Deprivation | | | | |
| Low | 4,153 (20.3) | 2,289 (19.4) | 4,720 (17.3) | 2,677 (16.8) |
| Moderate | 12,753 (62.2) | 7,423 (63.0) | 17,147 (63.0) | 10,160 (63.8) |
| High | 2,997 (14.6) | 1,705 (14.5) | 4,429 (16.3) | 2,538 (15.9) |
| Missing | 603 (2.9) | 374 (3.2) | 937 (3.4) | 555 (3.5) |
| Multilingual | | | | |
| Yes | 14,018 (68.4) | 1,743 (14.8) | 7,402 (27.2) | 2,103 (13.2) |
| No | 6,471 (31.6) | 4,054 (34.4) | 19,816 (72.8) | 5,741 (36.0) |
| Missing$^{\phi}$ | 17 (0.1) | 5,994 (50.8) | 15 (0.1) | 8,086 (50.8) |
| **General Health** | | | | |
| Sense of belonging | | | | |
| Very strong | 5,075 (24.7) | 2,766 (23.5) | 7,156 (26.3) | 3,987 (25.0) |
| Somewhat strong | 8,951 (43.7) | 5,518 (46.8) | 12,011 (44.1) | 7,657 (48.1) |
| Somewhat weak | 3,805 (18.6) | 2,075 (17.6) | 5,083 (18.7) | 2,709 (17.0) |
| Very weak | 1,628 (7.9) | 859 (7.3) | 2,271 (8.3) | 1,095 (6.9) |
| Missing | 1,047 (5.1) | 573 (4.9) | 712 (2.6) | 482 (3.0) |
| Self-perceived stress | | | | |
| Not at all stressful | 4,871 (23.8) | 2,616 (22.2) | 5,102 (18.7) | 2,616 (16.4) |
| Not very stressful | 6,420 (31.3) | 3,635 (30.8) | 8,554 (31.4) | 4,772 (30.0) |
| A bit stressful | 6,491 (31.7) | 3,867 (32.8) | 9,394 (34.5) | 5,739 (36.0) |
| Quite a bit stressful | 2,154 (10.5) | 1,348 (11.4) | 3,353 (12.3) | 2,279 (14.3) |
| Extremely stressful | 485 (2.4) | 270 (2.3) | 728 (2.7) | 459 (2.9) |
| Missing | 85 (0.4) | 55 (0.5) | 102 (0.4) | 65 (0.4) |
| Self-rated health | | | | |
| Excellent | 3,038 (14.8) | 1,692 (14.3) | 4,009 (14.7) | 2,474 (15.5) |
| Very Good | 6,241 (30.4) | 3,873 (32.8) | 8,467 (31.1) | 5,532 (34.7) |

| | | | | |
|---|---|---|---|---|
| Good | 6,385 (31.1) | 3,771 (32.0) | 8,469 (31.1) | 4,718 (29.6) |
| Poor | 3,376 (16.5) | 1,752 (14.9) | 4,521 (16.6) | 2,275 (14.3) |
| Fair | 1,442 (7.0) | 685 (5.8) | 1,737 (6.4) | 911 (5.7) |
| Missing | 24 (0.1) | 18 (0.2) | 30 (0.1) | 20 (0.1) |
| **Health Behaviors** | | | | |
| Smoking | | | | |
| Current smoker | 3,747 (18.3) | 2,527 (21.4) | 11,320 (41.6) | 6,396 (40.2) |
| Pack-years, median [IQR] | 40 [23.5, 55.7] | 36.5 [22, 52.5] | 28.8 [17.6, 47] | 27.0 [16.1, 45.3] |
| Former smoker ≥5 years | 9,898 (48.3) | 5,340 (45.3) | 7,544 (27.7) | 4,758 (29.9) |
| Pack-years, median [IQR] | 19.5 [7.2, 36.8] | 18 [7, 35] | 11.2 [5, 27] | 11.0 [5.0, 26.0] |
| Former smoker <5 years | 1,351 (6.6) | 627 (5.3) | 1,329 (4.9) | 673 (4.2) |
| Pack-years, median [IQR] | 46.9 [27.6, 61.3] | 45.5 [24.9, 59.4] | 34.1 [18, 53.1] | 33.8 [19.8, 52.9] |
| Non-smoker | 3,305 (16.1) | 1,877 (15.9) | 3,972 (14.6) | 2,244 (14.1) |
| Missing | 2,205 (10.8) | 1,420 (12.0) | 3,068 (11.3) | 1,859 (11.7) |
| Alcohol | | | | |
| Current drinker | 16,364 (79.8) | 9,436 (80.0) | 18,729 (68.8) | 11,200 (70.3) |
| Number of drinks last week, median [IQR] | 3 [0, 8] | 1 [0, 4] | 3 [0, 9] | 1 [0, 4] |
| Former drinker[§] | 3,532 (17.2) | 2319 (19.7) | 6,582 (24.2) | 4,682 (29.4) |
| Non-drinker[§] | 545 (2.7) | 0 (0) | 1,856 (6.8) | 0 (0) |
| Missing | 65 (0.3) | 36 (0.3) | 66 (0.2) | 48 (0.3) |
| Diet | | | | |
| Fruit and vegetable consumption, median [IQR] | 4.3 [3.1, 5.7] | 4.1 [2.8, 5.5] | 5.0 [3.7, 6.6] | 4.8 [3.4, 6.4] |
| Missing | 623 (3.0) | 481 (4.1) | 409 (1.5) | 330 (2.1) |
| Juice consumption, median [IQR] | 1.0 [0.1, 1.0] | 0.6 [0.1, 1.0] | 1.0 [0.1, 1.0] | 0.4 [0.0, 1.0] |
| Missing | 658 (3.2) | 515 (4.4) | 468 (1.7) | 382 (2.4) |
| Potato consumption, median [IQR] | 0.4 [0.3, 0.9] | 0.4 [0.1, 0.6] | 0.4 [0.1, 0.7] | 0.3 [0.1, 0.6] |
| Missing | 688 (3.4) | 539 (4.6) | 522 (1.9) | 416 (2.6) |
| Physical Activity | | | | |
| Daily energy expenditure (METs), median [IQR] | 1.5 [0.5, 3.0] | 1.6 [0.6, 3.0] | 1.1 [0.3, 2.3] | 1.2 [0.4, 2.5] |
| Missing | 824 (4.0) | 425 (3.6) | 417 (1.5) | 247 (1.6) |
| **Functional Measures** | | | | |
| Personal hygiene and care | | | | |
| Needs help | 656 (3.2) | 389 (3.3) | 1,154 (4.2) | 694 (4.4) |
| Does not need help | 19,838 (96.7) | 11,398 (96.7) | 26,073 (95.7) | 15,232 (95.6) |
| Missing | 12 (0.1) | 4 (0.0) | 6 (0.0) | 4 (0.0) |
| Locomotion in the home | | | | |
| Needs help | 460 (2.2) | 227 (1.9) | 781 (2.9) | 420 (2.6) |
| Does not need help | 20,037 (97.7) | 11,561 (98.0) | 26,444 (97.1) | 15,507 (97.3) |
| Missing | 9 (0.0) | 3 (0.0) | 8 (0.0) | 3 (0.0) |
| Meal preparation | | | | |
| Needs help | 833 (4.1) | 485 (4.1) | 1,394 (5.1) | 824 (5.2) |
| Does not need help | 19,665 (95.9) | 11,299 (95.8) | 25,829 (94.8) | 15,103 (94.8) |
| Missing | 8 (0.0) | 7 (0.1) | 10 (0.0) | 3 (0.0) |
| Running errands | | | | |
| Needs help | 1,413 (6.9) | 825 (7.0) | 4,127 (15.2) | 2,301 (14.4) |
| Does not need help | 19,083 (93.1) | 10,962 (93.0) | 23,098 (84.8) | 13,621 (85.5) |
| Missing | 10 (0.0) | 4 (0.0) | 8 (0.0) | 8 (0.1) |
| Ordinary housework | | | | |
| Needs help | 1,704 (8.3) | 998 (8.5) | 4,474 (16.4) | 2,552 (16.0) |
| Does not need help | 18,784 (91.6) | 10,780 (91.4) | 22,746 (83.5) | 13,368 (83.9) |
| Missing | 18 (0.1) | 13 (0.1) | 13 (0.0) | 10 (0.1) |
| Finances | | | | |
| Needs help | 534 (2.6) | 416 (3.5) | 1,128 (4.1) | 788 (4.9) |
| Does not need help | 15,593 (76.0) | 11,367 (96.4) | 20,183 (74.1) | 15,131 (95.0) |
| Missing[ψ] | 4,379 (21.4) | 8 (0.1) | 5,922 (21.7) | 11 (0.1) |
| Number of activities need help | | | | |

| | | | | |
|---|---|---|---|---|
| None | 14272 (69.6) | 10324 (87.6) | 16460 (60.4) | 12364 (77.6) |
| 1 | 816 (4.0) | 642 (5.4) | 2227 (8.2) | 1637 (10.3) |
| 2 | 351 (1.7) | 302 (2.6) | 1143 (4.2) | 827 (5.2) |
| 3 | 246 (1.2) | 192 (1.6) | 688 (2.5) | 474 (3.0) |
| 4 | 166 (0.8) | 128 (1.1) | 380 (1.4) | 285 (1.8) |
| 5 | 143 (0.7) | 92 (0.8) | 242 (0.9) | 180 (1.1) |
| 6 | 112 (0.5) | 87 (0.7) | 151 (0.6) | 131 (0.8) |
| Missing[ψ] | 4400 (21.5) | 24 (0.2) | 5942 (21.8) | 32 (0.2) |
| **Health Conditions** | | | | |
| Heart disease | | | | |
| Yes | 3,876 (18.9) | 2,133 (18.1) | 4,022 (14.8) | 1,982 (12.4) |
| No | 16,569 (80.8) | 9,613 (81.5) | 23,151 (85.0) | 13,891 (87.2) |
| Missing | 61 (0.3) | 45 (0.4) | 60 (0.2) | 57 (0.4) |
| Stroke | | | | |
| Yes | 751 (3.7) | 382 (3.2) | 859 (3.2) | 471 (3.0) |
| No | 19,741 (96.3) | 11,389 (96.6) | 26,351 (96.8) | 15,442 (96.9) |
| Missing | 14 (0.1) | 20 (0.2) | 23 (0.1) | 17 (0.1) |
| Diabetes | | | | |
| Yes | 3,147 (15.3) | 2,203 (18.7) | 3,142 (11.5) | 2,120 (13.3) |
| No | 17,341 (84.6) | 9,578 (81.2) | 24,078 (88.4) | 13,793 (86.6) |
| Missing | 18 (0.1) | 10 (0.1) | 13 (0.0) | 17 (0.1) |
| Mood disorder | | | | |
| Yes | 832 (4.1) | 754 (6.4) | 1,816 (6.7) | 1,581 (9.9) |
| No | 15,293 (74.6) | 11,020 (93.5) | 19,489 (71.6) | 14,335 (90.0) |
| Missing[¶] | 4,381 (21.4) | 17 (0.1) | 5,928 (21.8) | 14 (0.1) |
| High blood pressure | | | | |
| Yes | 7,509 (36.6) | 5,014 (42.5) | 11,533 (42.3) | 7,169 (45.0) |
| No | 12,950 (63.2) | 6,751 (57.3) | 15,671 (57.5) | 8,737 (54.8) |
| Missing | 47 (0.2) | 26 (0.2) | 29 (0.1) | 24 (0.2) |
| COPD | | | | |
| Yes | 1,432 (7.0) | 774 (6.6) | 2,069 (7.6) | 1,344 (8.4) |
| No | 17,430 (85.0) | 10,999 (93.3) | 23,035 (84.6) | 14,544 (91.3) |
| Missing[σ] | 1,644 (8.0) | 18 (0.2) | 2,129 (7.8) | 42 (0.3) |
| Epilepsy | | | | |
| Yes | 73 (0.4) | 0 (0) | 112 (0.4) | 0 (0) |
| No | 14,503 (70.7) | 0 (0) | 19,505 (71.6) | 0 (0) |
| Missing[ω] | 5,930 (28.9) | 11,791 (100) | 7,616 (28.0) | 15,930 (100) |
| Body mass index, median [IQR] | 26.4 [24.2, 29.2] | 26.9 [24.4, 29.7] | 25.5 [22.8, 29.0] | 25.8 [23.0, 29.6] |
| Missing | 268 (1.3) | 491 (4.2) | 880 (3.2) | 801 (5.0) |
| **Design** | | | | |
| Survey year | | | | |
| 2001 | 4,370 (21.3) | - | 5,910 (21.7) | - |
| 2003 | 5,200 (25.4) | - | 6,932 (25.5) | - |
| 2005 | 5,009 (24.4) | - | 6,778 (24.9) | - |
| 2007/08 | 5,927 (28.9) | - | 7,613 (28.0) | - |
| 2009/10 | - | 5,801 (49.2) | - | 7,848 (49.3) |
| 2011/12 | - | 5,990 (50.8) | - | 8,082 (50.7) |

[*] N and percentage unless otherwise specified

[†] Canadian Community Health Survey cycles conducted in 2000/01, 2003, 2005, and 2007/08

[‡] Canadian Community Health Survey cycles conducted in 2009/10 and 2011/12

[ϕ] Multilingualism cannot be derived in CCHS 2011/12

[§] In CCHS 2007/8, former drinkers can only be differentiated from non-drinkers in an optional module performed in only some regions; in 2009/10 and 2011/12, former drinkers cannot be differentiated from non-drinkers

[ψ] Needing help with finances was not collected in CCHS 2001

[¶] Mood disorder not ascertained in CCHS 2001

[σ] COPD was not available for survey respondents from 2007

[ω] Epilepsy not linked in CCHS 2007/8, 2009/10 and 2011/12

**Table 3-3.** DemPoRT goodness of fit summary statistics in the initial development model in the derivation and validation data, the full model derived from the combined derivation and validation data, and the reduced models*

| | Development | Validation | Combined | Reduced |
|---|---|---|---|---|
| **Male Model** | | | | |
| **Discrimination** | | | | |
| C-statistic (95% CI) | 0.82 (0.80, 0.83) | 0.83 (0.81, 0.85) | 0.82 (0.81, 0.84) | 0.82 (0.81, 0.84) |
| Ratio of 95 to 5 risk percentile | 28.6 (0.128/0.004) | 26.1 (0.119/0.005) | 36.5 (0.134/0.004) | 36.2 (0.134/0.004) |
| **Calibration** | | | | |
| Observed vs. predicted[1] | -0.76% | 4.21% | -0.61% | -0.61% |
| 5-year cumulative incidence (observed) (95% CI) | 0.044 (0.041, 0.047) | 0.045 (0.041, 0.048) | 0.044 (0.042 0.047) | 0.044 (0.042, 0.047) |
| 5-year risk (predicted) | 0.045 | 0.043 | 0.045 | 0.045 |
| Calibration slope and intercept | 0.7859, 0.0098 | 0.7799, 0.0080 | 0.824, 0.0081 | 0.8285, 0.0079 |
| **Overall Performance** | | | | |
| Brier Score (scaled) | 0.081 | 0.068 | 0.086 | 0.086 |
| Nagelkerke $R^2$ | 0.105 | 0.123 | 0.100 | 0.100 |
| **Female Model** | | | | |
| **Discrimination** | | | | |
| C-statistic (95% CI) | 0.82 (0.81, 0.83) | 0.83 (0.81, 0.85) | 0.83 (0.82, 0.83) | 0.83 (0.82, 0.83) |
| Ratio of 95 to 5 risk percentile | 54.3 (0.171/0.003) | 50.6 (0.167/0.003) | 64.8 (0.178/0.003) | 64.6 (0.178/0.003) |
| **Calibration** | | | | |
| Observed vs. predicted* | -1.07% | -10.58% | -0.78% | -0.78% |
| 5-year cumulative incidence (observed) (95% CI) | 0.060 (0.057, 0.062) | 0.053 (0.050, 0.057) | 0.057 (0.055, 0.060) | 0.057 (0.055, 0.060) |
| 5-year risk (predicted) | 0.060 | 0.059 | 0.058 | 0.058 |
| Calibration slope and intercept | 0.7671, 0.0145 | 0.8666, 0.0128 | 0.8320, 0.0102 | 0.8335, 0.0101 |
| **Overall Performance** | | | | |
| Brier Score (scaled) | 0.107 | 0.102 | 0.111 | 0.111 |
| Nagelkerke $R^2$ | 0.147 | 0.132 | 0.133 | 0.133 |

Abbreviations: CI = confidence interval

* Three types of performance tests were examined[22]. **1) Discrimination** is the ability of a predictive model to differentiate between those who experience the outcome from those who do not. *C-statistic* is a rank order statistic for predictions against true outcomes[24]. The statistic ranges from 0 to 1; a value of 0.5 indicates the model is no better than random prediction, while a value of 1 indicates the model perfectly predicts whose who will develop the outcome of interest and who will not. *Ratio of 95 to 5 risk percentile* is a measure indicating the spread of the predicted risks, where a higher ratio indicates a more discriminating algorithm. For example, a ratio of 20 indicates that the absolute risk of the event of interest is 20 times higher for a person in the 95$^{th}$ percentile of risk than for a person in the 5$^{th}$ percentile of risk. **2) Calibration** reflects agreement between

the observed outcomes and the predictions. Calibration (or accuracy) describes how well the predicted probability of disease agrees with the observed outcomes. *Observed versus predicted (O vs. P)* is the relative difference between the observed incidence and the predicted risk, calculated as (Observed – Predicted)/Observed *100. The absolute values for this calculation are the *observed 5-year cumulative incidence* and the *predicted 5-year risk*. A 1% difference indicates that 1% more events were observed than were predicted. This tables show overall O vs. P. Appendix 6 and 7 show O vs. P for specific subgroups. The *calibration slope and intercept* indicate the slope and intercept of the calibration plots. Figure 1 displays the calibration plots for validation data. **3) Overall performance measures.** *Brier score (scaled)* is a measure of overall agreement between observed and predicted risk with values between 0 and 1[33]. This scaled Brier score is very similar to Pearson's $R^2$ statistic. *Nagelkerke* $R^2$ is a measure of the amount of variation in risk between individuals in the data that is explained by the model, with values from 0 to 1[31]. Larger values indicate that more variation is explained.

who do not (Male c-statistic = 0.83, 95% confidence interval (CI): 0.81, 0.85; Female c-statistic = 0.83, 95% CI: 0.81, 0.85 in validation data). Discrimination remained stable across derivation, validation and pooled data, and in the reduced model. Within the validation data, the predicted number of dementia events somewhat differed from the observed number of events (percent difference between the 5-year observed cumulative incidence and the predicted risk, males: 4.21%; females: -10.58%), while they were very similar within the reduced models (males: -0.61%; females: -0.78%).

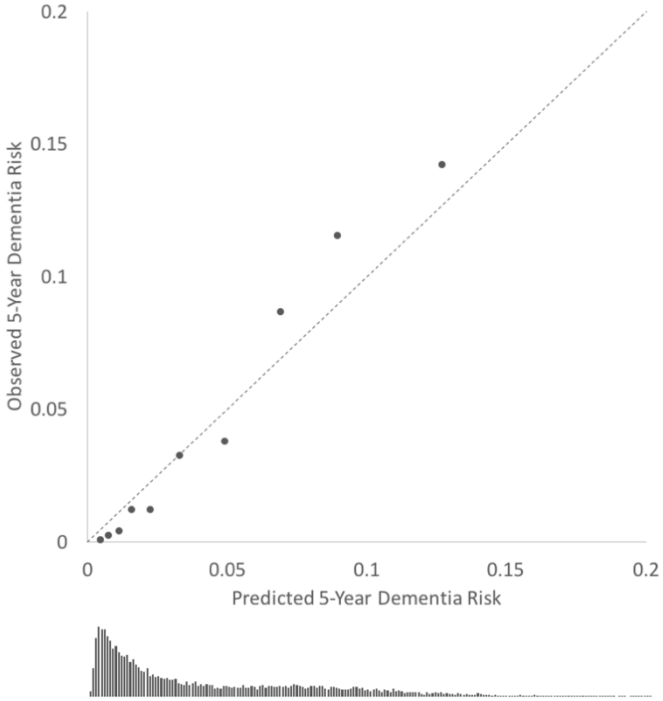
Calibration across deciles of predicted risk is presented in **Figure 3-1**. Calibration slopes in the validation data were 0.7859 among males and 0.8666 among females. Among males, the model was well-calibrated in 68 of 88 pre-defined policy-relevant population sub-groups (**Appendix 3-6**), having no more than a 20% difference in predicted versus observed risk, evaluated among sub-groups where at least 5% of individuals developed dementia. In females, the model was well-calibrated in 86 of 98 subgroups (**Appendix 3-7**). Both the male and female models underestimate dementia risk among those at older ages, those need help with daily activities and who have a history of stroke. Well-calibrated subgroups include behavioral risk exposure categories, many sociodemographic groups, stress, self-rated health, and by diabetes and hypertension status.

## 3.5 DISCUSSION

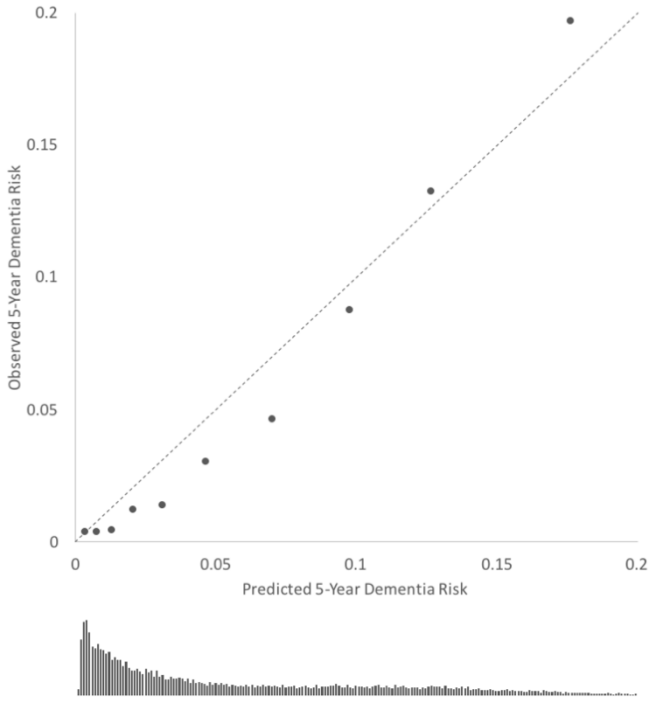
The Dementia Population Risk Tool is a discriminating and well-calibrated algorithm for predicting 5-year incidence of dementia among community-dwelling individuals, developed using self-reported risk factor information from national population health survey data.

**Figure 3-1.** Calibration plots for the full model validation data; mean predicted five year risk of dementia versus observed dementia incidence for (A) males and (B) females by decile of predicted risk. Histograms display the relative distribution of the predicted risk in the population

A




B



DemPoRT is a valuable tool for population health planning and policy development as it performs well across many population subgroups of importance to clinicians and policy makers.

Inclusion of health behaviour variables facilitates the development and evaluation of primary prevention strategies, and inclusion of sociodemographic variables enables evaluation of dementia burden and prevention strategies with an equity perspective[26]. As it was created using routinely collected population health survey data, DemPoRT can be easily applied to newer cycles of the CCHS or to other, similar national health surveys to produce population-level estimates of dementia incidence. All variables were centered on their means, facilitating application of the models to new settings by allowing for re-calibration to new populations with different risk factor distributions.

Population-level evaluation of disease including the development of disease projections is best performed using multivariable predictive risk algorithms[39,40]; DemPoRT is uniquely suited for this application. A recent review identified 39 studies describing risk algorithms to predict dementia among those in late life[41]. Sixteen studies described models with good discrimination (c-statistic of 0.80-0.89), while two studies described models with excellent discrimination (c-statistic: 0.91[13] and 0.93[6]). Of these, all but one[42] require neuropsychological (N = 16) or genetic (N = 4) testing; measures that are not available at the population-level. Furthermore, most were developed in highly defined populations with small sample sizes and few dementia events, limiting their generalizability. Seven studies included any socio-demographic or lifestyle variables, most including only education (N = 5). The only algorithm that does not require neuropsychological or genetic testing, and includes socio-demographic and lifestyle variables, is

the Dementia Risk Score (DRS), which predicts 5-year dementia risk among individuals 60 to 79 years of age using primary care data (c-statistic: 0.84)[42]. It includes measures for smoking (ascertained as current vs. non-current smoker), heavy alcohol use (yes vs. no), depression (yes vs. no), social deprivation (quintiles), and aspirin use (yes vs. no) in addition to various disease states. Linear age, age squared, linear BMI and BMI squared were included as continuous variables. The DRS may be useful for population health planning in the United Kingdom, as it was developed using over 800,000 patients from a nationally representative primary care database. However, its reliance on clinical data limits generalizability and the lack of some known sociodemographic and health behaviour risk factors (e.g. education, ethnicity, diet, physical activity) limits its application and potential usefulness in evaluating dementia prevention strategies.

DemPoRT can also be used by individual patients or their clinicians to assess dementia risk and inform decisions about lifestyle modification. See **Box 3-1** for an example of DemPoRT for individual use. As all variables are self-reported, dementia risk can be evaluated both within and outside of the clinical setting; an online calculator to facilitate this use is in development and will be made available at ProjectBigLife.ca. Variable centering enables unbiased risk assessment in individuals who provide only partial responses and will allow for dynamic risk calculation as an individual completes the questionnaire, providing a more interactive and engaging experience.

Our team has developed online health calculators as a knowledge translation tool for other risk algorithms, including for cardiovascular disease, which was developed using the same health survey data as DemPoRT[19]. The cardiovascular disease risk tool is integrated into the Heart and

**Box 3-1.** Example of DemPoRT for individual use

**75 year-old Female**

**Sociodemographic Factors**
- Secondary school graduate
- Married

**Health Behaviours**
- Former smoker, quit 10 years ago with 20 pack-years smoking history
- 2 drinks/week
- 5 fruit and vegetables/day; 1 potato/day; 1 juice/day
- Physical activity unknown

**General Health**
- Considers life a bit stressful

**Functional Measures**
- Needs help managing finances

**Chronic Conditions**
- High blood pressure
- Diabetic
- No heart disease, history of stroke, mood disorders, chronic obstructive pulmonary disease or epilepsy
- Body mass index = 30 $kg/m^2$

**5-year Dementia Risk = 9.7%**

Stroke Foundation of Canada website to provide individualized risk calculations for their eHealth Risk Assessment program. Knowledge translation of DemPoRT is facilitated by this online tool, in addition to the numerous online files.

Most predictive algorithms for dementia that include only sociodemographic and health behaviour variables perform poorly[41]. We attribute the favorable performance of DemPoRT to greater model complexity including the use of more variables, interaction terms, and flexible functions for continuous variables. Predictive model development generally prioritizes simplicity and parsimony with the goal of producing algorithms that are easier to interpret and use. More complex algorithms do not have to be more burdensome in their application, particularity if they are implemented as reflexive online tools and if unbiased calculations can be performed with partial responses. Overfitting is also often associated with increased model complexity. Our approach to model pre-specification limited this risk in the development of the DemPoRT models, which performed well in validation. While increasing model complexity may be considered unnecessary due to the generally marginal increase in overall discrimination with the addition of risk factors beyond the most predictive, added complexity also has the potential to improve model discrimination among individuals and population subgroups. Complex algorithms like DemPoRT therefore have potential to support both clinical and population-based precision medicine[19].

Favorable performance of DemPoRT may also be due in part to the choice of modelling technique. Given the late-life onset of dementia, it is important to consider the competing risk of death when developing a dementia risk prediction model, as failure to do so can result in risk

overestimation[43]. Although Cox proportional hazard modelling can be used, it has been suggested Fine and Gray subdistribution hazard models, which model the subdistribution hazard function rather than the hazard function, are better suited to prediction purposes[27]. We are only aware of one other dementia model that has used Fine and Gray regression[15].

### 3.5.1 Limitations

As the case ascertainment algorithm is imperfect, and only ascertains physician-diagnosed dementia, some individuals with dementia are not being identified[44,45]. DemPoRT's performance may therefore be overestimated, and prevalence projections using DemPoRT will underestimate the true number of individuals with dementia. As DemPoRT uses self-reported predictor information ascertained at baseline, model performance may be improved with more accurate, longitudinal predictor assessment – however, model performance was favorable regardless and the use of self-report data enhances application potential. Like the previously mentioned DRS model developed using routinely-available primary care data[42], DemPoRT underestimated dementia risk at older ages – likely due to the unavailability of variables with additional predictive ability among older adults. Other models have had success developing algorithms for older adults using neurophysiological testing[7,11]. That said, DemPoRT's underestimation can be corrected for in dementia projections using calibration techniques.

## 3.6 CONCLUSIONS

DemPoRT is the first multivariable predictive risk algorithm for dementia designed specifically for population use, with favorable performance despite using only self-reported population-level data and without the use of neuropsychological or genetic testing. It is discriminating and able to

predict dementia risk across a range of health profiles. DemPoRT will be used to answer key policy questions with respect to the future burden of dementia in Canada and will support the development and evaluation of population-level dementia strategies.

### 3.7 FUNDING

This work is supported by the Canadian Institute for Health Research (CIHR) operating grant MOP 142237, and by ICES, which is funded by an annual grant for the Ontario Ministry of Health and Long-Term Care (MOHLTC). SF was supported by a doctoral award from CIHR. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by CIHR, ICES or the Ontario MOHLTC is intended or should be inferred.

## 3.8 REFERENCES


1. Alzheimer's Disease International. *World Alzheimer Report 2018 - The State of the Art of Dementia Research: New Frontiers*.
2. Public Health Agency of Canada. *A Dementia Strategy for Canada: Together We Aspire*.
3. World Health Organization. *Dementia: A Public Health Priority*.; 2012. doi:978 92 4 156445 8
4. Australian Health Ministers Advisory Council. *National Framework for Action on Dementia 2015-2019*. Canberra, Australia: Deparment of Health; 2015.
5. Assistant Secretary for Planning and Evaluation U.S. Department of Health and Human Services. *National Plans to Address Alzheimer's Disease*. Washington, D.C.; 2012.
6. Mura T, Baramova M, Gabelle A, et al. Predicting dementia using socio-demographic characteristics and the Free and Cued Selective Reminding Test in the general population. *Alzheimers Res Ther*. 2017;9(1):21. doi:10.1186/s13195-016-0230-x
7. Wolfsgruber S, Jessen F, Wiese B, et al. The CERAD neuropsychological assessment battery total score detects and predicts alzheimer disease dementia with high diagnostic accuracy. *Am J Geriatr Psychiatry*. 2014;22(10):1017-1028. doi:10.1016/j.jagp.2012.08.021
8. Grober E, Sanders AE, Hall C, Lipton RB. Free and Cued Selective Reminding Identifies Very Mild Dementia in Primary Care. *Alzheimer Dis Assoc Disord*. 2010;24(3):1. doi:10.1097/WAD.0b013e3181cfc78b
9. Jessen F, Wiese B, Bickel H, et al. Prediction of Dementia in Primary Care Patients. Deli M, ed. *PLoS One*. 2011;6(2):e16852. doi:10.1371/journal.pone.0016852
10. Barnes DE, Covinsky KE, Whitmer RA, Kuller LH, Lopez OL, Yaffe K. Predicting risk of dementia in older adults: The late-life dementia risk index. *Neurology*. 2009;73(3):173-179. doi:10.1212/WNL.0b013e3181a81636
11. Derby CA, Burns LC, Wang C, et al. Screening for predementia AD: Time-dependent operating characteristics of episodic memory tests. *Neurology*. 2013;80(14):1307-1314. doi:10.1212/WNL.0b013e31828ab2c9
12. Meng X, D'Arcy C, Morgan D, Mousseau DD. Predicting the risk of dementia among Canadian seniors: a useable practice-friendly diagnostic algorithm. *Alzheimer Dis Assoc Disord*. 2013;27(1):23-29. doi:10.1097/WAD.0b013e318247a0dc

13. Jungwirth S, Sonja AE, Ae Z, et al. Prediction of Alzheimer dementia with short neuropsychological instruments. doi:10.1007/s00702-009-0318-6
14. Reitz C, Tang M-X, Schupf N, Manly JJ, Mayeux R, Luchsinger JA. A summary risk score for the prediction of Alzheimer disease in elderly persons. *Arch Neurol*. 2010;67(7):835-841. doi:10.1001/archneurol.2010.136
15. Licher S, Leening MJG, Yilmaz P, et al. Development and Validation of a Dementia Risk Prediction Model in the General Population: An Analysis of Three Longitudinal Studies. *Am J Psychiatry*. December 2018:appi.ajp.2018.1. doi:10.1176/appi.ajp.2018.18050566
16. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol*. 2006;5(9):735-741. doi:10.1016/S1474-4422(06)70537-3
17. Anstey KJ, Cherbuin N, Herath PM. Development of a New Method for Assessing Global Risk of Alzheimer's Disease for Use in Population Health Approaches to Prevention. 2013.
18. Fisher S, Hsu A, Mojaverian N, et al. Dementia Population Risk Tool (DemPoRT): Study protocol for a predictive algorithm assessing dementia risk in the community. *BMJ Open*. 2017;7(10). doi:10.1136/bmjopen-2017-018018
19. Manuel DG, Tuna M, Bennett C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: The Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ*. 2018;190(29):E871-E882. doi:10.1503/cmaj.170914
20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol*. 2015;67(6):1142-1151. doi:10.1016/j.eururo.2014.11.025
21. Béland Y. Canadian Community Health Survey - Methodological overview. *Heal Reports*. 2002;13(3):9-14.
22. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W73. doi:10.7326/M14-0698
23. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of Physician-Diagnosed

Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimer's Dis*. 2016;54(1):337-349. doi:http://dx.doi.org/10.3233/JAD-160105

24. Harrell FE. *Regression Modeling Strategies with Applicaitons to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer; 2001.
25. Steyerberg EW. *Clinical Prediction Models*. New York: Springer; 2009.
26. Manuel DG, Perez R, Sanmartin C, 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):1-27. doi:10.1371/journal.pmed.1002082
27. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
28. Van Houwelingen JC, Le CS. Predictive value of statistical models. *Stat Med*. 1990;9(11):1303-1325.
29. Gelman A. Struggles with Survey Weighting and Regression Modeling 1. *Stat Sci*. 2007;22(2):153-164. doi:10.1214/088342306000000691
30. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: A simulation study based on clinical data. *Stat Med*. 2002;21(24):3803-3822. doi:10.1002/sim.1422
31. Nagelkerke NJD. A Note on a General Definition of the Coefficient of Determination. *Biometrika*. 1991;78(3):691-692.
32. Arkes HR, Dawson N V, Speroff T, et al. The covariance decomposition of the probability score and its use in evaluating prognostic estimates. SUPPORT Investigators. *Med Decis Making*. 1995;15(2):120-131. doi:10.1177/0272989X9501500204
33. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. 2010;63(8):938-939; author reply 939. doi:10.1016/j.jclinepi.2009.11.009
34. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546.
35. Manuel DG, Perez R, Bennett C, et al. *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity and Stress on Health and Life Expectancy in Ontario. An ICES/PHO Report*. Toronto; 2012.
36. R Core Team. R: A language and environment for statistical computing. 2016.

37. Gerds TA, Blanche P, Mogense UB, Ozenne B. *RiskRegression: Risk Regression Models and Prediction Scores for Survial Analysis with Competing Risks. R Package.*
38. Harrell FE. *Hmisc: Harrell Miscellaneous. R Package*.
39. Manuel DG, Rosella LC, Hennessy D, Sanmartin C, Wilson K. Predictive risk algorithms in a population setting: an overview. *J Epidemiol Community Health*. 2012;66(10):859-865. doi:10.1136/jech-2012-200971
40. Manuel DG, Lim J, Tanuseputro P, et al. Revisiting Rose: strategies for reducing coronary heart disease. *BMJ*. 2006;332(7542):659-662. doi:10.1136/BMJ.332.7542.659
41. Hou X, Feng L, Zhang can, cao X, Tan L, Yu J-T. Models for predicting risk of dementia: a systematic review. *J Neurol Neurosurg Psychiatry*. 2019;90:373-379. doi:10.1136/jnnp-2018-318212
42. Walters K, Hardoon S, Petersen I, et al. Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Med*. 2016;14(1):1-12. doi:10.1186/s12916-016-0549-y
43. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*. 2012;31(11-12):1089-1097. doi:10.1002/sim.4384
44. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence. *Aging Ment Health*. 2011;15(8):978-984. doi:10.1080/13607863.2011.596805
45. Kosteniuk JG, Morgan DG, O'Connell ME, et al. Incidence and prevalence of dementia in linked administrative health data in Saskatchewan, Canada: a retrospective cohort study. *BMC Geriatr*. 2015;15(1):73. doi:10.1186/s12877-015-0075-3

# CHAPTER 4: Risk of dementia and the impact of unhealthy lifestyle behaviours on disease burden in Canada: A population-level modeling study

## 4.0 PREFACE

This chapter presents application of the DemPoRT algorithm and will be submitted for publication with the following author list:

> Stacey Fisher, Douglas G Manuel, Carol Bennett, Yulric Sequeira, Meltem Tuna, Anan Bader Eddeen, Mahsa Jessri, Amy Hsu, Laura Rosella, Monica Taljaard, Geoffrey Anderson, Peter Tanuseputro

Supporting information is provided in appendices at the end of the thesis (Section 7.2). The use of data in this project was authorized under section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a Research Ethics Board.

SF was responsible for project conception, study design, data analysis and interpretation of the results, and drafting and revision of the manuscript. DGM and PT were responsible for the grant application, and contributed to project conception, study design, protocol development and result interpretation. CB, AH, MJ, LR, MTaljaard and GA contributed to the design of the study and result interpretation. YS contributed to data analysis and result interpretation. MTuna and ABE contributed to the study design and provided data/statistical support. All authors provided critical reviews of the manuscript.

## 4.1 ABSTRACT

**Objective:** The Dementia Population Risk Tool (DemPoRT) health algorithm was used to estimate dementia risk, the number of future incident dementia cases, and the impact of unhealthy lifestyle behaviours on dementia burden in Canada.

**Methods:** The validated DemPoRT algorithm was applied to respondents of the 2013/14 Canadian Community Health Survey who were 55+ years of age (N = 62,528 respondents representing almost 10 million Canadians) to estimate the five-year population risk of dementia and number of new incident cases in Canada. The impact of smoking, poor diet and physical inactivity on dementia risk in the presence of socio-demographic and other risk factors was evaluated using the DemPoRT algorithm, a cause-deleted risk factor approach, and risk factor estimates derived from five increasingly adjusted, sex-specific Cox proportional hazards regression models.

**Results:** The five-year estimated dementia risk was 3.8% for men and 5.2% for women, corresponding to an estimated 452,000 incident cases of dementia in Canada between 2013/14 and 2018/19. An estimated 31% (56,500 out of 181,000 incident cases) of dementia risk among males and 47% (127,000 out of 271,000 incident cases) of risk among females was associated with smoking, poor diet and physical inactivity. Physical activity had the largest impact on dementia risk among males (13.3% attributable risk and 24,000 cases), while smoking had the largest impact among females (19.9% attributable risk and 54,000 cases).

**Conclusion:** Poor health behaviours have a large impact on the population dementia risk in Canada. These findings emphasize the importance of reducing the prevalence of risk behaviours for preventing dementia and can be used to inform resource allocation and primary prevention strategies.

## 4.2 INTRODUCTION

As life expectancies increase and populations age, the burden of disease from dementia is becoming an increasingly significant concern worldwide. In Canada, an estimated 600,000 individuals are currently living with dementia[1], and approximately nine people over the age of 65 are diagnosed with dementia every hour[2]. Although there is no cure or effective therapies to stop dementia progression, there is increasing evidence that healthy lifestyle choices may prevent or delay dementia onset. In meta-analyses, high fruit and vegetable intake (compared to low intake) and high levels of physical activity (compared to a sedentary lifestyle) have been associated with a decreased risk of dementia and cognitive decline, while current smoking, compared to never smoking, has been associated with an almost 80% increased risk of dementia[3–6]. To prepare for the growing burden of dementia and inform implementation of preventative strategies, it is important to understand the population risk of disease, who will develop dementia in the future, and what sub-groups of the population are at the highest risk and have the highest absolute burden.

The Dementia Population Risk Tool (DemPoRT) is a validated population risk algorithm for predicting dementia risk among community-dwelling individuals 55+ years of age (see Chapter 3). Uniquely designed for population health planning purposes, the DemPoRT algorithm predicts five-year dementia risk using population-level data including detailed health behaviour and socio-demographic information. The purpose of this study is to use the DemPoRT algorithm to estimate the population risk and future incident cases of dementia in Canada, and evaluate the impact of unhealthy lifestyle behaviours on the dementia burden.

## 4.3 METHODS

### *4.3.1 Estimation of 5-year dementia risk*

The DemPoRT model was applied to the 2013/14 national Canadian Community Health Survey (CCHS) public use dataset to estimate 5-year risk of dementia for all eligible survey respondents at least 55 years of age. The CCHS is a cross-sectional health survey developed by Statistics Canada to collect data about the health status, health determinants and health care use of Canadians. It is representative of 98% of the Canadian population age 12 years and older living in private dwellings; details of the survey methodology have been previously published[7]. The person-level response rate in the 2013/14 cycle was 87.3%[8]. The sample size of the application data was 62,528 (26,725 males and 35,803 females), representing 9,961,000 Canadians (4,759,000 males and 5,202,000 females). Unlinked, publicly-available CCHS data was used to demonstrate how the DemPoRT model can be widely applied for dementia risk evaluation. Bootstrap weights could not be used in the description of population characteristics as they are not available for the public use version of the CCHS, however, similar analyses have reported tight confidence intervals[9,10].

Five-year risk of dementia for each individual in the application data was calculated using the validated DemPoRT model. DemPoRT is sex-specific predictive algorithm developed to predict 5-year dementia risk for community-dwelling adults 55+ years of age using self-reported information on socio-demographic characteristics, general and chronic health conditions, health behaviours and physical functioning. Subdistribution hazard models, which consider competing risk of death, were developed using CCHS data (survey years 2001 to 2012; N = 75,460) linked to administrative health data for dementia ascertainment. DemPoRT is discriminating (C-statistic

males: 0.83; females: 0.82) and well-calibrated across a wide range of population subgroups. Full derivation and validation details can be found in a previous study (Chapter 3).

Application of the DemPoRT model to a data source different than that it was derived on required the use of strategies to handle differences between the DemPoRT model variables and the available data. Model variables not available in the public use CCHS data (epilepsy and multilingualism) were set to their population mean values (mean imputed); as all variables were centered on their means prior to model derivation, this allowed for the model to continue to produce population dementia risk estimates. For categorical variables that were not available with the same level of detail as specified in DemPoRT, the DemPoRT beta coefficients were weighted using the variable prevalence in the Canadian population. For continuous variables that were only available as categorical variables in the CCHS public use application data, sex-specific mean values from the population were used for each category. For example, males within the 70 to 74 age group we assigned an age of 71.9, the mean age among males in this age group in the Canadian population. All other missing values were mean imputed. Five-year risk and number of incident dementia cases were summarized using CCHS sample weights and reported within population subgroups.

### *4.3.2 Association of unhealthy behaviours and dementia*

Risk factor estimates from the DemPoRT models are not appropriate for assessing the impact of health behaviours on dementia burden as DemPoRT was developed for predictive purposes and includes mediating variables. Therefore, associations between smoking, alcohol consumption, poor diet and physical inactivity and incidence of dementia were estimated using separate, sex-

specific, cause-specific Cox proportional hazard regression models, considering death as a competing risk event, and the 2001-2012 Ontario sample of the CCHS share file. Five increasingly adjusted models were run to assess the impact of confounder adjustment on the health behaviour effects. Model 1 included the health behaviour variables and age only; model 2, 3, 4 and 5 additionally included demographic variables, social variables, chronic conditions and upstream factors. Models 4 and 5 also included interactions between age and health behaviours, as well as between age and chronic conditions. Continuous variables were modelled using restricted cubic splines, and all variables were centered on their means. Full specification of the five new models can be found in **Appendix 4-1**. Beta coefficients from these models are more appropriate for assessing the impact of unhealthy behaviours on the burden of dementia. To describe the relative risk of dementia associated with each health behaviour, dementia risk for a hypothetical "average" individual with all covariate values set to their reference values except for the health behaviour of interest was estimated and compared to the dementia risk for an individual with all values at reference. Uncertainty intervals (UI) around the 'best' relative risk estimate (from model 3) were created using the minimum and maximum estimates from the five increasingly adjusted models (usually from models 1 and 5).

*4.3.3 Impact of unhealthy behaviours on dementia risk and incident cases*

The impact of unhealthy behaviours on dementia risk and on the number of 5-year incident dementia cases was estimated using a cause-deleted risk factor approach and beta coefficients external to the model, using an unpublished method developed by Manuel DG *et al* (**Appendix 4-2**). In short, regression coefficients associated with the health behaviours of interest in the original DemPoRT algorithm were replaced with the health behaviour beta coefficients from

model 3 described above. These new models were used to calculate two new predicted risks: one using the original risk profile, and the second using a counterfactual 'healthy reference' profile in which unhealthy behaviour values were changed to healthy values. Individuals with healthier than reference exposures were not changed. The difference between these two risks is the cause-deleted effect, and the original risk minus the cause-deleted effect is the cause-deleted risk. Further details of this method are available in **Appendix 4-2**. The healthy reference profile was defined as: never smoker, consumption of fruit and vegetables at least 5 times per day, consumption of potatoes no more than 5 times per week and juice no more than once per day, and an average of least 3 METs of physical activity per day. These healthy reference values were selected using review of national and international healthy living guidelines.

The impact of unhealthy behaviours on the Canadian burden of dementia was evaluated for smoking, diet and physical activity alone and considering the three risk factors together. As all levels of alcohol consumption were found to be protective of dementia incidence, the impact of alcohol was not evaluated. The impact on dementia risk was calculated by averaging the cause-deleted effects using the CCHS sample weights, while the impact on the number of incident cases was calculated by summing the products of the cause-deleted effects and the sample weights. The strongest and weakest beta coefficients (usually from models 1 and 5) were used to create uncertainty intervals around the best estimate (model 3), providing information about how the health behaviour - dementia association changes with differing model adjustment.

## 4.4 RESULTS

Five-year estimated dementia risk in the 2013/14 Canadian population was 3.8% for men and 5.2% for women, corresponding to 452,000 incident dementia cases between 2013/14 and 2018/19. Population characteristics, five-year baseline dementia risk and the number of incident cases by population sub-groups are reported in **Table 4-1**. Risk and dementia cases increased with age, with an approximately 0.5% risk among those 55 to 59 years of age contributing to 3% of incident dementia cases. Over a third of the incident dementia burden was among those 80 years and older, where risk was 13.0% in males and 16.2% in females. Individuals with less than a high school education had a higher dementia risk than those with more education, however the largest proportion of incident dementia cases was among those with more education; 47% (N = 85,800) of cases among males and 37% (N = 100,000) among females were among post-secondary graduates. Dementia risk varied by health behaviour characteristics, with the highest risks among former smokers, light or non-drinkers, those with poor diets (fewer fruit and vegetables and greater consumption of juice and potatoes) and those who do less physical activity. Among males, the burden of dementia was larger among long-term former smokers than current smokers, while among females, burden was largest among non-smokers. Dementia burden among both males and females was largest among those who do little physical activity Dementia risk additionally varied by stress, self-rated health, needing help with activities of daily living and BMI. The population subgroups with the largest risks, apart from those in older age groups, were those who need help with activities of daily living and females who had heart disease or were underweight.

**Table 4-1.** Weighted distribution of baseline characteristics, 5-year incidence and burden of dementia among Canadians 55 years of age or older in the 2013/14 CCHS cohort by socio-demographic factors, health behaviours, general health, functional measures and health conditions

| | Males | | | Females (%) | | |
|---|---|---|---|---|---|---|
| **Characteristic[1]** | **Population Proportion** | **5-year Dementia Risk (%)** | **Number of Incident Cases (%)** | **Population Proportion** | **5-year Dementia Risk (%)** | **Number of Incident Cases (%)** |
| **Overall** | 4,759,000 | 3.8 | 181,000 | 5,202,000 | 5.2 | 272,000 |
| **Socio-demographic Factors** | | | | | | |
| Age | | | | | | |
| 55 to 59 | 27.2 | 0.5 | 6,150 (3.4) | 24.7 | 0.6 | 7,110 (2.6) |
| 60 to 64 | 22.8 | 1.2 | 13,100 (7.2) | 21.1 | 1.6 | 18,000 (6.6) |
| 65 to 69 | 18.2 | 2.6 | 22,900 (12.7) | 18.3 | 3.4 | 32,200 (11.8) |
| 70 to 74 | 12.9 | 5.7 | 34,700 (19.2) | 13.2 | 6.6 | 45,600 (16.8) |
| 75 to 79 | 9.3 | 9.7 | 42,800 (23.6) | 9.7 | 11.6 | 58,700 (21.6) |
| 80+ | 9.6 | 13.4 | 61,600 (34.0) | 13.0 | 16.2 | 110,000 (40.4) |
| Ethnicity | | | | | | |
| White | 85.5 | 3.9 | 151,000 (83.4) | 86.2 | 5.4 | 230,000 (84.6) |
| Visible minority | 14.5 | 3.3 | 22,500 (12.4) | 13.8 | 4.0 | 27,500 (10.1) |
| Education | | | | | | |
| Post-secondary graduate | 56.1 | 3.3 | 85,800 (47.4) | 49.3 | 4.0 | 100,000 (36.8) |
| Some post-secondary | 3.4 | 3.3 | 5,210 (2.9) | 3.5 | 4.9 | 8,770 (3.2) |
| High school graduate | 18.0 | 3.2 | 26,400 (14.6) | 23.3 | 4.8 | 56,700 (20.8) |
| Less than high school | 22.6 | 5.7 | 59,200 (31.0) | 23.9 | 8.1 | 98,500 (36.2) |
| Marital status | | | | | | |
| Now married/Common-law | 78.0 | 3.7 | 137,000 (75.7) | 60.1 | 3.9 | 121,000 (44.5) |
| Separated/Divorced/Widowed | 15.0 | 4.9 | 35,100 (19.4) | 33.1 | 7.8 | 134,000 (49.3) |
| Single | 7.0 | 2.6 | 8,520 (4.7) | 6.8 | 4.4 | 15,600 (5.7) |
| **Health Behaviors** | | | | | | |
| Smoking | | | | | | |
| Non-smoker | 23.6 | 3.8 | 41,700 (23.0) | 41.4 | 5.9 | 126,000 (46.3) |
| Former smoker quit ≥5 years | 39.5 | 4.4 | 82,000 (45.3) | 26.6 | 5.0 | 68,300 (25.1) |
| Pack-years, median [IQR] | 35.9 [20.5, 51.2] | - | - | 26.9 [16.8, 43.7] | - | - |
| Former smoker quit <5 years | 20.8 | 3.5 | 34,700 (19.2) | 19.3 | 4.8 | 48,000 (17.6) |
| Pack-years, median [IQR] | 45.2 [27.6, 56.9] | - | - | 29.5 [18.4, 49.4] | - | - |
| Current smoker | 16.0 | 2.6 | 19,600 (10.8) | 12.7 | 3.6 | 23,800 (8.8) |
| Pack-years, median [IQR] | 21.2 [11.6, 36.5] | - | - | 13.8 [7.0, 24.3] | - | - |
| Alcohol Consumption[2] | | | | | | |
| Median drinks per week [IQR] | 0.0 [0.0, 4.0] | - | - | 0.0 [0.0. 1.0] | - | - |
| Light or non-drinker | 74.0 | 4.1 | 137,000 (75.7) | 83.6 | 5.4 | 226,000 (83.1) |
| Moderate drinker | 21.3 | 3.2 | 30,500 (16.9) | 13.5 | 4.2 | 28,200 (10.4) |
| Heavy drinker | 4.8 | 2.5 | 5,500 (3.0) | 2.9 | 4.0 | 5,880 (2.2) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Fruit and vegetable consumption | | | | | | |
| Median [IQR] | 2.9 [1.9, 4.2] | - | - | 3.9 [2.6, 5.4] | - | - |
| 5+ | 27.1 | 3.4 | 39,600 (21.9) | 46.9 | 4.4 | 97,300 (35.8) |
| <5 | 72.9 | 3.6 | 111,000 (61.3) | 53.1 | 5.7 | 143,000 (52.6) |
| Daily juice consumption | | | | | | |
| Median [IQR] | 0.4 [0.1, 1.0] | - | - | 0.2 [0.0, 1.0] | - | - |
| <1 | 59.7 | 3.4 | 89,500 (49.4) | 65.2 | 4.2 | 134,000 (49.3) |
| 1+ | 40.3 | 3.9 | 70,600 (39.0) | 34.8 | 6.9 | 119,000 (43.8) |
| Weekly potato consumption | | | | | | |
| Median [IQR] | 2.1 [0.7, 4.2] | - | - | 2.1 [0.7, 4.2] | - | - |
| <6 | 84.9 | 3.2 | 122,000 (67.4) | 87.6 | 4.6 | 196,000 (72.1) |
| 6+ | 15.1 | 5.6 | 37,600 (20.8) | 12.4 | 9.0 | 55,000 (20.2) |
| Physical Activity[3] | | | | | | |
| Median [IQR] | 1.6 [0.6, 3.0] | - | - | 1.3 [0.5, 2.7] | - | - |
| ≥ 3 METs/day | 27.2 | 3.1 | 38,600 (21.3) | 21.6 | 4.0 | 43,000 (15.8) |
| 1 to < 3 METs/day | 38.2 | 3.5 | 61,100 (33.8) | 38.5 | 4.7 | 91,600 (33.7) |
| < 1 MET/day | 34.6 | 4.1 | 64,800 (35.8) | 39.9 | 6.2 | 125,000 (46.0) |
| **General Health** | | | | | | |
| Self-perceived stress | | | | | | |
| Not at all stressful | 22.1 | 4.6 | 48,500 (26.8) | 17.3 | 6.8 | 61,200 (22.5) |
| Not very stressful | 27.3 | 3.8 | 49,300 (27.2) | 27.3 | 5.3 | 74,800 (27.5) |
| A bit stressful | 34.6 | 3.6 | 58,100 (32.1) | 37.5 | 4.8 | 93,900 (34.5) |
| Quite a bit stressful | 13.6 | 3.1 | 19,700 (10.9) | 14.5 | 4.1 | 30,800 (11.3) |
| Extremely stressful | 2.5 | 3.4 | 3,980 (2.2) | 3.3 | 4.5 | 7,690 (2.8) |
| Self-rated health | | | | | | |
| Excellent | 15.1 | 2.9 | 20,600 (11.4) | 17.3 | 4.1 | 37,000 (13.6) |
| Very Good | 33.6 | 3.2 | 51,000 (28.2) | 32.6 | 4.7 | 79,800 (29.3) |
| Good | 32.6 | 4.1 | 63,400 (35.0) | 31.4 | 5.4 | 88,000 (32.4) |
| Fair | 13.4 | 5.2 | 33,200 (18.3) | 13.5 | 6.7 | 47,100 (17.3) |
| Poor | 5.3 | 5.9 | 12,600 (7.0) | 5.1 | 6.9 | 18,200 (6.7) |
| **Functional Measures** | | | | | | |
| Number of activities needing help | | | | | | |
| None | 88.6 | 3.3 | 140,000 (77.3) | 79.5 | 4.3 | 177,000 (65.1) |
| 1 | 4.6 | 7.5 | 16,300 (9.0) | 8.9 | 7.6 | 35,100 (12.9) |
| 2 | 2.1 | 6.5 | 6,580 (3.6) | 4.4 | 10.3 | 23,200 (8.5) |
| 3 | 1.6 | 7.2 | 5,510 (3.0) | 3.0 | 9.0 | 13,800 (5.1) |
| 4 | 1.1 | 7.8 | 4,170 (2.3) | 1.8 | 10.5 | 9,650 (3.5) |
| 5 | 1.0 | 10.3 | 4,830 (2.7) | 1.5 | 9.0 | 6,960 (2.6) |
| 6 | 0.9 | 7.1 | 3,000 (1.7) | 1.0 | 8.0 | 4,220 (1.6) |
| **Health Conditions** | | | | | | |
| Heart disease | 14.4 | 5.9 | 40,100 (22.2) | 9.6 | 8.7 | 43,100 (15.8) |
| Stroke | 2.9 | 6.4 | 8,710 (4.8) | 2.5 | 8.1 | 10,600 (3.9) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Diabetes | 16.5 | 4.8 | 37,300 (20.6) | 12.8 | 6.2 | 41,100 (15.1) |
| Mood disorder | 6.0 | 3.5 | 9,830 (5.4) | 9.5 | 4.2 | 20,600 (7.6) |
| High blood pressure | 39.4 | 4.5 | 84,600 (46.7) | 39.3 | 6.5 | 132,000 (48.6) |
| COPD | 5.7 | 5.5 | 14,700 (8.1) | 6.4 | 6.9 | 22,800 (8.4) |
| Body mass index | | | | | | |
| Median [IQR] | 26.8 [24.2, 29.5] | - | - | 25.5 [22.7, 29.3] | - | - |
| Underweight | 0.8 | 6.9 | 2,400 (1.3) | 2.8 | 10.0 | 13,700 (5.0) |
| Normal weight | 32.0 | 4.4 | 62,700 (34.6) | 42.5 | 5.9 | 121,000 (44.5) |
| Overweight | 45.3 | 3.4 | 69,700 (38.5) | 33.7 | 4.8 | 79,000 (29.0) |
| Obese | 21.9 | 2.5 | 5,600 (3.1) | 21.0 | 3.0 | 9,980 (3.7) |

[1] Column percentage, unless otherwise indicated; percentages do not total to 100% as missing values are not reported

[2] Weekly consumption of alcohol greater than 15 drinks for males or 10 for females considered heavy, from 4 to 15 drinks for males or 3 to 10 drinks for females considered moderate, and less than 4 drinks for males or less than 3 drinks for females considered light/non-drinker

[3] Average metabolic equivalent of task (METs) per day

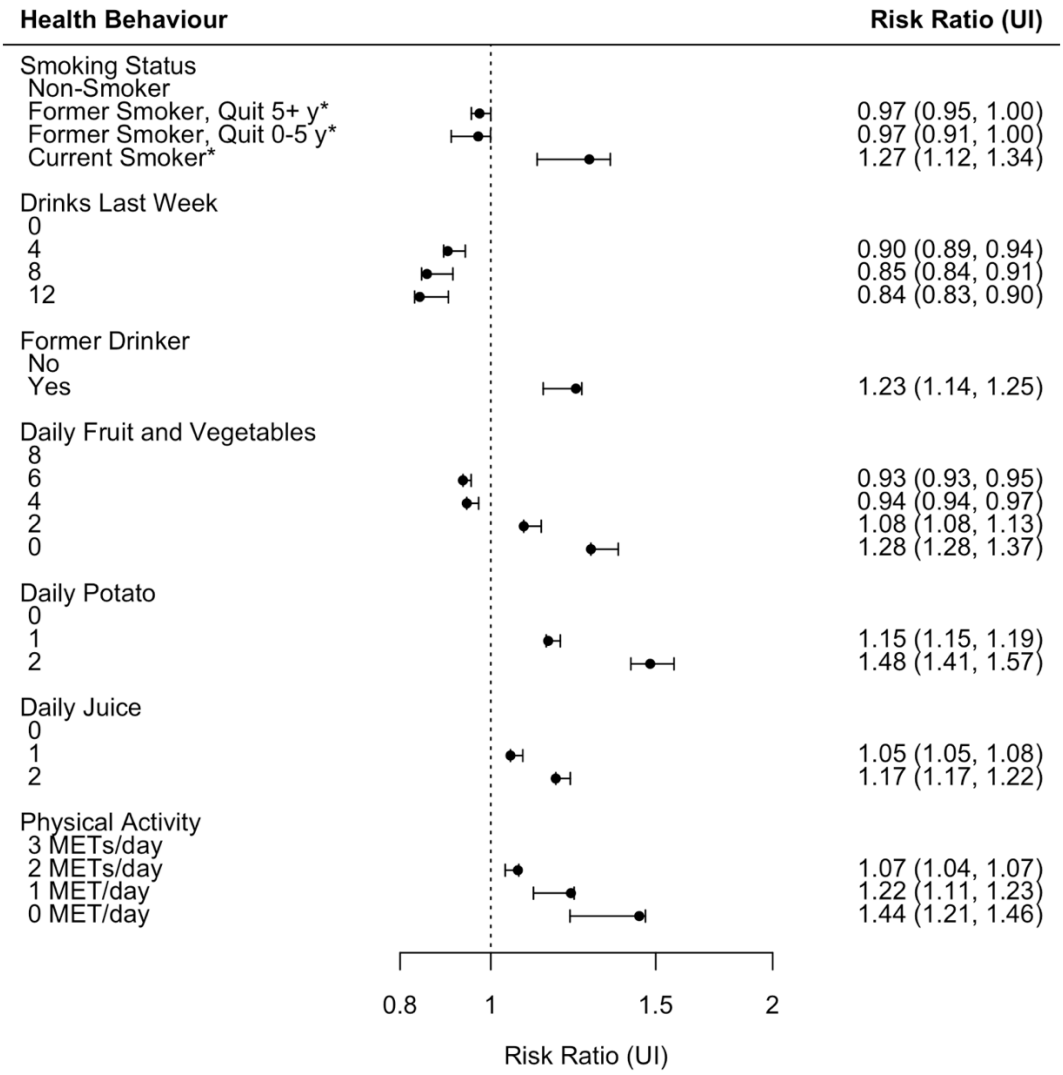
*Association between health behaviours and dementia risk*

Relative risks of dementia associated with smoking, alcohol consumption, poor diet and physical inactivity for males and females are presented in **Figures 4-1 and 4-2**, respectively. Among males, current smokers with a 21.2 pack-year smoking history had a 27% (RR = 1.27, UI: 1.21, 1.34) increased risk of dementia compared to non-smokers, while there was no association with former smoking. Among females, both current and former smoking was associated with increased risk of dementia. Former drinking was associated with a 23% (UI: 1.14, 1.25) and a 28% (UI: 1.17, 1.29) increased risk of dementia in males and females, respectively. A higher number of drinks consumed in the last week was protective of dementia incidence in both males and females. Consumption of few fruit and vegetables, more potatoes and more juice was associated with higher risks of dementia in males and females. Little physical activity was also associated with increased dementia risk, with physical inactivity (average of 0 METs/day) associated with a 44% (RR = 1.44, UI: 1.21, 1.46) and a 29% (RR = 1.29, UI: 1.12, 1.30) increased risk among males and females, respectively. Beta coefficients for models 1 to 5 are available in **Appendix 4-3 and 4-4**.

*The burden of health behaviours on dementia risk in Canada, 2013/14*

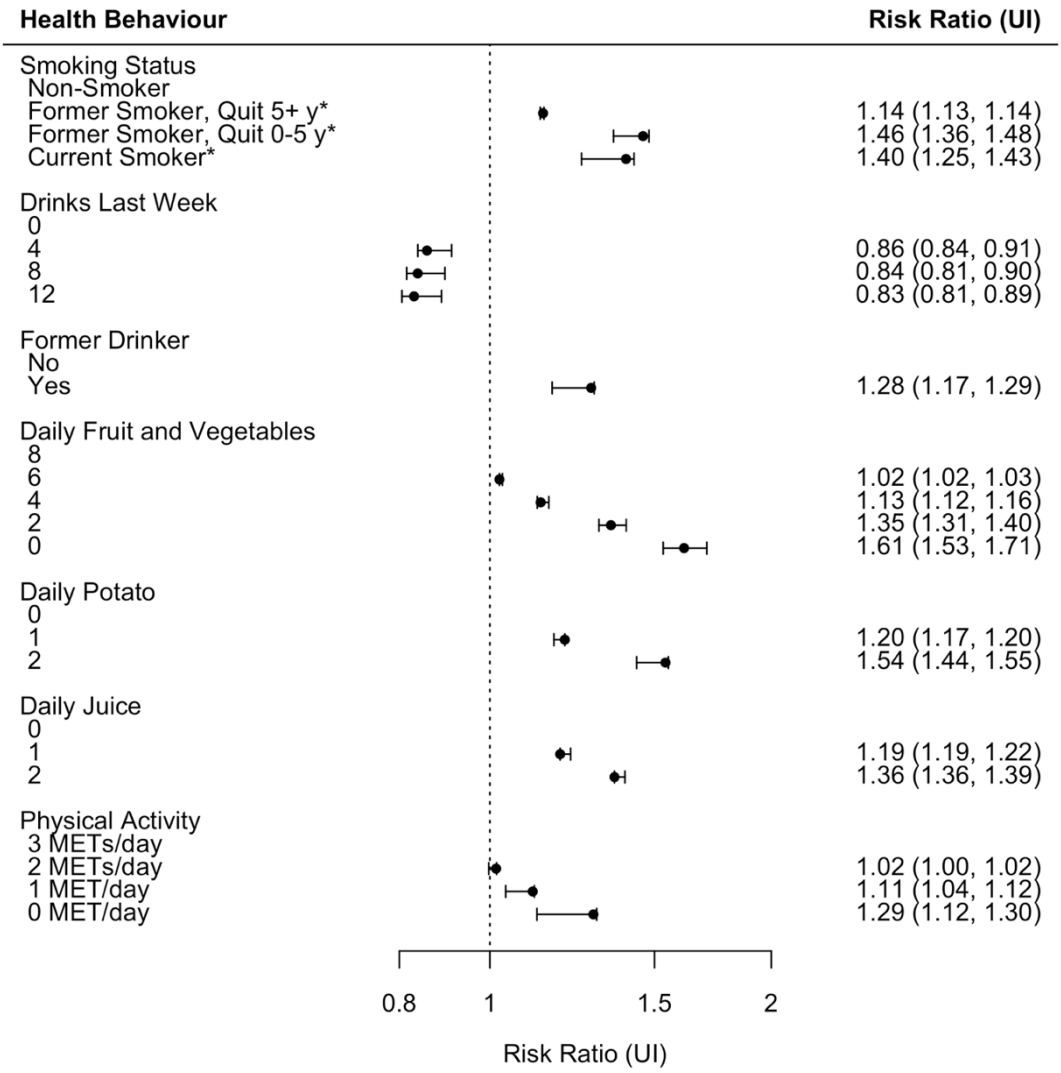
**Figure 4-3** shows the estimated impact of unhealthy behaviours on 5-year dementia risk and the number of incident dementia cases in the 2013/14 Canadian population. An estimated 31% of dementia risk among males (56,500 out of 181,000 incident cases) and 47% of risk among females (127,000 out of 271,000 incident cases) was associated with smoking, poor diet and inactivity. Therefore, an estimated 40.5% of dementia cases (183,000 of 452,000) accumulated from 2013/14 to 2018/19 may be associated with exposure to smoking, physical inactivity or

**Figure 4-1.** Relative risk ratios for health behaviours by mean risk exposure – males



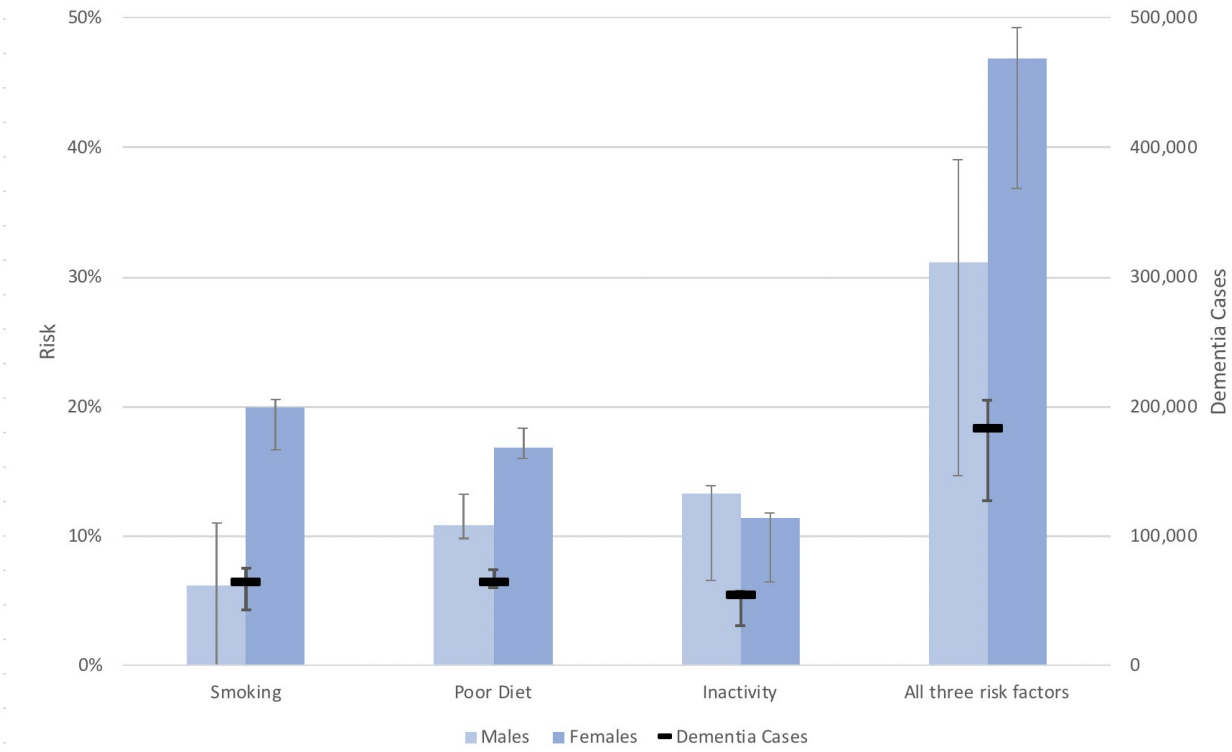

* With a smoking history equal to the median pack-years in the male population: former smoker quit 5+ years = 35.9, former smoker quit 0 to 5 years = 45.2, current smokers = 21.2

**Figure 4-2.** Relative risk ratios for health behaviours by mean risk exposure – females




* With a smoking history equal to the median pack-years in the female population: former smoker quit 5+ years = 26.9, former smoker quit 0 to 5 years = 29.5, current smokers = 13.8

**Figure 4-3.** Impact of unhealthy behaviours on 5-year dementia population risk and the number of incident dementia cases among Canadians aged 55 and older, 2013/14. Absolute 5-year dementia risk was 3.8% for males and 5.2% for females. The total predicted incident dementia cases accumulated between 2013/14 and 2018/19 is 452,000. Uncertainty intervals were created using the minimum and maximum estimates from five increasingly adjusted models (see **Appendix 4-5).**



poor diet. This estimate may range from 28.0% (127,000 cases) to 45.2% (205,000 cases), depending on the strength of the relationships between the health behaviours and dementia incidence. Physical inactivity was the leading risk factor in males (13.3% risk and 24,000 cases), while smoking was the leading risk factor in females (19.9% risk and 54,000 cases). Poor diet had a large impact on dementia risk among both males and females (males: 10.9% risk, 19,700 cases; females: 16.9% risk, 45,700 cases). Detailed estimates from the five increasingly adjusted models are available in **Appendix 4-5**.

## 4.5 DISCUSSION

In this study, a validated prediction algorithm was used to describe risk of dementia in the Canadian population 55+ years of age and to evaluate the impact of unhealthy lifestyle behaviours on the burden of dementia. There were an estimated 452,000 new cases of dementia between 2013/14 and 2018/19, of which an estimated 40% were associated with smoking, poor diet and physical inactivity. The largest contributors to the burden of dementia were physical inactivity among males and smoking among females.

At the individual level, disease prevention is best accomplished by reducing or eliminating the risk factors associated with the largest relative risks[11]. In contrast, prevention of disease in a population is best accomplished by reducing or eliminating the causes of disease incidence by targeting population subgroups with the largest disease burden. This is often more effective than targeting only those at high risk, as a large number of people with a small risk can give rise to more disease cases than a small number of people with a larger risk[11]. In the present study both dementia risk and burden increased with increasing age and was higher among women than men.

Individuals with less than a high school education had a higher risk than those with more education, however the largest burden of dementia was among those with post-secondary education. Although risks were generally higher among those with poor health behaviours, there is substantial burden among those with healthy lifestyles. As Canada and other countries develop population prevention strategies for dementia, they will need to consider not only which groups are at the highest risk, but also which groups have the largest disease burden.

A large proportion of dementia risk among both males and females was associated with unhealthy behaviours. Several studies have estimated the burden of modifiable risk factors on dementia risk using Levin's population attributable fraction (PAF) approach, which is commonly calculated using relative risk estimates from independent epidemiology studies or meta-analyses, risk factor prevalence from population health surveys, and outcome counts[12]. Using this method, a Lancet commission report on dementia estimated that 31% of Alzheimer's disease in the USA is attributable to seven modifiable risk factors (smoking, physical inactivity, diabetes, depression, midlife hypertension, midlife obesity, low educational attainment)[13]. Physical inactivity was estimated to have the largest PAF estimate, 21%, as it had both the largest population prevalence and relative risk (32.5% prevalence, RR = 1.82). One in ten cases of dementia were estimated to be attributable to smoking (20.6% prevalence, RR = 1.59). Similar estimates have been reported in Europe[13,14], while an Australian study[15] reported that 48.4% of dementia was attributed to the same seven risk factors, of which only 4.3% was attributed to smoking alone due to a lower smoking prevalence and use of a smaller relative risk estimate.

Although useful in jurisdictions with limited population-level data, PAF methods have many limitations, largely related to the use of summary measures of relative risk, exposure prevalence and outcome counts[16]. The multivariable predictive risk approach used in the present paper is conceptually similar to the PAF approach but has several advantages that largely come from the use of a single, individual-level data source[16–18]. Most importantly, individual-level data allows for direct estimation of dementia incidence among those who are exposed and unexposed to the risk factors of interest, rather than indirectly using PAF methods. The multivariable approach is therefore better able to capture the complexity of the disease-risk factor relationships, compared to the use of simplistic relative risk measures. For example, the effect of smoking on Alzheimer's disease risk in the Lancet report[13] was evaluated as a simple relative risk comparing current smokers to non-smokers (RR = 1.59), ignoring the effects of smoking history and the number of daily cigarettes consumed. In contrast, in the present study all health behaviours were modelled using flexible functions and allowed to vary by age. Smoking was modelled continuously using pack-years, with a smoking status variable to capture current, former and never smoking history. Estimation of effects within the population, rather than from meta-analyses, is also often advantageous, as relative risks from meta-analysis may not be generalizable to the population of interest if the included studies were performed within highly defined populations, differ in how they define the exposure of interest or if estimates were inconsistently adjusted for confounding factors[16].

Our study is also unique in that we improved upon the multivariable predictive risk approach by separately estimating the relative risks of the poor health behaviours instead of using the beta coefficients from the predictive model[17]. As DemPoRT was developed for predictive purposes

and includes mediating and downstream variables of the health behaviour – dementia risk relationships (e.g. BMI and functional measures), the beta coefficients from the predictive model are likely over-adjusted and do not necessarily describe the relationships of interest. Additionally, the use of uncertainty intervals provides an indication of how model misspecification may affect the estimates.

Poor diet, which has been hypothesized to increase incidence of dementia through obesity or oxidative stress[19,20], was estimated to have a large impact on the burden of dementia in Canada. In the present study, diet was evaluated using frequency of fruit and vegetable consumption, associated with a good diet, and potato and fruit juice consumption, which have been associated with poor outcomes[17,21,22]. To our knowledge, no one has evaluated the impact of diet on the population burden of dementia, perhaps due to the lack of population-level measures, focusing instead on the burden of obesity[13–15]. Dementia prevention efforts should consider strategies that include the promotion of healthy diets.

There are several limitations to this study. The total Canadian burden of dementia is likely underestimated as the CCHS does not consider individuals in long-term care homes, where dementia prevalence is high, and because respondents to health surveys tend to be healthier than the general population[23,24]. The total burden of dementia may also be underestimated as the DemPoRT algorithm was developed to predict physician-diagnosed dementia, and most individuals with dementia in Canada have not received a formal dementia diagnosis[25], and therefore were not captured. There is additionally potential for residual confounding in the estimation of the relative risks, however the use of uncertainty intervals provides some

information about the potential range of uncertainty. As the CCHS is a cross-sectional, self-report health survey, there is potential for misclassification error in the assessment of risk factor information. Alcohol consumption was not associated with development of dementia, consistent with findings from prospective cohort studies, however it is unclear whether this is due to selection bias, or if alcohol consumption is protective of dementia incidence[26].

## 4.6 CONCLUSION

These results suggest that unhealthy lifestyle behaviours including smoking, poor diet and physical inactivity have a large impact on the burden of disease from dementia. To our knowledge, this is the first time an algorithmic approach has been used to estimate the burden of dementia and the impact of unhealthy lifestyle behaviours on dementia incidence at the population level. This information can be used to inform population health planning and resource allocation, as well as the implementation of targeted primary prevention strategies.

## 4.7 FUNDING

This work is supported by the Canadian Institutes of Health Research operating grant MOP 142237, and by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). SF is supported by a doctoral award from the Canadian Institutes of Health Research (CIHR). The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES, the Ontario MOHLTC or CIHR is intended or should be inferred.

## 4.8 REFERENCES


1. Alzheimer Society of Canada. *Prevalence and Monetary Costs of Dementia in Canada*. Toronto; 2016.
2. Public Health Agency of Canada. *A Dementia Strategy for Canada: Together We Aspire*.
3. Sofi F, Valecchi D, Bacci D, et al. Physical activity and risk of cognitive decline: a meta-analysis of prospective studies. *J Intern Med*. 2011;269(1):107-117. doi:10.1111/j.1365-2796.2010.02281.x
4. Jiang X, Huang J, Song D, Deng R, Wei J, Zhang Z. Increased Consumption of Fruit and Vegetables Is Related to a Reduced Risk of Cognitive Impairment and Dementia: Meta-Analysis. *Front Aging Neurosci*. 2017;9:18. doi:10.3389/fnagi.2017.00018
5. Anstey KJ, von Sanden C, Salim A, O'Kearney R. Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. *Am J Epidemiol*. 2007;166(4):367-378. doi:10.1093/aje/kwm116
6. Kivipelto M, Solomon A, Ahtiluoto S, et al. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER): Study design and progress. *Alzheimer's Dement*. 2013;9(6):657-665. doi:10.1016/j.jalz.2012.09.012
7. Béland Y. Canadian Community Health Survey - Methodological overview. *Heal Reports*. 2002;13(3):9-14.
8. *Canadian Community Health Survey, 2013-2014: Annual Component; Study Documentation*. Ottawa, ON: Statistics Canada; 2017.
9. Rosella LC, Lebenbaum M, Li Y, Wang J, Manuel DG. Risk distribution and its influence on the population targets for diabetes prevention. *Prev Med (Baltim)*. 2014;58(1):17-21. doi:10.1016/j.ypmed.2013.10.007
10. O'Neill M, Kornas K, Rosella L. The future burden of obesity in Canada: a modelling study. *Can J Public Health*. August 2019. doi:10.17269/s41997-019-00251-y
11. Rose G. Sick Individuals and Sick Populations. *Int J Epidemiol*. 1985;14(1):32-38. doi:10.1093/ije/14.1.32
12. Levin ML, Bertell R. RE: "Simple estimation of population attributable risk from case-control studies." *Am J Epidemiol*. 1978;108(1):78-79.
13. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: An analysis of population-based data. *Lancet Neurol*.

2014;13(8):788-794. doi:10.1016/S1474-4422(14)70136-X

14. Mayer F, Di Pucchio A, Lacorte E, et al. An Estimate of Attributable Cases of Alzheimer Disease and Vascular Dementia due to Modifiable Risk Factors: The Impact of Primary Prevention in Europe and in Italy. *Dement Geriatr Cogn Dis Extra*. 2018;8(1):60-71. doi:10.1159/000487079
15. Ashby-Mitchell K, Burns R, Shaw J, Anstey KJ. Proportion of dementia in Australia explained by common modifiable risk factors. *Alzheimers Res Ther*. 2017;9(1):11. doi:10.1186/s13195-017-0238-x
16. Tanuseputro P, Perez R, Rosella L, et al. Improving the estimation of the burden of risk factors: an illustrative comparison of methods to measure smoking-attributable mortality. *Popul Health Metr*. 2015;13:5. doi:10.1186/s12963-015-0039-z
17. Manuel DG, Perez R, Sanmartin C, 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):1-27. doi:10.1371/journal.pmed.1002082
18. Manuel DG, Rosella LC, Hennessy D, Sanmartin C, Wilson K. Predictive risk algorithms in a population setting: an overview. *J Epidemiol Community Health*. 2012;66(10):859-865. doi:10.1136/jech-2012-200971
19. Napoli N, Shah K, Waters DL, Sinacore DR, Qualls C, Villareal DT. Effect of weight loss, exercise, or both on cognition and quality of life in obese older adults. *Am J Clin Nutr*. 2014;100(1):189-198. doi:10.3945/ajcn.113.082883
20. Lourida I, Soni M, Thompson-Coon J, et al. Mediterranean Diet, Cognitive Function, and Dementia. *Epidemiology*. 2013;24(4):479-489. doi:10.1097/EDE.0b013e3182944410
21. Hung H-C, Joshipura KJ, Jiang R, et al. Fruit and Vegetable Intake and Risk of Major Chronic Disease. doi:10.1093/jnci/djh296
22. Hansen L, Dragsted LO, Olsen A, et al. Fruit and vegetable intake and risk of acute coronary syndrome. *Br J Nutr*. 2010;104(2):248-255. doi:10.1017/S0007114510000462
23. Keyes KM, Rutherford C, Popham F, Martins SS, Gray L. How healthy are survey respondents compared with the general population?: Using survey-linked death records to compare mortality outcomes. *Epidemiology*. 2018;29(2):299-307. doi:10.1097/EDE.0000000000000775

24. Gorman E, Leyland AH, McCartney G, et al. Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *Am J Epidemiol*. 2014;180(9):941-948. doi:10.1093/aje/kwu207
25. Sternberg SA, Wolfson C, Baumgarten M. Undetected dementia in community-dwelling older people: the Canadian Study of Health and Aging. *J Am Geriatr Soc*. 2000;48(11):1430-1434.
26. Anstey KJ, Mack HA, Cherbuin N. Alcohol Consumption as a Risk Factor for Dementia and Cognitive Decline: Meta-Analysis of Prospective Studies. *Am J Geriatr Psychiatry*. 2009;17(7):542-555. doi:10.1097/JGP.0b013e3181a2fd07

# CHAPTER 5: Life expectancy and health care utilization of individuals with dementia in Ontario, Canada: A population-based study

## 5.0 PREFACE

Stacey Fisher, Douglas G Manuel, Sarah Spruin, Monica Taljaard, Geoffrey Anderson, Peter Tanuseputro. In this chapter, life expectancy and health care use of individuals with dementia in Ontario is investigated. It will be submitted with the following author list:

> Stacey Fisher, Douglas G Manuel, Sarah Spruin, Geoffrey Anderson, Monica Taljaard, Peter Tanuseputro

Supporting information is provided in appendices at the end of the thesis (Section 7.3) and at http://tiny.cc/osfDHCU. The use of data in this project was authorized under section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a Research Ethics Board.

SF was responsible for project conception, study design, data analysis, interpretation of the results, and drafting and revision of the manuscript. DGM and PT contributed to project conception, study design, and result interpretation. SS contributed to study design and data analysis. GA and MT contributed to the design of the study and result interpretation. All authors provided critical reviews of the manuscript and reviewed the final version.

## 5.1 ABSTRACT

**Introduction:** Life expectancy estimates for individuals with dementia vary widely, and there is limited information about the extent of formal health care use over the course of this disease. This study assesses the life expectancy and formal health care experience of individuals after diagnosis with dementia and describes the average annual health care utilization provided to patients from 2014 to 2017 in Ontario, Canada.

**Methods:** Using administrative health data of all individuals with dementia from April 2014 to March 2017 and period life table methodology, we examined the life expectancy and average health care use of individuals from dementia diagnosis to death. Dementia was ascertained by a validated case ascertainment definition, supplemented by information from long-term care, home care and complex continuing care assessments. Health care was categorized as inpatient care, outpatient care, home care or long-term care. The total average population-level health care use per year was evaluated by summing the total number of days of health care received by individuals in the study period, stratified by the type of care and stage of disease.

**Results:** Life expectancy at dementia diagnosis was 8.7 years and 9.8 years for men and women diagnosed prior to age 75, of which 3.7 years (42%) and 4.7 years (49%) was spent receiving care, respectively. Life expectancy was 4.4 and 5.2 years for men and women diagnosed after the age of 75, of which 2.2 years (50%) and 3.1 years (60%) was spent receiving health care, respectively. Women receive proportionally more long-term care and home care compared to men, while men receive more inpatient and outpatient care. The rate of health care use grows from the year prior to dementia diagnosis to end-of-life; in the year prior to dementia diagnosis individuals receive 20 days of health care per 100-person days, while those in the end-of-life

receive 79 days per 100 person-days. Among those at the end-of-life, individuals experience 55 LTC days per 100 person-days and 7.8 inpatient days per 100 person-days.

**Conclusion:** The burden of formal care for dementia on individuals and the health care system as a whole is substantial. After diagnosis with dementia, individuals receive health care for half of their remaining days of life, a large proportion of which is in long-term care facilities. The results of this study can be used by physicians to inform conversations with patients, their families and caregivers around what to expect after a dementia diagnosis, and by health care planners for population health planning.

## 5.2 INTRODUCTION

Dementia is the fifth leading cause of death globally, accounting for 2.4 million deaths in 2016[1]. Survival time varies considerably among patients as mortality has been associated with sex, age at diagnosis, dementia subtype, disease severity, comorbid conditions and functional impairment[2–8]. Estimates of average life expectancy after diagnosis therefore vary considerably with differences in study population characteristics.

Dementia is also associated with substantial morbidity and high health care use. After diagnosis, many individuals remain in private homes in the community, supported by informal care from family and friends that is often supplemented with formal home care services[9]. As the disease progresses however, many individuals will transition to long-term care facilities. In Canada, two thirds of those with dementia die in long-term care, of which 72% resided for more than a year prior to death[10]. Persons with dementia are also more likely to visit the emergency room and be admitted to hospital compared to similarly aged individuals without dementia[11–16], and hospital admission is associated with poorer health outcomes, including longer length of stay[17,18] and higher mortality[19]. Although care needs among individuals with dementia are known to be higher than for those without, the formal care experience from diagnosis to end-of-life is undescribed.

It is important for patients and their families, caregivers and health planners to have an understanding of life expectancy and the likely care experience of dementia patients from diagnosis to end-of-life. The objective of the current study was to examine life expectancy and the formal care experience of individuals after diagnosis with dementia to death using population-level administrative health data and life table methodology. This study uniquely adds

to the literature by providing a longitudinal, population perspective of the dementia care experience and by evaluating health care use across health sectors, including long-term care, home care, inpatient and outpatient care.

## 5.3 METHODS

This study was performed in Ontario, Canada using administrative health care data housed ICES, which has anonymized records of all health care encounters for every Ontario resident. Ontario is Canada's most populous province with over 14 million residents, of which 30% (4.3 million) are 55+ years of age[20].

### *5.3.1 Study design and data sources*

De-identified, individual-level administrative health data at ICES was used to identify all individuals with prevalent health care-diagnosed dementia at any time between April 1, 2014 to March 31, 2017. Dementia was defined using an ICES-derived cohort, which uses a validated case ascertainment definition based on physician billing, hospitalization records, and drug dispensing records[21]. Due to known underdiagnosis of dementia,[22,23] this definition was supplemented with information from long-term care, home care and complex continuing care (CCC) assessments (dementia flag AND Cognitive Performance Scale score $\geq$2) using the Continuing Care Reporting System (CCRS) and the Resident Assessment Instrument - Home Care database, respectively. CPS has been validated against the Mini-Mental State Exam (MMSE)[24–27]. Dementia diagnosis date was assigned as the date of first dementia ascertainment by the above definition, looking back to January 1991. Individuals were excluded if they were younger than 55 years of age at the time of diagnosis, were not eligible for Ontario's health care

insurance plan one year prior to dementia diagnosis or study start, or were older than 105. Age, sex and date of death were obtained from the Registered Persons Database.

Health care utilization was measured in days of care and categorized in to four types: long-term care, home care, inpatient care, and outpatient care. Long-term care use was defined as days in a 24-hour residential care home and identified using the CCRS. Home care was defined as care received from publicly-funded home care services and identified using the Home Care Database. Inpatient care was defined as care received in a hospital, including emergency room visits and same-day surgeries or procedures, inpatient mental health care, complex continuing care, or inpatient rehabilitation services. These inpatient care days were identified using the Discharge Abstract Database, the Same-Day Surgery database, the Ontario Mental Health Reporting System, the CCC database and the National Rehabilitation Reporting System. Outpatient care was identified using Ontario Health Insurance Plan (OHIP) billing records from the OHIP Claims database. More information about each data source is available in **Table 5-1**. Days were never double counted; if an individual received more than one type of care on a single day a hierarchy was used (inpatient care > long-term care > home care > outpatient care).

### *5.3.2 Life table analyses*

Life expectancy and the average health care utilization of individuals from dementia diagnosis to death was estimated using complete period life table methodology[28]. This method uses a cross-sectional sample of individuals with dementia to create a hypothetical cohort and assumes that all individuals are subject to the same time since diagnosis-specific mortality rates observed among individuals in the cross-sectional study period.

**Table 5-1.** Databases used to retrieve health care utilization information

| Type of Care | Database | Description |
|---|---|---|
| Long-Term Care | Continuing Care Reporting System (CCRS) | Resident information from over 600 publicly-funded residential care homes that provide 24-hour nursing and personal care |
| Home Care | Home Care Database (HCD) | Data from the Ontario Association of Community Care Access Centers (OACCAC), responsible for coordinating publicly-funded home care services |
| Inpatient Care | Discharge Abstract Database (DAD) | In-patient hospital stay information |
| | Same-Day Surgery (SDS) | Information about same-day surgeries and procedures |
| | National Ambulatory Care Reporting System (NACRS) | Records of emergency room visits |
| | Ontario Mental Health Reporting System (OMHRS) | Records of inpatient mental health care services |
| | Complex Continuing Care (CCC) | Information on hospitalized patients who are deemed to be in a non-acute state. These facilities (or units within a larger facility may also provide palliative care to individuals who are at the end-of-life. |
| | National Rehabilitation Reporting System (NRS) | Data from participating adult inpatient rehabilitation facilities and programs across Ontario |
| Outpatient Care | Ontario Health Insurance Plan (OHIP) Claims Database | OHIP billings made by physicians (which includes both in-patient and outpatient services) and by other health professionals including oral surgeons, chiropractors, optometrists and physiotherapists |

To estimate life expectancy, mortality rates associated with individual years since diagnosis intervals were evaluated using only data within the study period, April 1, 2014 to March 31, 2017. For example, an individual diagnosed with dementia in April 2010 who died in April 2016 contributed to the calculation of the mortality rates for the 4-5 and 5-6 years since diagnosis time intervals. These mortality rates were stratified by sex and age at diagnosis (<75 years at diagnosis or 75+ years at diagnosis). Chiang's method for ending life tables was used at the final time since diagnosis interval of 15+ years[28]. Average health care utilization of individuals with dementia from diagnosis to death was evaluated using a modified version of Sullivan's life table method[29], where the average fraction of life lived in a particular setting was used in place of the prevalence of an unhealthy state. Average fraction of life lived in long-term care, home care, inpatient care, outpatient care and receiving no care was evaluated as the average proportion of time spent in the interval receiving each type of care, stratified by sex and age, using only data within the study period.

*5.3.3 Population-level health care use analyses*

Rates of population-level health care use was evaluated by summing the total number of days of care received by individuals in the study period, divided by the total person-days, stratified by the type of care (long-term care, home care, inpatient and outpatient) and stage of disease (pre-diagnosis, diagnosis, progression, end-of-life). Pre-diagnosis was defined as the 1-year period prior to dementia diagnosis. Diagnosis was defined as the 1-year period after dementia diagnosis. Progression was defined as all time between 1-year post-diagnosis and 1-year prior to death. End-of-life was defined as the 1-year period prior to death. If an individual died within 1-year of

diagnosis, health care utilization during that time was classified as occurring in the end-of-life stage.

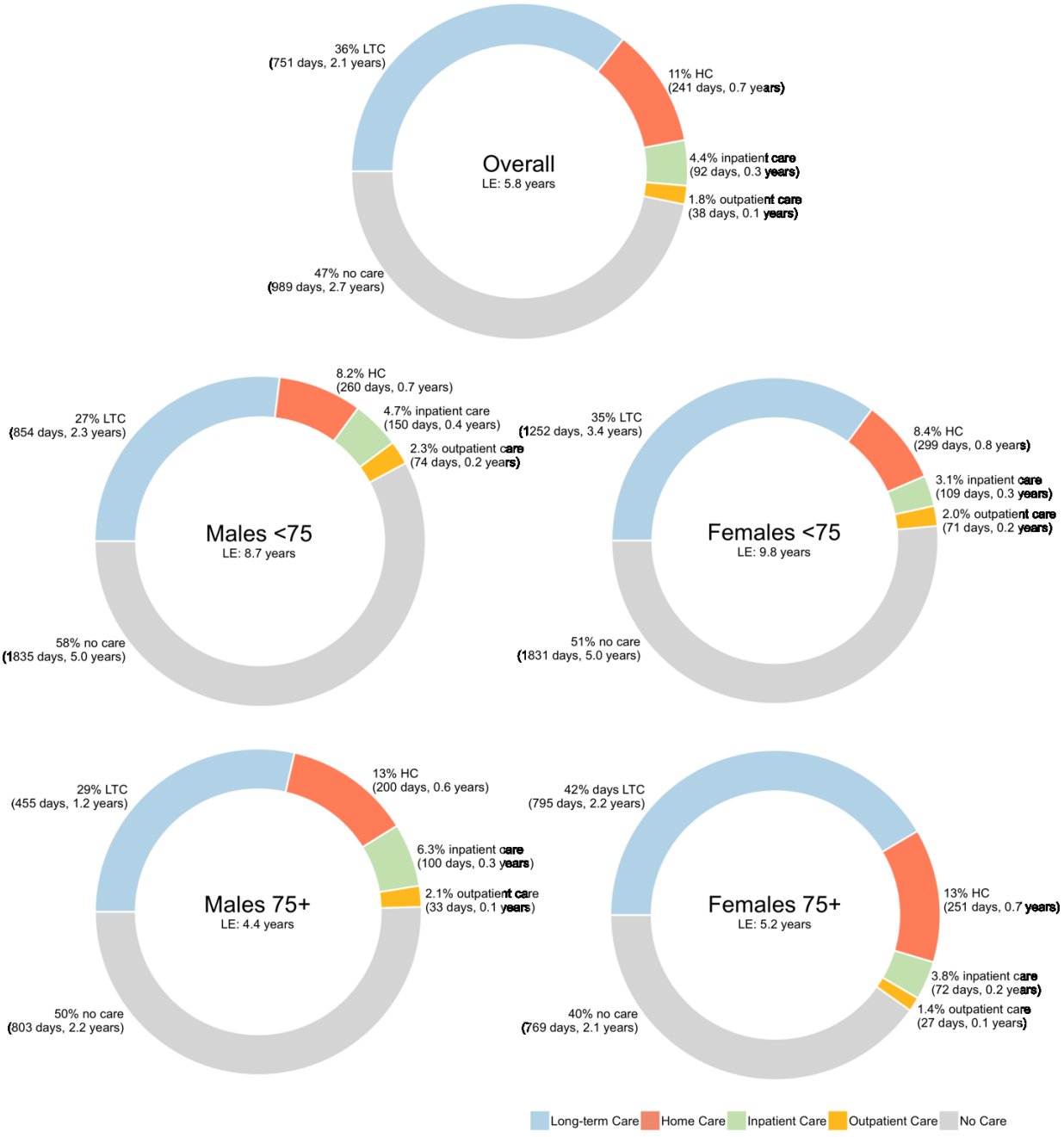
The annual population-level health care use by Ontarians with dementia, and those in the year prior to dementia diagnosis, was evaluated by summing the total number of days of care received by individuals in the study period, stratified by the type of care (long-term care, home care, inpatient and outpatient), stage of disease (pre-diagnosis, diagnosis, progression, end-of-life), and year (2014/15, 2015/16, 2016/17). Health care days were then averaged across study years.

## 5.4 RESULTS

There were 94,000 deaths among individuals with dementia in Ontario from 2014 to 2017 within 555,000 person-years of follow-up, therefore the overall mortality rate was 169 per 1000 person-years. In the year of diagnosis, the overall mortality rate was 172 per 1000 person-years, decreasing to 136 per 1000 person-years in the second year, and increasing to 207 per 1000 person-years for those who survive to 15 or more years since diagnosis. This pattern is similar for both males and females diagnosed prior to or after age 75.

Life expectancy and the average number of days individuals with dementia received long-term care, home care, inpatient care, outpatient care and no care in a year, stratified by sex and age at diagnosis, is displayed in **Figure 5-1.** Overall life expectancy at dementia diagnosis was 5.8 years. Among men and women diagnosed prior to age 75, life expectancy was 8.7 and 9.8 years, of which an average of 3.7 years (1,338 days; 42%) and 4.7 years (1,731 days; 49%) was spent

**Figure 5-1.** Life expectancy and average number of days receiving long-term care, home care, inpatient care, outpatient care and no care from diagnosis to death overall, and stratified by sex and age at diagnosis



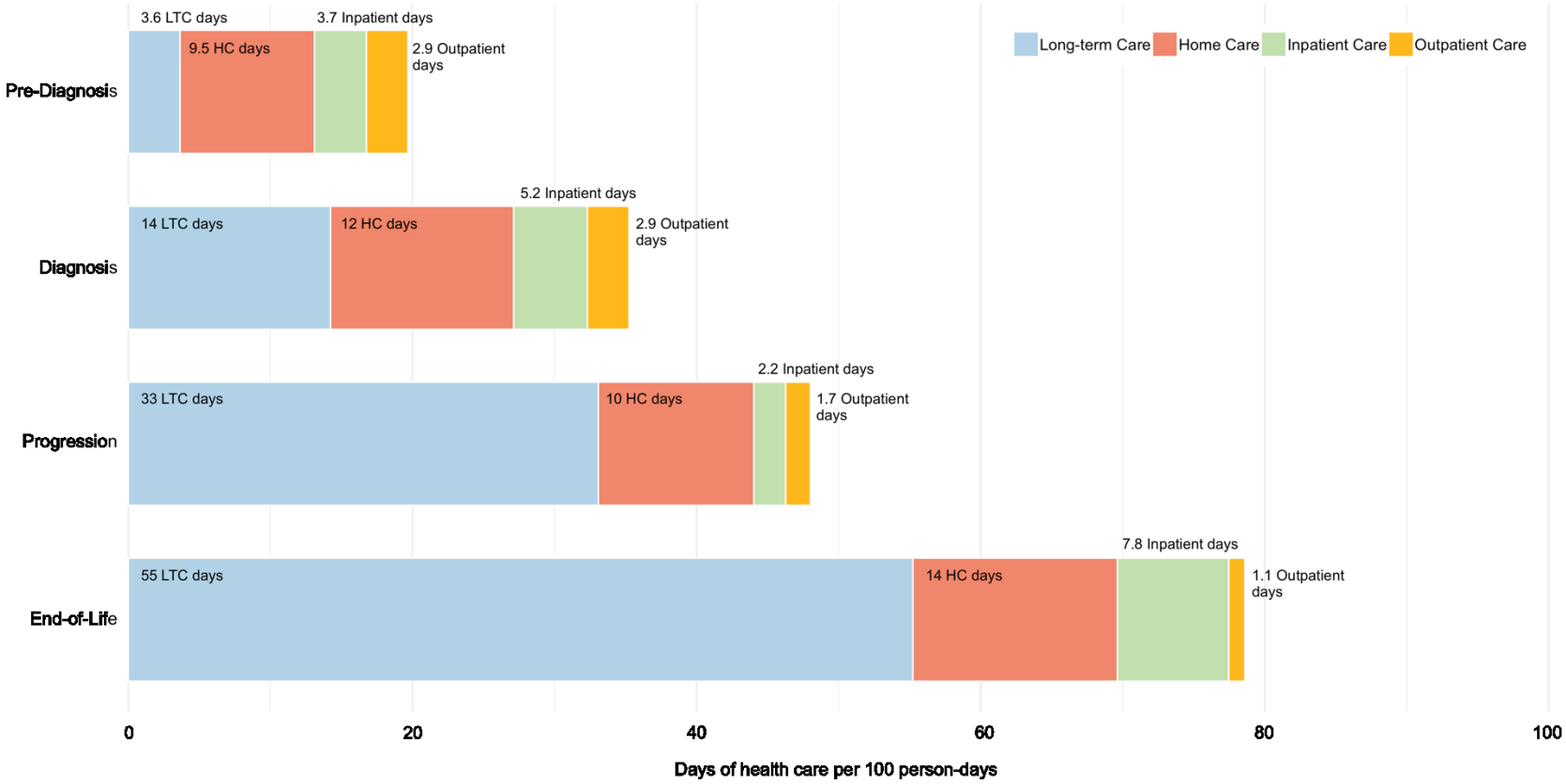

Abbreviations: HC, home care; LE, life expectancy; LTC, long-term care

receiving care. Among men and women diagnosed after age 75, life expectancy was 4.4 and 5.2 years, of which an average of 2.2 years (788 days; 50%) and 3.1 years (1,145 days; 60%) was spent receiving care. In both age groups, women had longer life expectancies and received proportionally more long-term care and home care than men, while men experienced proportionally more inpatient and outpatient care. Survival curves can be found in **Appendix 5-1.** Years since diagnosis-specific mortality rates, life expectancy and days of health care overall and stratified by sex and age at diagnosis are available in **Appendix 5-2**. Full life tables are available at http://tiny.cc/osfDHCU.

The rate of health care use grows from the year prior to dementia diagnosis to the end-of-life **(Figure 5-2)**. In the pre-diagnosis, diagnosis, progression and end-of-life stages individuals receive 20, 35, 48 and 79 days of health care per 100 person-days, respectively. The rate of long-term care use increases with disease progression, from 3.6 days per 100 person-days among those in the year prior to diagnosis, to 55 days per 100 person-days among those in the year prior to death. Home care and inpatient care use are highest in the year of diagnosis (13 days and 5.2 days per 100 person-days, respectively) and at the end-of-life (14 days and 7.8 days per 100 person-days, respectively).

The total health care utilization of individuals with dementia in Ontario in an average year, stratified by stage of disease, is displayed in **Appendix 5-3**. An estimated 34.1 million days of health care are received by individuals with dementia in Ontario in an average year. Another 2.5 million days of health care are received by those who will be diagnosed with dementia within a year. A large proportion of all care is received by individuals in the year prior to death (22%,

**Figure 5-2.** Average number of days of long-term care, home care, inpatient care and outpatient care received by individuals with dementia in Ontario per 100 person-days, by stage of disease




Abbreviations: HC, home care; LTC, long-term care

8,140,000 days). However, the majority of health care in this population, 61% (22,300,000 days), is received by individuals in the progression stage of the disease, defined as a year after diagnosis to a year prior to death, which lasts an average of 3.8 years. Long-term care makes up 18% of all care received by those in the year prior to diagnosis, increasing to 70% among those in the progression and end-of-life stages.

## 5.5 DISCUSSION

Life expectancy from dementia diagnosis and subsequent health care use was evaluated in Ontario, Canada. Overall life expectancy was 5.8 years and was longer for women and those diagnosed prior to the age of 75. Individuals with dementia spent, on average, half of their life receiving formal care, a large proportion of which was in long-term care facilities. Women diagnosed prior to age 75 were found to spend an average of 3.4 years (35% of their 9.8 year life expectancy) in long-term care. The burden of dementia on the health care system is significant throughout the course of the disease.

Dementia is a progressive and chronic disease. Individuals become increasingly frail as their cognitive and functional abilities decline, leading to higher health care needs that often require long-term care admission. More persons with dementia are remaining in the community than in the past[30], and there have been calls to continue to enhance the capacity of community-based care[31,32]. Caring for individuals with dementia at home reduces system spending and is consistent with the desire of most people to age at home[33–35]. However, a recent Ontario study of people newly admitted to LTC found that the majority had high levels of need and could not be cost-effectively cared for in the community[36]. Informal caregivers often experience shame during the

transition of their loved one to long-term care, and describe the transition as traumatic[37], suggesting that carers may be unaware that most individuals with dementia eventually require institutionalization due to their high care requirements[37,38]. It is important for patients, their caregivers and families to be informed of the progressive nature of the disease and the likely need for long-term care.

From the health care system perspective, the burden of dementia is significant, and grows from the year prior to dementia diagnosis to end-of-life. Although the highest use of health care is in their last year of life, a considerable amount of care is provided to individuals along the disease trajectory, including in the year prior to diagnosis. Furthermore, we have shown that a large proportion of the care provided to those with dementia is designed to meet the needs of individuals to perform basic day-to-day functions (i.e. activities of daily living (ADL) and instrumental ADLs). This care – delivered through the long-term care and home care sectors – becomes an increasingly prominent facet as death approaches, with more than 2/3 of days in the last year of life having such service delivery.

The population perspective of this study produces uniquely relevant life expectancy estimates from health care diagnosis of dementia. Median survival time from dementia diagnosis was reported to range from 3.2 to 6.6 years in a recent review[2]. This variation is likely due in large part to differences in study populations, as many patient factors have been associated with dementia survival. Differences in mean age across studies, for example, can lead to substantial differences in survival estimates, as younger age at diagnosis is associated with longer survival[4,5]. Additionally, survival studies performed using highly selected groups of patients,

such as those that have been referred to specialist clinics, are subject to substantial selection bias as these patients are not representative of the general dementia population. Study design differences, including length of follow-up and inclusion of only prevalent cases, can also have a substantial impact on survival estimates[39,40]. In the present study, life expectancy from dementia diagnosis was estimated using mortality rates derived from all incident and prevalent dementia cases in Ontario during the study period. Selection bias is therefore minimal, as all individuals with dementia were included. Loss to follow-up and emigration are not sources of bias as dementia patients were not followed, rather annual mortality rates were calculated, and life expectancy estimated using life table methodology. The present results are therefore directly applicable to newly diagnosed individuals with dementia in the general population in Ontario and may be generalized to populations with similar characteristics.

A major strength of this study is the use of large, linked population-level data of all health care encounters of all individuals with dementia in Ontario. The results of this study therefore reflect the life expectancy and health care experience of the Ontario dementia population. As period life table methodology was used and mortality rates and health care utilization were estimated from 2014 to 2017, the results are based on recent data that is representative of the care currently experienced by those with dementia. These estimates are therefore more indicative of the experience of patients diagnosed with dementia in recent years compared to life expectancy estimates from older cohort studies. However, as dementia diagnosis was evaluated using health care data rather than explicit physician evaluation, and as dementia is generally under-recorded in primary care[22,41], diagnosis dates may be mis-specified. This may have led to an underestimation of life expectancy and underestimation of health care use.

## 5.6 CONCLUSION

The burden of health care for dementia on individuals and the health care system is substantial. To our knowledge this is the first study to estimate the proportion of life spent receiving care across health sectors in dementia patients from diagnosis to death. Physicians can use these results of this study to inform conversations with patients, their families and caregivers around what to expect after a dementia diagnosis. As these results are population-based and representative of the Ontario dementia population, these results can also be used for population health planning for dementia.

## 5.7 REFERENCES


1. GBD 2016 Dementia Collaborators E, Szoeke CEI, Vollset SE, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 2019;18(1):88-106. doi:10.1016/S1474-4422(18)30403-4
2. Brodaty H, Seeher K, Gibson L. Dementia time to death: a systematic literature review on survival time and years of life lost in people with dementia. *Int psychogeriatrics*. 2012;24(7):1034-1045. doi:10.1017/S1041610211002924
3. Strand BH, Knapskog A-B, Persson K, et al. Survival and years of life lost in various aetiologies of dementia, mild cognitive impairment (MCI) and subjective cognitive decline (SCD) in Norway. Ginsberg SD, ed. *PLoS One*. 2018;13(9):e0204436. doi:10.1371/journal.pone.0204436
4. Xie J, Brayne C, Matthews FE, Medical Research Council Cognitive Function and Ageing Study collaborators. Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up. *BMJ*. 2008;336(7638):258-262. doi:10.1136/bmj.39433.616678.25
5. Waring SC, Doody RS, Pavlik VN, Massman PJ, Chan W. Survival among patients with dementia from a large multi-ethnic population. *Alzheimer Dis Assoc Disord*. 19(4):178-183. http://www.ncbi.nlm.nih.gov/pubmed/16327343. Accessed August 21, 2019.
6. Todd S, Barr S, Roberts M, Passmore AP. Survival in dementia and predictors of mortality: a review. *Int J Geriatr Psychiatry*. 2013;28(11):1109-1124. doi:10.1002/gps.3946
7. Mueller C, Soysal P, Rongve A, et al. Survival time and differences between dementia with Lewy bodies and Alzheimer's disease following diagnosis: A meta-analysis of longitudinal studies. *Ageing Res Rev*. 2019;50:72-80. doi:10.1016/J.ARR.2019.01.005
8. Price A, Farooq R, Yuan J-M, Menon VB, Cardinal RN, O'Brien JT. Mortality in dementia with Lewy bodies compared with Alzheimer's dementia: a retrospective naturalistic cohort study. *BMJ Open*. 2017;7(11):e017504. doi:10.1136/bmjopen-2017-017504
9. Wong SL, Gilmour H, Ramage-Morin PL. Alzheimer's disease and other dementias in Canada. *Heal reports*. 2016;27(5):11-16. http://www.ncbi.nlm.nih.gov/pubmed/27192206.

Accessed August 22, 2019.

10. Hsu A, Chalifoux M, Spruin S, Bronskill S, Manuel D, Tanuseputro P. Health care among people with dementia at the end of life. *Submitt to Alzheimer's Dement*.
11. Tonelli M, Wiebe N, Straus S, et al. Multimorbidity, dementia and health care in older people:a population-based cohort study. *C Open*. 2017;5(3):E623-E631. doi:10.9778/cmajo.20170052
12. Zhao Y, Kuo T-C, Weir S, Kramer MS, Ash AS. Healthcare costs and utilization for Medicare beneficiaries with Alzheimer's. *BMC Health Serv Res*. 2008;8(1):108. doi:10.1186/1472-6963-8-108
13. Feng Z, Coots LA, Kaganova Y, Wiener JM. Hospital and ED use among Medicare beneficiaries with dementia varies by setting and proximity to death. *Health Aff (Millwood)*. 2014;33(4):683-690. doi:10.1377/hlthaff.2013.1179
14. LaMantia MA, Stump TE, Messina FC, Miller DK, Callahan CM. Emergency Department Use Among Older Adults With Dementia. *Alzheimer Dis Assoc Disord*. 2016;30(1):35-40. doi:10.1097/WAD.0000000000000118
15. Maxwell CJ, Amuah JE, Hogan DB, et al. Elevated Hospitalization Risk of Assisted Living Residents With Dementia in Alberta, Canada. *J Am Med Dir Assoc*. 2015;16(7):568-577. doi:10.1016/j.jamda.2015.01.079
16. Phelan EA, Borson S, Grothaus L, Balch S, Larson EB. Association of Incident Dementia With Hospitalizations. *JAMA*. 2012;307(2):165. doi:10.1001/jama.2011.1964
17. Connolly S, O'Shea E. The impact of dementia on length of stay in acute hospitals in Ireland. *Dementia*. 2015;14(5):650-658. doi:10.1177/1471301213506922
18. Draper B, Karmel R, Gibson D, Peut A, Anderson P. The Hospital Dementia Services Project: age differences in hospital stays for older people with and without dementia. *Int Psychogeriatrics*. 2011;23(10):1649-1658. doi:10.1017/S1041610211001694
19. Lehmann J, Michalowsky B, Kaczynski A, et al. The Impact of Hospitalization on Readmission, Institutionalization, and Mortality of People with Dementia: A Systematic Review and Meta-Analysis. *J Alzheimer's Dis*. 2018;64(3):735-749. doi:10.3233/JAD-171128
20. Statistics Canada. Table 17-10-0005-01 - Population estimates on July 1st, by age and sex. doi:https://doi.org/10.25318/1710000501-eng

21. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimer's Dis*. 2016;54(1):337-349. doi:http://dx.doi.org/10.3233/JAD-160105
22. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence. *Aging Ment Health*. 2011;15(8):978-984. doi:10.1080/13607863.2011.596805
23. Kosteniuk JG, Morgan DG, O'Connell ME, et al. Incidence and prevalence of dementia in linked administrative health data in Saskatchewan, Canada: a retrospective cohort study. *BMC Geriatr*. 2015;15(1):73. doi:10.1186/s12877-015-0075-3
24. Paquay L, Lepeleire J De, Schoenmakers B, Ylieff M, Fontaine O, Buntinx F. Comparison of the diagnostic accuracy of the Cognitive Performance Scale (Minimum Data Set) and the Mini-Mental State Exam for the detection of cognitive impairment in nursing home residents. *Int J Geriatr Psychiatry*. 2007;22(4):286-293. doi:10.1002/gps.1671
25. Hartmaier SL, Sloane PD, Guess HA, Koch GG, Mitchell CM, Phillips CD. Validation of the Minimum Data Set Cognitive Performance Scale: agreement with the Mini-Mental State Examination. *J Gerontol A Biol Sci Med Sci*. 1995;50(2):M128-33. http://www.ncbi.nlm.nih.gov/pubmed/7874589. Accessed December 5, 2018.
26. Morris JN, Fries BE, Mehr DR, et al. MDS Cognitive Performance Scale. *J Gerontol*. 1994;49(4):M174-82. http://www.ncbi.nlm.nih.gov/pubmed/8014392. Accessed December 5, 2018.
27. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the Minimum Data Set for nursing home resident assessment and care screening (MDS). *Gerontologist*. 1995;35(2):172-178. http://www.ncbi.nlm.nih.gov/pubmed/7750773. Accessed December 5, 2018.
28. Chiang C. *The Life Table and Its Applications*. Malabar, Florida: Robert E. Krieger Publ. Co.; 1984.
29. Sullivan DF. A single index of mortality and morbidity. *HSMHA Health Rep*. 1971;86(4):347-354. http://www.ncbi.nlm.nih.gov/pubmed/5554262. Accessed August 15, 2019.
30. The Alzheimer Society of Canada. *Rising Tide: The Impact of Dementia on Canadian*

*Society*.; 2010.

31. Canadian Nurses Association. *Dementia in Canada: Recommendations to Support Care for Canada's Aging Population*.; 2016. www.cna-aiic.ca. Accessed August 20, 2019.
32. National Academies of Sciences Engineering and Medicine. *Families Caring for an Aging America*. Washington, D.C; 2016.
33. Wiles JL, Leibing A, Guberman N, Reeve J, Allen RES. The Meaning of "Aging in Place" to Older People. *Gerontologist*. 2012;52(3):357-366. doi:10.1093/geront/gnr098
34. Groff AC, Colla CH, Lee TH. Days Spent at Home - A Patient-Centered Goal and Outcome. *N Engl J Med*. 2016;375(17):1610-1612. doi:10.1056/NEJMp1607206
35. Barnato AE, Herndon MB, Anthony DL, et al. Are Regional Variations in End-of-Life Care Intensity Explained by Patient Preferences? *Med Care*. 2007;45(5):386-393. doi:10.1097/01.mlr.0000255248.79308.41
36. Tanuseputro P, Hsu A, Kuluski K, et al. Level of Need, Divertibility, and Outcomes of Newly Admitted Nursing Home Residents. *J Am Med Dir Assoc*. 2017;18(7):616-623. doi:10.1016/j.jamda.2017.02.008
37. Kelsey SG, Laditka SB, Laditka JN. Caregiver perspectives on transitions to assisted living and memory care. *Am J Alzheimers Dis Other Demen*. 2010;25(3):255-264. doi:10.1177/1533317509357737
38. Afram B, Verbeek H, Bleijlevens MHC, Hamers JPH. Needs of informal caregivers during transition from home towards institutional care in dementia: a systematic review of qualitative studies. *Int Psychogeriatrics*. 2015;27(6):891-902. doi:10.1017/S1041610214002154
39. Habbema JDF, Dippel DWJ. Survivors-only bias in estimating survival in Alzheimer's disease and vascular dementias. *Neurology*. 1986;36(7):1009-1009. doi:10.1212/WNL.36.7.1009-b
40. Wolfson C, Wolfson DB, Asgharian M, et al. A reevaluation of the duration of survival after the onset of dementia. *N Engl J Med*. 2001;344(15):1111-1116. doi:10.1056/NEJM200104123441501
41. Kosteniuk JG, Morgan DG, O'Connell ME, et al. Incidence and prevalence of dementia in linked administrative health data in Saskatchewan, Canada: a retrospective cohort study. *BMC Geriatr*. 2015;15(1):73. doi:10.1186/s12877-015-0075-3

# CHAPTER 6: DISCUSSION

The purpose of this thesis was to support population health decision-making for dementia, including informing health resource planning and the development of preventative strategies. Together, this work provides population health planners and policy-makers with a tool to project dementia incidence and important information about the burden and health care use of dementia. I first pre-specified the analysis plan and predictor variables for the development of a predictive algorithm for dementia in the community (Chapter 2). Pre-specification is an important step in the development of predictive algorithms to avoid data-driven approaches and type 1 error which can lead to overfitting and poor performance in external validation. Next, this model, the Dementia Population Risk Tool (DemPoRT), was developed and internally validated using self-reported information from health surveys, and without the use of neuropsychological or genetic testing widely used in clinical dementia algorithms (Chapter 3). We demonstrated that DemPoRT is well-calibrated across a wide range of policy-relevant subgroups, making it appropriate for population health planning. An example of this is demonstrated in Chapter 4, where I described the distribution of dementia risk across policy-relevant subgroups and found that a large proportion of the burden of disease from dementia in Canada is associated with unhealthy lifestyle behaviours. Lastly, I explored the previously-undescribed formal care experience of those with dementia from diagnosis to death, and the health care burden across health sectors and by stage of disease (Chapter 5).

## 6.1 GENERAL LIMITATIONS

There are a number of general limitations to this thesis that are important to acknowledge, related to misclassification of dementia and the DemPoRT predictor variables, as well as considerations related to the causal inference from predictive models.

*Misclassification of dementia*

Firstly, all studies included in this thesis used a case ascertainment algorithm for physician-diagnosed dementia[1]. When validated against family physician EMR data this definition was found to have a sensitivity of 79.3% and a specificity of 99.1%. The positive and negative predictive values were 80.4% and 99.0%, respectively, with a total prevalence of 4.4% in the validation sample population, which is similar to the application population. Therefore, few individuals (0.9%) who did not have dementia were falsely identified as having dementia, and few (1.0%) of those identified as not having dementia actually had dementia. However, approximately 20.7% of those who have dementia were not being identified by the algorithm, and 19.6% of those identified by the algorithm didn't actually have dementia. Family physician EMR data, however, is not a perfect gold standard for physician-diagnosed dementia. In the validation study, the majority of the false positives were due to problems inherent with using EMR as the gold standard[1]. Therefore, the specificity, as well as the positive and negative predictive values may be underestimated. To decrease the number of false negatives and increase sensitivity for the investigation of life expectancy and health care use in Chapter 5, the validated dementia case ascertainment definition was supplemented with information from home care and long-term care assessments. This was not done for the earlier work as the case ascertainment priority for development of the DemPoRT algorithm was high specificity, as false positives are

more of a thread to algorithm development than false negatives, and because the home care data was not available for prior to 2008.

Regardless, there is likely a sizable amount of dementia misclassification that may be biasing this work. As administrative data was used to identify dementia cases, only health-care diagnosed dementia can be identified, and individuals who do not seek care are not captured. Dementia is known to be generally underdiagnosed[2,3] as individuals are often reluctant to seek care for issues with cognitive functioning, and those with informal support in the community are able to manage at home for longer. Therefore, individuals with less severe dementia symptoms, those who have significant home supports, or have poor access to health care may be missed or identified later in the disease trajectory. This outcome misclassification likely reduces the discrimination and calibration of the DemPoRT algorithm, although performance remains acceptable. Dementia burden estimates also therefore only include health care-diagnosed dementia and underestimate total burden.

Health care-diagnosed dementia may also be under ascertained by administrative data as physician service claims allow for only one diagnostic code per claim, and individuals with dementia often have many comorbidities that may be treated in a single visit. Also, physicians are that not fee-for-service may not always submit comprehensive billing claims. Although the use of administrative data for identifying dementia is less accurate than objective testing combined with expert judgement, the use of administrative data is less costly, more practical, and provides an opportunity to identify individuals with dementia at the population-level.

We were also not able to differentiate between different subtypes of dementia. The validation study of the dementia ascertainment algorithm used in this work was not able to describe the distribution of subtypes identified, therefore it is unclear if some subtypes are ascertained better than others. Some risk factors may be specific to particular subtypes, and prevention strategies may work better in some subtypes compared to others. This has potential implications for dementia risks ascertained by DemPoRT, and for life expectancy estimates and health care use from dementia diagnosis. The literature generally suggests that Alzheimer's disease and vascular dementia share many of the same risk factors, however dementia with Lewy bodies and frontotemporal dementia are less studied. Life expectancy risk is known to differ by dementia subtype[4], and therefore total health care use will also likely differ. Identification of dementia subtype often requires neuroimaging and biomarker assessment, however, and is not practical at the population-level. Subtype specific analyses are also complicated by the fact that many patients present with mixed dementia[5–7].

With the new Canadian Longitudinal Study on Aging (CLSA), there is potential in the future to evaluate the dementia case ascertainment algorithm with a gold standard that captures all dementia, rather than just health care-diagnosed disease. The CLSA is a 20-year longitudinal study of over 50,000 community dwelling adults age 45 to 85 years with collection of biological, psychological, social, lifestyle and economic characteristics, including collection of detailed cognitive measures to evaluate memory and executive functioning over time[8]. Baseline data was collected from 2012 to 2015. The CLSA can also be used to inform re-calibration of dementia burden estimates from the existing algorithms and investigate dementia subtypes.

*Misclassification of predictor variables*

Another limitation of this work is misclassification of the DemPoRT predictor variables, which can lead to deterioration of model performance. All CCHS variables are collected by self-report. Although demographic characteristics like education are likely measured with little error, it has been established that self-reported health behaviour variables such as diet, physical activity and alcohol consumption may be ascertained with substantial bias when evaluated via self-report[9]. Diet is also not well captured in the CCHS, as limited information is collected, and food frequency is assessed without consideration of serving sizes. The CCHS also only collects information about leisure time physical activity, and not physical activity performed at home, work or from active transportation. Additionally, predictor variables were only evaluated at baseline, as necessitated by the cross-sectional nature of the CCHS. Therefore, any changes in exposure would not be captured, such as quitting smoking, adoption of improved physical activity or eating habits, or reduction in alcohol consumption. Improved ascertainment of health behaviour variables would likely improve performance of the DemPoRT model, however, it is difficult to ascertain population-level information about health behaviours from means other than from national health surveys, and the DemPoRT model performs well in spite of these limitations.

*Causal inference from predictive models*

Predictive models are not designed for causal inference. As the goal is prediction, the focus of should be on the overall risk calculation. As variables are included in a predictive model based on predictive performance (and not statistical uncertainty) and mediator variables are often included, individual beta coefficients should not be interpreted. However, if a predictive model

has been determined to be discriminating and well-calibrated in a particular population, risk estimates from the model can be used at the population level to inform the development of prevention strategies, or at the individual level to inform decisions about lifestyle modification. Population-level estimates of disease risk are important for population level planning; there is also substantial interest in using the DemPoRT algorithm for this purpose.

Interpretation of the impact of unhealthy behaviours on the burden of dementia in Chapter 4 must also be performed with caution. In this study, the DemPoRT model was used to estimate the population burden of dementia and the potential impact of unhealthy behaviours. The impact of risk factors on disease burden is most commonly performed using PAF methods, however in this study we employed a multivariable modeling approach that has been shown to have many advantages over PAF methods, as discussed previously in the literature[10] and in Chapter 4. However, for the hazards from the predictive model to be used for this purpose, the predictive model needs to be developed including only exposures and covariates that are strongly suspected of being causally associated with the outcome[10,12] usually at the expense of predictive performance. In this study, we improved upon the existing multivariable approach by using beta coefficients from separate Cox proportional hazard regression models, integrated into predictive model using the method described in Appendix 4-2. This method is novel and has not be fully tested, however we are confident that it is an improvement over the usual multivariable approach using a predictive model that may include non-causal factors.

With any approach, however, one needs to be cautious about making causal inferences from the results, due to the potential for residual confounding. When using a PAF approach, this

confounding arises from the relative risk estimate used to summarize the relationship between the risk factor and outcome, which is usually pulled from a meta-analysis. As the multivariable approach uses individual-level data rather than summary estimates, risk factors are able to be considered more complexly which may reduce residual confounding, however the accuracy of the estimates are dependent on proper model specification and is conditional on model assumptions.

## 6.2 IMPLICATIONS FOR POLICY AND PRACTICE

Dementia risk and health behaviour attribution results from Chapter 4, and dementia health care use from Chapter 5 can be used to inform health care planning for dementia, and support Canada's national dementia strategy[13], in many ways. One of the pillars identified by the strategy as essential for effective implementation is enhanced surveillance and data, which is necessary to understand the scope of dementia and to help focus research efforts and health care resources. The relative attribution of health behaviours to dementia burden informs the development of prevention strategies, and baseline population risk and the future number of incident cases described across subgroups can be used to target these strategies. Estimates of total health care use from diagnosis to death can be used by physicians in conversations with those newly-diagnosed with dementia, and the rate of health care across sectors and disease stages informs resource allocation.

The DemPoRT tool itself also has implications for influencing policy and practice. As the DemPoRT algorithm is both discriminating and well-calibrated in a wide range of sub-groups of interest to policy-makers, the Public Health Agency of Canada has expressed interest in

integrating it into Statistics Canada's Population Health Model (POHEM). POHEM is a population-based longitudinal microsimulation model that dynamically models births, deaths and migration, in addition to individual disease incidence and progression, health determinants and exposure to risk factors[14]. Integration of DemPoRT into POHEM will facilitate the dynamic projection of future dementia incidence and prevalence in Canada, with many advantages over crude extrapolation and macrosimulation approaches[15], including evaluation of how changing risk factor prevalence may affect future dementia burden. It will also be used to study potential dementia prevention strategies as part of Canada's new national dementia strategy[13]. Cost and caregiving information will also be included and used to project both direct and indirect health care costs associated with dementia using POHEM.

Public health units have also expressed interest in using DemPoRT to aid in the development of regional dementia strategies – some units currently use a similar algorithm for evaluation of diabetes risk[16]. As DemPoRT was developed in Canada it is likely to be the most applicable to Canadian settings, however it can also be applied to other populations. Many other countries have similar population health surveys that collect similar information to which DemPoRT could be applied. Validation in new populations is recommended, and re-calibration should be performed if necessary.

DemPoRT is also being developed into a web-based, patient-centered health risk calculator that will enable users to assess the impact of their lifestyle choices on their risk of developing dementia. It will be added to our existing lineup of health calculators (ProjectBigLife.ca), which were developed to translate population-based research into individual estimates of health

outcomes including all-cause mortality[17] and cardiovascular disease[18]. We have found these calculators to be an effective engagement and knowledge translation tool for both the general public and knowledge users – most of the Ontario Provincial Ministers of Health have competed the life expectancy calculator, and the cardiovascular calculator is currently being used on the Canadian Heart and Stroke Foundation for cardiovascular disease prevention education. Online health calculators are a novel way to translate complex prognostic risk algorithms into engaging and meaningful risk communication tools for the public.

This thesis also has implications for general predictive algorithm dissemination and practical application. In health, predictive algorithms are most commonly disseminated by publication in a research journal, where the algorithm is described in a human-readable format. For this algorithm to be used by others the article must be read, interpreted and 'hard-coded' by analysts – variables often need to be manipulated to reproduce the final study variables, and max and min values and missing or out-of-range values need to be addressed, all of which is time-consuming and error prone. These steps are never published in a way that can be easily implemented. To facilitate dissemination and exchange of predictive algorithms in health, my team and I have been implementing standards from the information technology industry to describe and facilitate application of our predictive algorithms. We are amongst the first to use these standards in the health sector. The GitHub link provided in the manuscript text for DemPoRT includes: 1) Lime Survey questionnaire files which describe the flow of questions necessary to ascertain the model variables; 2) transformation code which describes how the responses to these questions are transformed to produce the final model variables, and 3) Predictive Modeling Mark-up Language (PMML) files describing the DemPoRT algorithms. By publishing DemPoRT with these

machine-readable files we have made it much easier for this algorithm to be used by other research teams, and in real-world settings such as within electronic medical records or decision aid applications.

## 6.3 FUTURE RESEARCH DIRECTIONS

Several follow-up studies will be conducted based on the findings of this thesis. There are plans to improve the DemPoRT model by adding other risk factors as they become available, such as sleep, traumatic brain injury and poor oral health[19,20], and by including other variables that are of interest to policy-makers. The Canadian Longitudinal Study of Aging, a national long-term study of 50,000 individuals age 45 to 85, will provide many opportunities to improve dementia prediction as well as the evaluation of health care use, and may be used to expand this work to include measures of dementia severity or subtype. I will also be evaluating if machine learning methods can be used to improve prediction of dementia for population health purposes, emphasizing an equity perspective that prioritizes vulnerable populations. External validation of DemPoRT is also a priority.

## 6.4 CONCLUSION

The burden of dementia on population health and the health care system is substantial. Taken together, this thesis provides population health planners and policy-makers with important information and tools that can be used for health resource planning and to inform dementia prevention strategies. This work supports Canada's national dementia strategy and its vision for "a Canada in which all people living with dementia and caregivers are valued and supported,

quality of life is optimized, and dementia is prevented, well understood, and effectively treated" [8].

## 6.5 REFERENCES


1. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimer's Dis*. 2016;54(1):337-349. doi:http://dx.doi.org/10.3233/JAD-160105
2. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence. *Aging Ment Health*. 2011;15(8):978-984. doi:10.1080/13607863.2011.596805
3. Kosteniuk JG, Morgan DG, O'Connell ME, et al. Incidence and prevalence of dementia in linked administrative health data in Saskatchewan, Canada: a retrospective cohort study. *BMC Geriatr*. 2015;15(1):73. doi:10.1186/s12877-015-0075-3
4. Fitzpatrick AL, Kuller LH, Lopez OL, Kawas CH, Jagust W. Survival following dementia onset: Alzheimer's disease and vascular dementia. *J Neurol Sci*. 2005;229-230:43-49. doi:10.1016/j.jns.2004.11.022
5. Aguero-Torres H, Kivipelto M, von Strauss E. Rethinking the dementia diagnoses in a population-based study: what is Alzheimer's disease and what is vascular dementia?. A study from the kungsholmen project. *Dement Geriatr Cogn Disord*. 2006;22(3):244-249. doi:10.1159/000094973
6. Schneider JA, Bennett DA. Where vascular meets neurodegenerative disease. *Stroke*. 2010;41(10 Suppl):S144-6. doi:10.1161/STROKEAHA.110.598326
7. Savva GM, Wharton SB, Ince PG, et al. Age, Neuropathology, and Dementia. *N Engl J Med*. 2009;360(22):2302-2309. doi:10.1056/NEJMoa0806142
8. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48(6):1752-1753j. doi:10.1093/ije/dyz173
9. Garriguet D, Colley RC. A comparison of self-reported leisure time physical activity and measured moderate-to-vigorous physical activity in adolescents and adults. *Heal Reports*. 2014;25(7):3-11.
10. Tanuseputro P, Perez R, Rosella L, et al. Improving the estimation of the burden of risk factors: an illustrative comparison of methods to measure smoking-attributable mortality. *Popul Health Metr*. 2015;13:5. doi:10.1186/s12963-015-0039-z
11. Gelman A. Multilevel (hierarchical) modeling: What It can and cannot do. *Technometrics*.

2006;48(3):432-435. doi:10.1198/004017005000000661

12. Shmueli G. To explain or to predict? *Stat Sci*. 2010;25(3):289-310. doi:10.1214/10-STS330
13. Public Health Agency of Canada. *A Dementia Strategy for Canada: Together We Aspire*.
14. Hennessy DA, Flanagan WM, Tanuseputro P, et al. The Population Health Model (POHEM): an overview of rationale, methods and applications. *Popul Health Metr*. 2015;13(24):1-12. doi:10.1186/s12963-015-0057-x
15. Norton S, Matthews FE, Brayne C. A commentary on studies presenting projections of the future prevalence of dementia. *BMC Public Health*. 2013;13(1):1. doi:10.1186/1471-2458-13-1
16. 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 Heal*. 2011;65(7):613-620. doi:10.1136/jech.2009.102244
17. Manuel DG, Perez R, Sanmartin C, 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):1-27. doi:10.1371/journal.pmed.1002082
18. Manuel DG, Tuna M, Bennett C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: The Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ*. 2018;190(29):E871-E882. doi:10.1503/cmaj.170914
19. Baumgart M, Snyder HM, Carrillo MC, Fazio S, Kim H, Johns H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's Dement*. 2015;11(6):718-726. doi:10.1016/j.jalz.2015.05.016
20. Kamer AR, Janal MN, de Leon M. Letter to the editor regarding: Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's Dement (Amsterdam, Netherlands)*. 2015;1(4):385-386. doi:10.1016/j.dadm.2015.08.003

# CHAPTER 7: APPENDICIES

## 7.1 CHAPTER 3 APPENDICIES

**Appendix 3-1.** TRIPOD checklist for prediction model development and validation

| Section/Topic | Item | | Checklist Item | Page |
|---|---|---|---|---|
| Title and abstract | | | | |
| Title | 1 | D;V | Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted. | 41 |
| Abstract | 2 | D;V | Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | 42, 43 |
| Introduction | | | | |
| Background and objectives | 3a | D;V | Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models. | 44 |
| | 3b | D;V | Specify the objectives, including whether the study describes the development or validation of the model or both. | 44, 45 |
| Methods | | | | |
| Source of data | 4a | D;V | Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. | 45-47 |
| | 4b | D;V | Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up. | 45, 47 |
| Participants | 5a | D;V | Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres. | 45, 46 |
| | 5b | D;V | Describe eligibility criteria for participants. | 45, 46 |
| | 5c | D;V | Give details of treatments received, if relevant. | NA |
| Outcome | 6a | D;V | Clearly define the outcome that is predicted by the prediction model, including how and when assessed. | 47 |
| | 6b | D;V | Report any actions to blind assessment of the outcome to be predicted. | 48 |
| Predictors | 7a | D;V | Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured. | 48-50 |
| | 7b | D;V | Report any actions to blind assessment of predictors for the outcome and other predictors. | 48 |
| Sample size | 8 | D;V | Explain how the study size was arrived at. | 45, protocol |
| Missing data | 9 | D;V | Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method. | 48, protocol |
| Statistical analysis methods | 10a | D | Describe how predictors were handled in the analyses. | 48-50 |
| | 10b | D | Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation. | 48, protocol |
| | 10c | V | For validation, describe how the predictions were calculated. | 48, protocol |
| | 10d | D;V | Specify all measures used to assess model performance and, if relevant, to compare multiple models. | 51 |
| | 10e | V | Describe any model updating (e.g., recalibration) arising from the validation, if done. | NA |
| Risk groups | 11 | D;V | Provide details on how risk groups were created, if done. | 51 |
| Development vs. validation | 12 | V | For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. | 45, 53-55 |

| | | | | |
|---|---|---|---|---|
| Results | | | | |
| Participants | 13a | D;V | Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful. | 52 |
| | 13b | D;V | Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. | 53-55 |
| | 13c | V | For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome). | 53-55 |
| Model development | 14a | D | Specify the number of participants and outcome events in each analysis. | 52-55 |
| | 14b | D | If done, report the unadjusted association between each candidate predictor and outcome. | NA |
| Model specification | 15a | D | Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point). | Appendix 3-5 and online files |
| | 15b | D | Explain how to the use the prediction model. | Appendix 3-5 and online files |
| Model performance | 16 | D;V | Report performance measures (with CIs) for the prediction model. | 57-58 |
| Model-updating | 17 | V | If done, report the results from any model updating (i.e., model specification, model performance). | NA |
| Discussion | | | | |
| Limitations | 18 | D;V | Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data). | 65 |
| Interpretation | 19a | V | For validation, discuss the results with reference to performance in the development data, and any other validation data. | 59, 61 |
| | 19b | D;V | Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence. | 59, 61 |
| Implications | 20 | D;V | Discuss the potential clinical use of the model and implications for future research. | 61, 62 |
| Other information | | | | |
| Supplementary information | 21 | D;V | Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets. | Throughout manuscript |
| Funding | 22 | D;V | Give the source of funding and the role of the funders for the present study. | 66 |

* Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

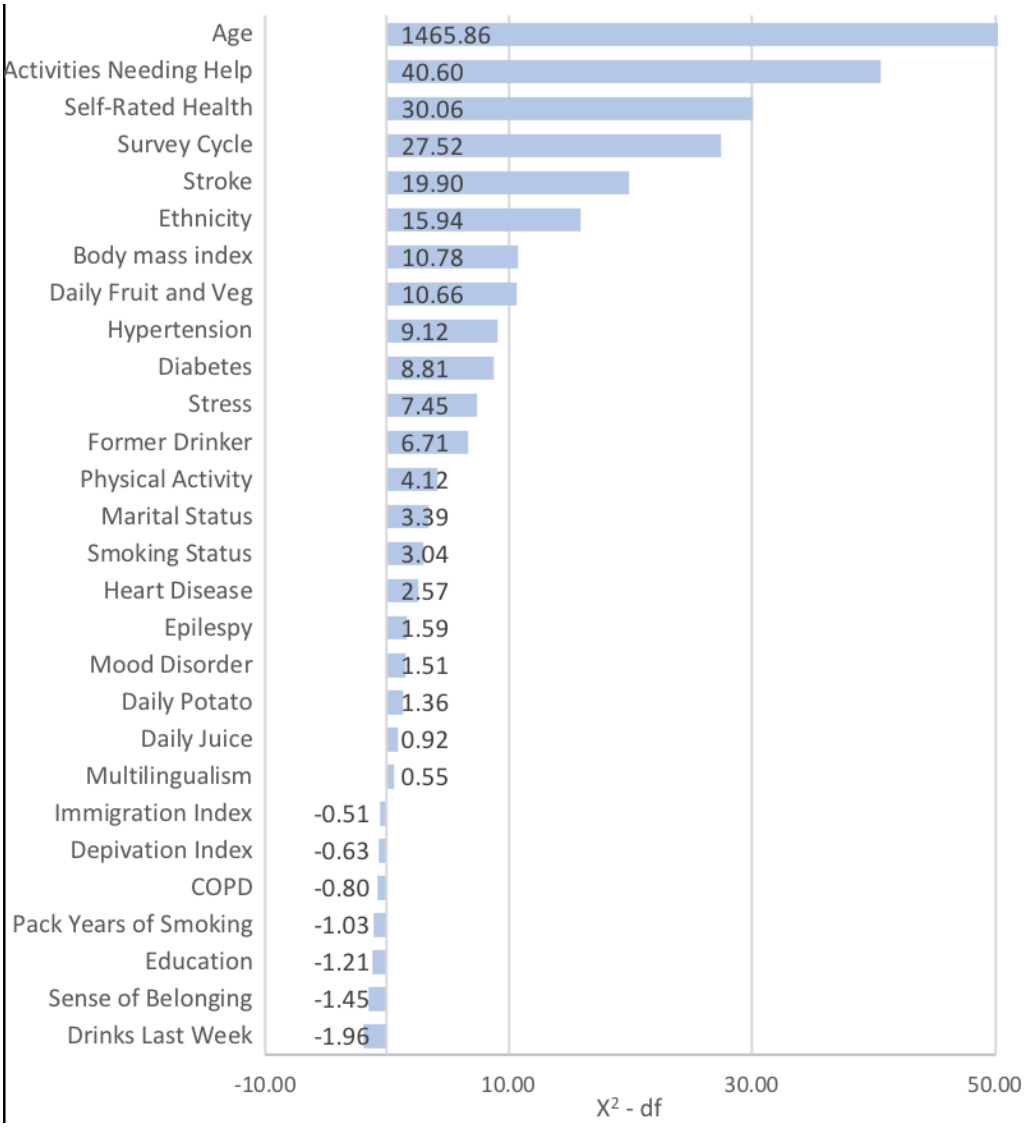
Source: http://www.equator-network.org/reporting-guidelines/tripod-statement/

Reference: Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Eur Urol,* 2015; 67:1142-1151.

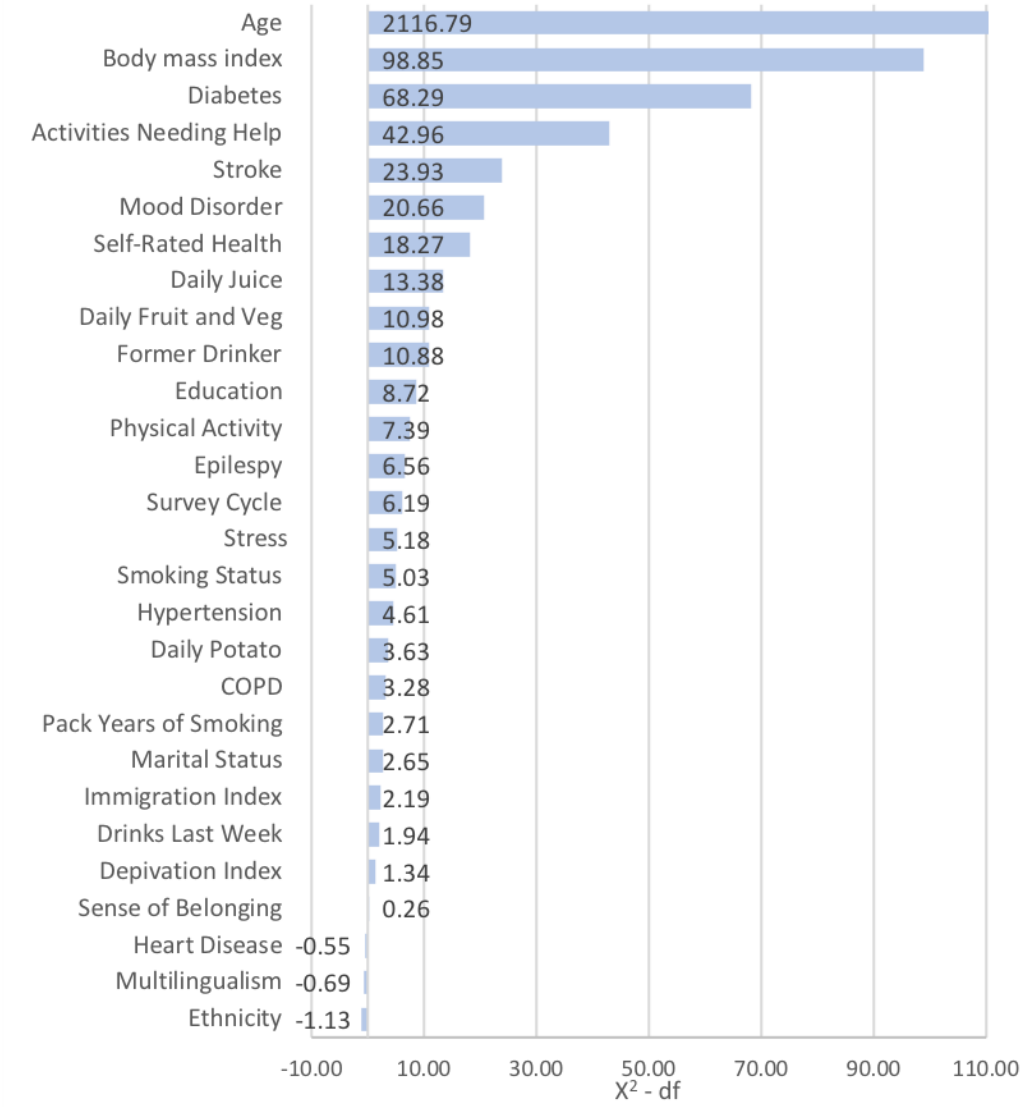
**Appendix 3-2.** Male and female partial correlation plots

- Partial correlation plots show each variable's contribution to the dementia risk calculation in the presence of the other variables
- The Wald chi square statistics, used as a measure of association for each predictor and time to dementia incidence, were penalized for degrees of freedom to level the playing field among predictors with varying degrees of freedom
- See Harrell FE. *Regression Modelling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001

Males:



Females:


Age 2116.79
Body mass index 98.85
Diabetes 68.29
Activities Needing Help 42.96
Stroke 23.93
Mood Disorder 20.66
Self-Rated Health 18.27
Daily Juice 13.38
Daily Fruit and Veg 10.98
Former Drinker 10.88
Education 8.72
Physical Activity 7.39
Epilespy 6.56
Survey Cycle 6.19
Stress 5.18
Smoking Status 5.03
Hypertension 4.61
Daily Potato 3.63
COPD 3.28
Pack Years of Smoking 2.71
Marital Status 2.65
Immigration Index 2.19
Drinks Last Week 1.94
Depivation Index 1.34
Sense of Belonging 0.26
Heart Disease -0.55
Multilingualism -0.69
Ethnicity -1.13
-10.00
10.00
30.00
50.00
70.00
90.00
110.00
X² - df

**Appendix 3-3.** DemPoRT subdistribution hazard ratios for full and reduced models – males[1]

| Variable | Description | sHR (95% CI) Full Model | Reduced Model |
|---|---|---|---|
| AgeC_rcs1 | First RCS component for age | 1.21 ( 1.11, 1.31 ) | 1.21 ( 1.11, 1.31 ) |
| AgeC_rcs2 | Second RCS component for age | 0.81 ( 0.41, 1.59 ) | 0.80 ( 0.41, 1.58 ) |
| AgeC_rcs3 | Third RCS component for age | 1.63 ( 0.31, 8.56 ) | 1.66 ( 0.32, 8.75 ) |
| AgeC_rcs4 | Fourth RCS component for age | 0.54 ( 0.15, 1.98 ) | 0.53 ( 0.15, 1.94 ) |
| EthnSEAsian4C_cat | Ethnicity: SE Asian/Chinese/Japanese/Korean/Filipino | 0.55 ( 0.37, 0.81 ) | 0.54 ( 0.37, 0.80 ) |
| EthnOther4C_cat | Ethnicity: Other/Multiple origin/ Unknown | 1.02 ( 0.81, 1.27 ) | 1.02 ( 0.82, 1.27 ) |
| EthnSWAsian4C_cat | Ethnicity: South Asian/Arab/West Asian | 0.67 ( 0.46, 0.98 ) | 0.66 ( 0.46, 0.96 ) |
| ImmigrantC_cat | Immigrant | 0.95 ( 0.87, 1.04 ) | - |
| EduHSGrad2C_cat | Education: High school graduate | 0.99 ( 0.88, 1.12 ) | - |
| EduSomePS2C_cat | Education: Some post-secondary school | 0.94 ( 0.79, 1.11 ) | - |
| EduPSGrad2C_cat | Education: Post-secondary graduate | 0.95 ( 0.88, 1.04 ) | - |
| MSDivorcedC_cat | Marital status: Separated/Divorced | 1.08 ( 0.94, 1.24 ) | - |
| MSWidowedC_cat | Marital status: Widowed | 0.96 ( 0.87, 1.07 ) | - |
| MSSingleC_cat | Marital status: Single | 0.92 ( 0.78, 1.08 ) | - |
| MultilingualC_cat | Multilingual | 1.08 ( 1.00, 1.18 ) | 1.07 ( 0.99, 1.16 ) |
| DepIndModC_cat | Deprivation: Moderate | 1.02 ( 0.93, 1.12 ) | - |
| DepIndHighC_cat | Deprivation: High | 1.06 ( 0.93, 1.20 ) | - |
| BelongSomeC_cat | Sense of Belonging: Somewhat strong | 1.00 ( 0.92, 1.09 ) | - |
| BelongSWeakC_cat | Sense of Belonging: Somewhat weak | 1.08 ( 0.96, 1.20 ) | - |
| BelongVWeakC_cat | Sense of Belonging: Very weak | 1.06 ( 0.92, 1.23 ) | - |
| StressNotVeryC_cat | Stress: Not at all stressful | 0.99 ( 0.90, 1.09 ) | 0.99 ( 0.90, 1.09 ) |
| StressABitC_cat | Stress: A bit stressful | 1.13 ( 1.02, 1.25 ) | 1.13 ( 1.02, 1.25 ) |
| StressQuiteC_cat | Stress: Quite a bit stressful | 1.25 ( 1.08, 1.44 ) | 1.25 ( 1.09, 1.45 ) |
| StressExtreamlyC_cat | Stress: Extremely stressful | 1.26 ( 0.96, 1.65 ) | 1.27 ( 0.98, 1.66 ) |
| SRHealthPoorC_cat | Self-rated health: Poor | 1.03 ( 0.90, 1.17 ) | 1.03 ( 0.91, 1.18 ) |
| SRHealthFairC_cat | Self-rated health: Fair | 1.18 ( 1.04, 1.34 ) | 1.19 ( 1.05, 1.35 ) |
| SRHealthGoodC_cat | Self-rated health: Good | 1.15 ( 0.99, 1.33 ) | 1.17 ( 1.01, 1.35 ) |
| SRHealthVGoodC_cat | Self-rated health: Very good | 1.09 ( 0.90, 1.32 ) | 1.11 ( 0.91, 1.35 ) |
| PackYearsC_rcs1 | First RCS component of pack years of smoking | 0.99 ( 0.99, 1.00 ) | 0.99 ( 0.99, 1.00 ) |
| PackYearsC_rcs2 | Second RCS component of pack years of smoking | 1.02 ( 1.00, 1.03 ) | 1.02 ( 1.00, 1.03 ) |
| SmokeFormer0to5C_cat | Smoking status: Former smoker quit <5 years ago | 1.04 ( 0.89, 1.20 ) | 1.04 ( 0.90, 1.21 ) |
| SmokeFormer5PlusC_cat | Smoking status: Former smoker quit ≥5 years ago | 0.95 ( 0.76, 1.17 ) | 0.95 ( 0.77, 1.18 ) |
| SmokeCurrentC_cat | Smoking status: Current smoker | 1.08 ( 0.88, 1.33 ) | 1.10 ( 0.90, 1.35 ) |
| DrinksLastWeekC_rcs1 | First RCS component of drinks last week | 0.99 ( 0.96, 1.01 ) | 0.98 ( 0.96, 1.01 ) |
| DrinksLastWeekC_rcs2 | Second RCS component of drinks last week | 1.05 ( 0.90, 1.22 ) | 1.06 ( 0.91, 1.23 ) |
| FormerDrinkerC_cat | Former drinker | 1.13 ( 0.99, 1.29 ) | 1.14 ( 0.99, 1.30 ) |
| FruitVegC_rcs1 | First RCS component of daily fruit and vegetable consumption | 0.94 ( 0.89, 1.00 ) | 0.94 ( 0.89, 0.99 ) |
| FruitVegC_rcs2 | Second RCS component of daily fruit and vegetable consumption | 1.07 ( 1.01, 1.14 ) | 1.08 ( 1.01, 1.15 ) |
| PotatoC_rcs1 | First RCS component of daily potato consumption | 0.97 ( 0.70, 1.34 ) | 0.97 ( 0.70, 1.34 ) |
| PotatoC_rcs2 | Second RCS component of daily potato consumption | 1.32 ( 0.84, 2.06 ) | 1.32 ( 0.85, 2.07 ) |
| JuiceC_rcs1 | First RCS component of daily juice consumption | 1.02 ( 0.89, 1.17 ) | 1.02 ( 0.89, 1.17 ) |
| JuiceC_rcs2 | Second RCS component of daily juice consumption | 1.06 ( 0.94, 1.19 ) | 1.06 ( 0.94, 1.18 ) |
| PhysicalActivityC_rcs1 | First RCS component of daily physical activity | 0.95 ( 0.89, 1.01 ) | 0.95 ( 0.89, 1.01 ) |
| PhysicalActivityC_rcs2 | Second RCS component of daily physical activity | 1.04 ( 0.96, 1.12 ) | 1.04 ( 0.97, 1.12 ) |

| | | | |
|---|---|---|---|
| HeartDisC_cat | Heart disease | 0.87 ( 0.76, 1.00 ) | 0.87 ( 0.76, 1.00 ) |
| StrokeC_cat | Stroke | 1.51 ( 1.20, 1.91 ) | 1.55 ( 1.23, 1.95 ) |
| DiabetesC_cat | Diabetes | 1.16 ( 1.01, 1.32 ) | 1.16 ( 1.01, 1.32 ) |
| MoodC_cat | Mood disorder | 1.19 ( 0.99, 1.44 ) | 1.21 ( 1.00, 1.45 ) |
| HypertensionC_cat | Hypertension | 0.92 ( 0.83, 1.03 ) | 0.92 ( 0.83, 1.02 ) |
| COPDC_cat | COPD | 0.91 ( 0.75, 1.10 ) | 0.91 ( 0.75, 1.11 ) |
| EpilepsyC_cat | Epilepsy | 1.55 ( 0.92, 2.60 ) | - |
| BMIC_rcs1 | First RCS component of body mass index | 0.98 ( 0.96, 1.00 ) | 0.98 ( 0.96, 1.00 ) |
| BMIC_rcs2 | Second RCS component of body mass index | 1.01 ( 0.98, 1.04 ) | 1.01 ( 0.98, 1.04 ) |
| NeedHelp1C_cat | Need help with one functional activity | 1.56 ( 1.29, 1.90 ) | 1.57 ( 1.29, 1.90 ) |
| NeedHelp2C_cat | Need help with two functional activities | 1.22 ( 0.86, 1.73 ) | 1.22 ( 0.87, 1.73 ) |
| NeedHelp3C_cat | Need help with three functional activities | 1.39 ( 0.99, 1.96 ) | 1.40 ( 0.99, 1.97 ) |
| NeedHelp4C_cat | Need help with four functional activities | 1.61 ( 1.08, 2.41 ) | 1.61 ( 1.08, 2.40 ) |
| NeedHelp5C_cat | Need help with five functional activities | 2.25 ( 1.55, 3.27 ) | 2.26 ( 1.56, 3.28 ) |
| NeedHelp6C_cat | Need help with six functional activities | 1.25 ( 0.72, 2.17 ) | 1.27 ( 0.73, 2.20 ) |
| SurveyCycle2C_cat | Survey year: 2003 | 0.85 ( 0.76, 0.94 ) | 0.84 ( 0.76, 0.94 ) |
| SurveyCycle3C_cat | Survey year: 2005 | 0.88 ( 0.79, 0.99 ) | 0.88 ( 0.79, 0.98 ) |
| SurveyCycle4C_cat | Survey year: 2007/08 | 0.62 ( 0.55, 0.70 ) | 0.62 ( 0.55, 0.70 ) |
| SurveyCycle5C_cat | Survey year: 2009/10 | 0.66 ( 0.58, 0.75 ) | 0.65 ( 0.57, 0.74 ) |
| SurveyCycle6C_cat | Survey year: 2011/12 | 0.59 ( 0.51, 0.68 ) | 0.59 ( 0.51, 0.68 ) |
| AgeCXPackYearsC_int | Interaction between age and pack years of smoking | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXSmokeFormer5PlusC_int | Interaction between age and former smoker quit ≥5 years | 0.99 ( 0.98, 1.01 ) | 0.99 ( 0.98, 1.01 ) |
| AgeCXSmokeFormer0to5C_int | Interaction between age and former smoker quit <5 years | 0.99 ( 0.97, 1.02 ) | 0.99 ( 0.97, 1.02 ) |
| AgeCXSmokeCurrentC_int | Interaction between age and current smoker | 0.99 ( 0.97, 1.01 ) | 0.99 ( 0.97, 1.01 ) |
| AgeCXDrinksLastWeekC_int | Interaction between age and drinks last week | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXFormerDrinkerC_int | Interaction between age and former drinker | 1.00 ( 0.99, 1.01 ) | 1.00 ( 0.99, 1.01 ) |
| AgeCXFruitVegC_int | Interaction between age and daily fruit and vegetable consumption | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXJuiceC_int | Interaction between age and daily juice consumption | 0.99 ( 0.98, 1.00 ) | 0.99 ( 0.98, 1.00 ) |
| AgeCXPotatoC_int | Interaction between age and daily potato consumption | 0.99 ( 0.97, 1.00 ) | 0.99 ( 0.97, 1.00 ) |
| AgeCXPhysicalActivityC_int | Interaction between age and physical activity | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXHeartDisC_int | Interaction between age and health disease | 1.00 ( 0.99, 1.01 ) | 1.00 ( 0.99, 1.01 ) |
| AgeCXStrokeC_int | Interaction between age and stroke | 1.00 ( 0.97, 1.02 ) | 0.99 ( 0.97, 1.02 ) |
| AgeCXDiabetesC_int | Interaction between age and diabetes | 0.99 ( 0.98, 1.00 ) | 0.99 ( 0.98, 1.00 ) |
| AgeCXMoodC_int | Interaction between age and mood disorder | 0.99 ( 0.97, 1.01 ) | 0.99 ( 0.97, 1.01 ) |
| AgeCXHypertensionC_int | Interaction between age and hypertension | 0.99 ( 0.98, 1.00 ) | 0.99 ( 0.98, 1.00 ) |
| AgeCXCOPDC_int | Interaction between age and COPD | 0.99 ( 0.97, 1.01 ) | 0.99 ( 0.97, 1.01 ) |
| AgeCXEpilepsyC_int | Interaction between age and epilepsy | 1.00 ( 0.93, 1.07 ) | - |
| AgeCXBMIC_int | Interaction between age and body mass index | 1.00 ( 1.00, 1.00 ) | 0.98 ( 0.96, 1.00 ) |
| AgeCXNeedHelp1C_int | Interaction between age and need help with one activity | 0.98 ( 0.96, 1.00 ) | 0.99 ( 0.97, 1.02 ) |
| AgeCXNeedHelp2C_int | Interaction between age and need help with two activities | 0.99 ( 0.97, 1.02 ) | 0.99 ( 0.96, 1.01 ) |
| AgeCXNeedHelp3C_int | Interaction between age and need help with three activities | 0.99 ( 0.96, 1.01 ) | 0.97 ( 0.93, 1.00 ) |
| AgeCXNeedHelp4C_int | Interaction between age and need help with four activities | 0.96 ( 0.93, 1.00 ) | 0.96 ( 0.94, 0.99 ) |
| AgeCXNeedHelp5C_int | Interaction between age and need help with five activities | 0.96 ( 0.94, 0.99 ) | 0.97 ( 0.93, 1.01 ) |
| AgeCXNeedHelp6C_int | Interaction between age and need help with six activities | 0.97 ( 0.93, 1.02 ) | 0.98 ( 0.96, 1.00 ) |

[1] All variables are centered on their means

Abbreviations: CI: confidence interval; COPD= chronic obstructive pulmonary disease; RCS= restricted cubic spline; sHR= subdistribution hazard ratio

**Appendix 3-4.** DemPoRT subdistribution hazard ratios for full and reduced models – females[1]

| Variable | Description | sHR (95% CI) Full Model | Reduced Model |
|---|---|---|---|
| AgeC_rcs1 | First RCS component for age | 1.28 ( 1.20, 1.36 ) | 1.28 ( 1.20, 1.36 ) |
| AgeC_rcs2 | Second RCS component for age | 0.55 ( 0.34, 0.87 ) | 0.55 ( 0.34, 0.87 ) |
| AgeC_rcs3 | Third RCS component for age | 3.04 ( 1.17, 7.89 ) | 3.02 ( 1.17, 7.83 ) |
| AgeC_rcs4 | Fourth RCS component for age | 0.43 ( 0.21, 0.88 ) | 0.43 ( 0.21, 0.88 ) |
| EthnSEAsian4C_cat | Ethnicity: SE Asian/Chinese/Japanese/Korean/Filipino | 0.92 ( 0.71, 1.19 ) | 1.28 ( 1.20, 1.36 ) |
| EthnOther4C_cat | Ethnicity: Other/Multiple origin/ Unknown | 1.02 ( 0.86, 1.22 ) | - |
| EthnSWAsian4C_cat | Ethnicity: South Asian/Arab/West Asian | 1.13 ( 0.84, 1.53 ) | - |
| ImmigrantC_cat | Immigrant | 0.96 ( 0.90, 1.03 ) | - |
| EduHSGrad2C_cat | Education: High school graduate | 0.94 ( 0.87, 1.02 ) | 0.94 ( 0.87, 1.01 ) |
| EduSomePS2C_cat | Education: Some post-secondary school | 0.89 ( 0.78, 1.01 ) | 0.88 ( 0.77, 1.00 ) |
| EduPSGrad2C_cat | Education: Post-secondary graduate | 0.90 ( 0.84, 0.97 ) | 0.90 ( 0.84, 0.96 ) |
| MSDivorcedC_cat | Marital status: Separated/Divorced | 1.11 ( 0.99, 1.23 ) | 1.11 ( 1.00, 1.24 ) |
| MSWidowedC_cat | Marital status: Widowed | 1.05 ( 0.98, 1.13 ) | 1.06 ( 0.99, 1.13 ) |
| MSSingleC_cat | Marital status: Single | 1.08 ( 0.94, 1.25 ) | 1.09 ( 0.95, 1.26 ) |
| MultilingualC_cat | Multilingual | 1.03 ( 0.96, 1.10 ) | - |
| DepIndModC_cat | Deprivation: Moderate | 1.01 ( 0.93, 1.09 ) | - |
| DepIndHighC_cat | Deprivation: High | 1.02 ( 0.93, 1.13 ) | - |
| BelongSomeC_cat | Sense of Belonging: Somewhat strong | 0.98 ( 0.92, 1.05 ) | - |
| BelongSWeakC_cat | Sense of Belonging: Somewhat weak | 1.05 ( 0.97, 1.14 ) | - |
| BelongVWeakC_cat | Sense of Belonging: Very weak | 1.04 ( 0.93, 1.16 ) | - |
| StressNotVeryC_cat | Stress: Not at all stressful | 0.93 ( 0.86, 1.00 ) | 0.93 ( 0.86, 1.00 ) |
| StressABitC_cat | Stress: A bit stressful | 1.01 ( 0.94, 1.09 ) | 1.02 ( 0.95, 1.10 ) |
| StressQuiteC_cat | Stress: Quite a bit stressful | 1.05 ( 0.95, 1.17 ) | 1.08 ( 0.97, 1.19 ) |
| StressExtreamlyC_cat | Stress: Extremely stressful | 1.10 ( 0.91, 1.33 ) | 1.13 ( 0.94, 1.36 ) |
| SRHealthPoorC_cat | Self-rated health: Poor | 1.00 ( 0.90, 1.10 ) | - |
| SRHealthFairC_cat | Self-rated health: Fair | 1.00 ( 0.91, 1.10 ) | - |
| SRHealthGoodC_cat | Self-rated health: Good | 1.08 ( 0.96, 1.20 ) | - |
| SRHealthVGoodC_cat | Self-rated health: Very good | 1.07 ( 0.92, 1.24 ) | - |
| PackYearsC_rcs1 | First RCS component of pack years of smoking | 1.01 ( 0.99, 1.02 ) | 1.01 ( 1.00, 1.02 ) |
| PackYearsC_rcs2 | Second RCS component of pack years of smoking | 0.95 ( 0.88, 1.02 ) | 0.95 ( 0.88, 1.02 ) |
| SmokeFormer5PlusC_cat | Smoking status: Former smoker quit ≥5 years ago | 1.03 ( 0.92, 1.15 ) | 1.03 ( 0.92, 1.15 ) |
| SmokeFormer0to5C_cat | Smoking status: Former smoker quit <5 years ago | 1.04 ( 0.88, 1.22 ) | 1.04 ( 0.88, 1.23 ) |
| SmokeCurrentC_cat | Smoking status: Current smoker | 1.13 ( 0.94, 1.35 ) | 1.13 ( 0.95, 1.36 ) |
| DrinksLastWeekC_rcs1 | First RCS component of drinks last week | 0.97 ( 0.94, 1.01 ) | 0.97 ( 0.93, 1.01 ) |
| DrinksLastWeekC_rcs2 | Second RCS component of drinks last week | 1.03 ( 0.97, 1.09 ) | 1.03 ( 0.97, 1.09 ) |
| FormerDrinkerC_cat | Former drinker | 1.14 ( 1.04, 1.25 ) | 1.14 ( 1.04, 1.25 ) |
| FruitVegC_rcs1 | First RCS component of daily fruit and vegetable consumption | 0.94 ( 0.90, 0.97 ) | 0.94 ( 0.90, 0.97 ) |
| FruitVegC_rcs2 | Second RCS component of daily fruit and vegetable consumption | 1.04 ( 1.00, 1.08 ) | 1.04 ( 1.00, 1.09 ) |
| PotatoC_rcs1 | First RCS component of daily potato consumption | 1.11 ( 0.87, 1.40 ) | 1.10 ( 0.87, 1.39 ) |
| PotatoC_rcs2 | Second RCS component of daily potato consumption | 1.12 ( 0.81, 1.57 ) | 1.13 ( 0.81, 1.58 ) |
| JuiceC_rcs1 | First RCS component of daily juice consumption | 1.19 ( 1.08, 1.30 ) | 1.18 ( 1.08, 1.30 ) |
| JuiceC_rcs2 | Second RCS component of daily juice consumption | 0.96 ( 0.88, 1.05 ) | 0.96 ( 0.88, 1.05 ) |
| PhysicalActivityC_rcs1 | First RCS component of daily physical activity | 0.96 ( 0.91, 1.02 ) | 0.95 ( 0.90, 1.01 ) |
| PhysicalActivityC_rcs2 | Second RCS component of daily physical activity | 1.06 ( 0.98, 1.14 ) | 1.07 ( 0.99, 1.15 ) |

| | | | |
|---|---|---|---|
| HeartDisC_cat | Heart disease | 1.06 ( 0.95, 1.17 ) | 1.07 ( 0.96, 1.19 ) |
| StrokeC_cat | Stroke | 1.43 ( 1.19, 1.73 ) | 1.45 ( 1.20, 1.74 ) |
| DiabetesC_cat | Diabetes | 1.49 ( 1.34, 1.67 ) | 1.51 ( 1.35, 1.68 ) |
| MoodC_cat | Mood disorder | 1.34 ( 1.19, 1.51 ) | 1.36 ( 1.20, 1.53 ) |
| HypertensionC_cat | Hypertension | 0.99 ( 0.91, 1.08 ) | 1.00 ( 0.92, 1.08 ) |
| COPDC_cat | COPD | 0.82 ( 0.72, 0.95 ) | 0.83 ( 0.73, 0.96 ) |
| EpilepsyC_cat | Epilepsy | 1.43 ( 1.01, 2.03 ) | 1.43 ( 1.01, 2.03 ) |
| BMIC_rcs1 | First RCS component of body mass index | 0.96 ( 0.95, 0.97 ) | 0.96 ( 0.94, 0.97 ) |
| BMIC_rcs2 | Second RCS component of body mass index | 1.02 ( 1.00, 1.04 ) | 1.02 ( 1.00, 1.04 ) |
| NeedHelp1C_cat | Need help with one functional activity | 1.27 ( 1.13, 1.44 ) | - |
| NeedHelp2C_cat | Need help with two functional activities | 1.45 ( 1.23, 1.70 ) | - |
| NeedHelp3C_cat | Need help with three functional activities | 1.56 ( 1.28, 1.89 ) | - |
| NeedHelp4C_cat | Need help with four functional activities | 1.54 ( 1.19, 1.98 ) | - |
| NeedHelp5C_cat | Need help with five functional activities | 1.35 ( 0.97, 1.89 ) | - |
| NeedHelp6C_cat | Need help with six functional activities | 1.82 ( 1.32, 2.51 ) | - |
| SurveyCycle2C_cat | Survey year: 2003 | 1.02 ( 0.94, 1.10 ) | 1.02 ( 0.94, 1.10 ) |
| SurveyCycle3C_cat | Survey year: 2005 | 0.87 ( 0.80, 0.95 ) | 0.87 ( 0.80, 0.94 ) |
| SurveyCycle4C_cat | Survey year: 2007/08 | 0.80 ( 0.73, 0.87 ) | 0.79 ( 0.72, 0.87 ) |
| SurveyCycle5C_cat | Survey year: 2009/10 | 0.70 ( 0.63, 0.77 ) | 0.69 ( 0.63, 0.77 ) |
| SurveyCycle6C_cat | Survey year: 2011/12 | 0.63 ( 0.56, 0.71 ) | 0.63 ( 0.56, 0.70 ) |
| AgeCXPackYearsC_int | Interaction between age and pack years of smoking | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXSmokeFormer5PlusC_int | Interaction between age and former smoker quit ≥5 years | 1.00 ( 0.99, 1.01 ) | 1.00 ( 0.99, 1.01 ) |
| AgeCXSmokeFormer0to5C_int | Interaction between age and former smoker quit <5 years | 0.99 ( 0.98, 1.01 ) | 0.99 ( 0.98, 1.01 ) |
| AgeCXSmokeCurrentC_int | Interaction between age and current smoker | 0.99 ( 0.97, 1.00 ) | 0.99 ( 0.97, 1.00 ) |
| AgeCXDrinksLastWeekC_int | Interaction between age and drinks last week | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXFormerDrinkerC_int | Interaction between age and former drinker | 0.99 ( 0.99, 1.00 ) | 0.99 ( 0.99, 1.00 ) |
| AgeCXFruitVegC_int | Interaction between age and daily fruit and vegetable consumption | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXJuiceC_int | Interaction between age and daily juice consumption | 0.99 ( 0.99, 1.00 ) | 0.99 ( 0.99, 1.00 ) |
| AgeCXPotatoC_int | Interaction between age and daily potato consumption | 0.99 ( 0.98, 1.00 ) | 0.99 ( 0.98, 1.00 ) |
| AgeCXPhysicalActivityC_int | Interaction between age and physical activity | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXHeartDisC_int | Interaction between age and health disease | 0.98 ( 0.97, 0.99 ) | 0.98 ( 0.97, 0.99 ) |
| AgeCXStrokeC_int | Interaction between age and stroke | 1.00 ( 0.98, 1.02 ) | 1.00 ( 0.98, 1.02 ) |
| AgeCXDiabetesC_int | Interaction between age and diabetes | 0.98 ( 0.96, 0.99 ) | 0.98 ( 0.96, 0.99 ) |
| AgeCXMoodC_int | Interaction between age and mood disorder | 1.00 ( 0.98, 1.01 ) | 1.00 ( 0.98, 1.01 ) |
| AgeCXHypertensionC_int | Interaction between age and hypertension | 0.99 ( 0.98, 0.99 ) | 0.99 ( 0.98, 0.99 ) |
| AgeCXCOPDC_int | Interaction between age and COPD | 0.99 ( 0.98, 1.01 ) | 0.99 ( 0.98, 1.01 ) |
| AgeCXEpilepsyC_int | Interaction between age and epilepsy | 0.97 ( 0.93, 1.00 ) | 0.97 ( 0.93, 1.00 ) |
| AgeCXBMIC_int | Interaction between age and body mass index | 1.00 ( 1.00, 1.00 ) | 1.00 ( 1.00, 1.00 ) |
| AgeCXNeedHelp1C_int | Interaction between age and need help with one activity | 0.98 ( 0.97, 0.99 ) | 0.98 ( 0.97, 0.99 ) |
| AgeCXNeedHelp2C_int | Interaction between age and need help with two activities | 0.98 ( 0.97, 1.00 ) | 0.98 ( 0.97, 1.00 ) |
| AgeCXNeedHelp3C_int | Interaction between age and need help with three activities | 0.96 ( 0.95, 0.98 ) | 0.96 ( 0.95, 0.98 ) |
| AgeCXNeedHelp4C_int | Interaction between age and need help with four activities | 0.98 ( 0.96, 0.99 ) | 0.97 ( 0.96, 0.99 ) |
| AgeCXNeedHelp5C_int | Interaction between age and need help with five activities | 0.98 ( 0.95, 1.00 ) | 0.98 ( 0.95, 1.00 ) |
| AgeCXNeedHelp6C_int | Interaction between age and need help with six activities | 0.94 ( 0.92, 0.96 ) | 0.94 ( 0.92, 0.96 ) |

[1] All variables are centered on their means

Abbreviations: CI: confidence interval; COPD= chronic obstructive pulmonary disease; RCS= restricted cubic spline; sHR= subdistribution hazard ratio

**Appendix 3-5.** DemPoRT reduced male and female formulas for risk

**DemPoRT formula for 5-year risk of dementia:**

$$5\ year\ dementia\ risk = 1 - \exp(-(h_0(5\ years) * \exp(Score)))$$

**5 year baseline hazards:**

$h_0(5\ years)_{male}$ = 0.03026372
$h_0(5\ years)_{female}$ = 0.08718073

**Beta coefficients:**
Available at https://github.com/Big-Life-Lab/predictive-algorithms

**Male DemPoRT risk score:**
$Score_{male}$ =

($\beta_{AgeC\_rcs1}$ * AgeC_rcs1) + ($\beta_{AgeC\_rcs2}$ * AgeC_rcs2) + ($\beta_{AgeC\_rcs3}$*AgeC_rcs3) + ($\beta_{AgeC\_rcs4}$ * AgeC_rcs4) + ($\beta_{EthnSEAsian4C\_cat}$ * EthnSEAsian4C_cat) + ($\beta_{EthnOther4C\_cat}$ * EthnOther4C_cat) + ($\beta_{EthnSWAsian4C\_cat}$ * EthnSWAsian4C_cat) + ($\beta_{MultilingualC\_cat}$ * MultilingualC_cat) + ($\beta_{StressNotVeryC\_cat}$ * StressNotVeryC_cat) + ($\beta_{StressABitC\_cat}$ * StressABitC_cat) + ($\beta_{StressQuiteC\_cat}$ * StressQuiteC_cat) + ($\beta_{StressExtreamlyC\_cat}$ * StressExtreamlyC_cat) + ($\beta_{SRHealthPoorC\_cat}$ * SRHealthPoorC_cat) + ($\beta_{SRHealthFairC\_cat}$ * SRHealthFairC_cat) + ($\beta_{SRHealthGoodC\_cat}$ * SRHealthGoodC_cat) + ($\beta_{SRHealthVGoodC\_cat}$ * SRHealthVGoodC_cat) + ($\beta_{PackYearsC\_rcs1}$ * PackYearsC_rcs1) + ($\beta_{PackYearsC\_rcs2}$ * PackYearsC_rcs2) + ($\beta_{SmokeFormer5PlusC\_cat}$ * SmokeFormer5PlusC_cat) + ($\beta_{SmokeFormer0to5C\_cat}$ * SmokeFormer0to5C_cat ) + ($\beta_{SmokeCurrentC\_cat}$ * SmokeCurrentC_cat) + ($\beta_{DrinksLastWeekC\_rcs1}$ * DrinksLastWeekC_rcs1) + ($\beta_{DrinksLastWeekC\_rcs2}$ * DrinksLastWeekC_rcs2) + ($\beta_{FormerDrinkerC\_cat}$ * FormerDrinkerC_cat) + ($\beta_{FruitVegC\_rcs1}$ * FruitVegC_rcs1) + ($\beta_{FruitVegC\_rcs2}$ * FruitVegC_rcs2) + ($\beta_{PotatoC\_rcs1}$ * PotatoC_rcs1) + ($\beta_{PotatoC\_rcs2}$ * PotatoC_rcs2) + ($\beta_{JuiceC\_rcs1}$ * JuiceC_rcs1) + ($\beta_{JuiceC\_rcs2}$ * JuiceC_rcs2) + ($\beta_{PhysicalActivityC\_rcs1}$ * PhysicalActivityC_rcs1) + ($\beta_{PhysicalActivityC\_rcs2}$ * PhysicalActivityC_rcs2) + ($\beta_{HeartDisC\_cat}$ * HeartDisC_cat) + ($\beta_{StrokeC\_cat}$ * StrokeC_cat) + ($\beta_{DiabetesC\_cat}$ * DiabetesC_cat) + ($\beta_{MoodC\_cat}$ * MoodC_cat) + ($\beta_{HypertensionC\_cat}$* HypertensionC_cat) + ($\beta_{COPDC\_cat}$* COPDC_cat) + ($\beta_{BMIC\_rcs1}$* BMIC_rcs1) + ($\beta_{BMIC\_rcs2}$ * BMIC_rcs2) + ($\beta_{NeedHelp1C\_cat}$ * NeedHelp1C_cat) + ($\beta_{NeedHelp2C\_cat}$ * NeedHelp2C_cat) + ($\beta_{NeedHelp3C\_cat}$ * NeedHelp3C_cat) + ($\beta_{NeedHelp4C\_cat}$ * NeedHelp4C_cat) + ($\beta_{NeedHelp5C\_cat}$ * NeedHelp5C_cat) + ($\beta_{NeedHelp16C\_cat}$ * NeedHelp6C_cat) + ($\beta_{SurveyCycle2C\_cat}$* SurveyCycle2C_cat) + ($\beta_{SurveyCycle3C\_cat}$ * SurveyCycle3C_cat) + ($\beta_{SurveyCycle4C\_cat}$ * SurveyCycle4C_cat) + ($\beta_{SurveyCycle5C\_cat}$ * SurveyCycle5C_cat) + ($\beta_{SurveyCycle6C\_cat}$ * SurveyCycle6C_cat) + ($\beta_{AgeCXPackYearsC\_int}$ * AgeCXPackYearsC_int) + ($\beta_{AgeCXSmokeFormer5PlusC\_int}$ * AgeCXSmokeFormer5PlusC_int) + ($\beta_{AgeCXSmokeFormer0to5C\_int}$ * AgeCXSmokeFormer0to5C_int) + ($\beta_{AgeCXSmokeCurrentC\_int}$ * AgeCXSmokeCurrentC_int) + + ($\beta_{AgeCXDrinksLastWeekC\_int}$ * AgeCXDrinksLastWeekC_int ) + ($\beta_{AgeCXFruitVegC\_int}$* AgeCXFruitVegC_int) + ($\beta_{AgeCXPotatoC\_int}$ * AgeCXPotatoC_int) + ($\beta_{AgeCXJuiceC\_int}$ * AgeCXJuiceC_int) + ($\beta_{AgeCXPhysicalActivityC\_int}$ * AgeCXPhysicalActivityC_int) + ($\beta_{AgeCXHeartDisC\_int}$ * AgeCXHeartDisC_int) + ($\beta_{AgeCXStrokeC\_int}$ * AgeCXStrokeC_int) +

($\beta_{AgeCXDiabetesC\_int}$ * AgeCXDiabetesC_int) + ($\beta_{AgeCXMoodC\_int}$ * AgeCXMoodC_int) + ($\beta_{AgeCXHypertensionC\_int}$ * AgeCXHypertensionC_int) + ($\beta_{AgeCXCOPDC\_int}$ * AgeCXCOPDC_int) + ($\beta_{AgeCXBMIC\_int}$ * AgeCXBMIC_int) + ($\beta_{AgeCXNeedHelp1C\_int}$ * AgeCXNeedHelp1C_int) + ($\beta_{AgeCXNeedHelp2C\_int}$ * AgeCXNeedHelp2C_int) + ($\beta_{AgeCXNeedHelp3C\_int}$ * AgeCXNeedHelp3C_int) + ($\beta_{AgeCXNeedHelp4C\_int}$ * AgeCXNeedHelp4C_int) + ($\beta_{AgeCXNeedHelp5C\_int}$ * AgeCXNeedHelp5C_int) + ($\beta_{AgeCXNeedHelp6C\_int}$ * AgeCXNeedHelp6C_int)

**Female DemPoRT risk score:**

$Score_{female}$ =

($\beta_{AgeC\_rcs1}$ * AgeC_rcs1) + ($\beta_{AgeC\_rcs2}$ * AgeC_rcs2) + ($\beta_{AgeC\_rcs3}$*AgeC_rcs3) + ($\beta_{AgeC\_rcs4}$ * AgeC_rcs4) + ($\beta_{EduHSGrad2C\_cat}$ * EduHSGrad2C_cat) + ($\beta_{EduSomePS2C\_cat}$ * EduSomePS2C_cat) + ($\beta_{EduPSGrad2C\_cat}$ * EduPSGrad2C_cat) + ($\beta_{MSDivorcedC\_cat}$ * MSDivorcedC_cat) + ($\beta_{MSWidowedC\_cat}$ * MSWidowedC_cat) + ($\beta_{MSSingleC\_cat}$ * MSSingleC_cat) + ($\beta_{StressNotVeryC\_cat}$ * StressNotVeryC_cat) + ($\beta_{StressABitC\_cat}$ * StressABitC_cat) + ($\beta_{StressQuiteC\_cat}$ * StressQuiteC_cat) + ($\beta_{StressExtreamlyC\_cat}$ * StressExtreamlyC_cat) + ($\beta_{PackYearsC\_rcs1}$ * PackYearsC_rcs1) + ($\beta_{PackYearsC\_rcs2}$ * PackYearsC_rcs2) + ($\beta_{SmokeFormer5PlusC\_cat}$ * SmokeFormer5PlusC_cat) + ($\beta_{SmokeFormer0to5C\_cat}$ * SmokeFormer0to5C_cat ) + ($\beta_{SmokeCurrentC\_cat}$ * SmokeCurrentC_cat) + ($\beta_{DrinksLastWeekC\_rcs1}$ * DrinksLastWeekC_rcs1) + ($\beta_{DrinksLastWeekC\_rcs2}$ * DrinksLastWeekC_rcs2) + ($\beta_{FormerDrinkerC\_cat}$ * FormerDrinkerC_cat) + ($\beta_{FruitVegC\_rcs1}$ * FruitVegC_rcs1) + ($\beta_{FruitVegC\_rcs2}$ * FruitVegC_rcs2) + ($\beta_{PotatoC\_rcs1}$ * PotatoC_rcs1) + ($\beta_{PotatoC\_rcs2}$ * PotatoC_rcs2) + ($\beta_{JuiceC\_rcs1}$ * JuiceC_rcs1) + ($\beta_{JuiceC\_rcs2}$ * JuiceC_rcs2) + ($\beta_{PhysicalActivityC\_rcs1}$ * PhysicalActivityC_rcs1) + ($\beta_{PhysicalActivityC\_rcs2}$ * PhysicalActivityC_rcs2) + ($\beta_{HeartDisC\_cat}$ * HeartDisC_cat) + ($\beta_{StrokeC\_cat}$ * StrokeC_cat) + ($\beta_{DiabetesC\_cat}$ * DiabetesC_cat) + ($\beta_{MoodC\_cat}$ * MoodC_cat) + ($\beta_{HypertensionC\_cat}$* HypertensionC_cat) + ($\beta_{COPDC\_cat}$* COPDC_cat) + ($\beta_{EpilepsyC\_cat}$* EpilepsyC_cat) + ($\beta_{BMIC\_rcs1}$* BMIC_rcs1) + ($\beta_{BMIC\_rcs2}$ * BMIC_rcs2) + ($\beta_{NeedHelp1C\_cat}$ * NeedHelp1C_cat) + ($\beta_{NeedHelp2C\_cat}$ * NeedHelp2C_cat) + ($\beta_{NeedHelp3C\_cat}$ * NeedHelp3C_cat) + ($\beta_{NeedHelp4C\_cat}$ * NeedHelp4C_cat) + ($\beta_{NeedHelp5C\_cat}$ * NeedHelp5C_cat) + ($\beta_{NeedHelp16C\_cat}$ * NeedHelp6C_cat) + ($\beta_{SurveyCycle2C\_cat}$* SurveyCycle2C_cat) + ($\beta_{SurveyCycle3C\_cat}$ * SurveyCycle3C_cat) + ($\beta_{SurveyCycle4C\_cat}$ * SurveyCycle4C_cat) + ($\beta_{SurveyCycle5C\_cat}$ * SurveyCycle5C_cat) + ($\beta_{SurveyCycle6C\_cat}$ * SurveyCycle6C_cat) + ($\beta_{AgeCXPackYearsC\_int}$ * AgeCXPackYearsC_int) + ($\beta_{AgeCXSmokeFormer5PlusC\_int}$ * AgeCXSmokeFormer5PlusC_int) + ($\beta_{AgeCXSmokeFormer0to5C\_int}$ * AgeCXSmokeFormer0to5C_int) + ($\beta_{AgeCXSmokeCurrentC\_int}$ * AgeCXSmokeCurrentC_int) +
($\beta_{AgeCXDrinksLastWeekC\_int}$ * AgeCXDrinksLastWeekC_int ) + ($\beta_{AgeCXFruitVegC\_int}$* AgeCXFruitVegC_int) + ($\beta_{AgeCXPotatoC\_int}$ * AgeCXPotatoC_int) + ($\beta_{AgeCXJuiceC\_int}$ * AgeCXJuiceC_int) + ($\beta_{AgeCXPhysicalActivityC\_int}$ * AgeCXPhysicalActivityC_int) + ($\beta_{AgeCXHeartDisC\_int}$ * AgeCXHeartDisC_int) + ($\beta_{AgeCXStrokeC\_int}$ * AgeCXStrokeC_int) + ($\beta_{AgeCXDiabetesC\_int}$ * AgeCXDiabetesC_int) + ($\beta_{AgeCXMoodC\_int}$ * AgeCXMoodC_int) + ($\beta_{AgeCXHypertensionC\_int}$ * AgeCXHypertensionC_int) + ($\beta_{AgeCXCOPDC\_int}$ * AgeCXCOPDC_int) + ($\beta_{AgeCXEpilepsyC\_int}$ * AgeCXEpilepsyC_int) + ($\beta_{AgeCXBMIC\_int}$ * AgeCXBMIC_int) + ($\beta_{AgeCXNeedHelp1C\_int}$ * AgeCXNeedHelp1C_int) + ($\beta_{AgeCXNeedHelp2C\_int}$ * AgeCXNeedHelp2C_int) + ($\beta_{AgeCXNeedHelp3C\_int}$ * AgeCXNeedHelp3C_int) +

($\beta_{AgeCXNeedHelp4C\_int}$ * AgeCXNeedHelp4C_int) + ($\beta_{AgeCXNeedHelp5C\_int}$ * AgeCXNeedHelp5C_int) + ($\beta_{AgeCXNeedHelp6C\_int}$ * AgeCXNeedHelp6C_int)

**Components of restricted cubic spline with j = 1…k knots:**

X_rcs1 = $X$

$$\text{X\_rcs}_{j+1} = \left(\frac{(X-knot_j)}{(knot_k-knot_1)^{2/3}}\right)_+^3 + (knot_{k-1} - knot_j)\left(\frac{(X-knot_k)}{(knot_k-knot_1)^{2/3}}\right)_+^3 - (knot_k - knot_j)\left(\frac{X-knot_{k-1}}{(knot_k-knot_1)^{2/3}}\right)_+^3 / (knot_k - knot_{k-1})$$

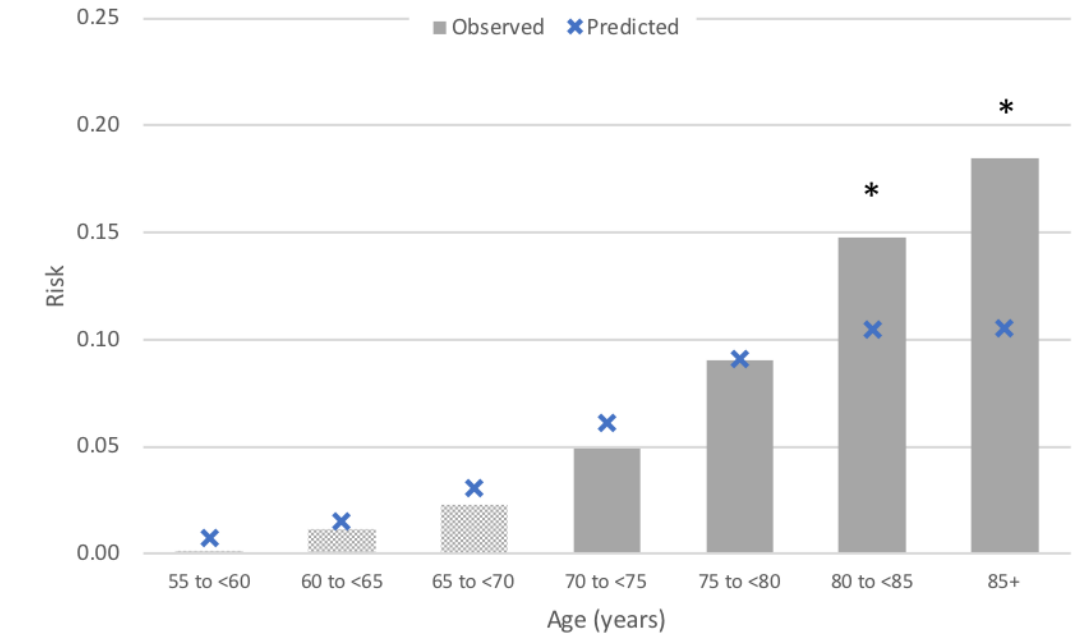
Where:

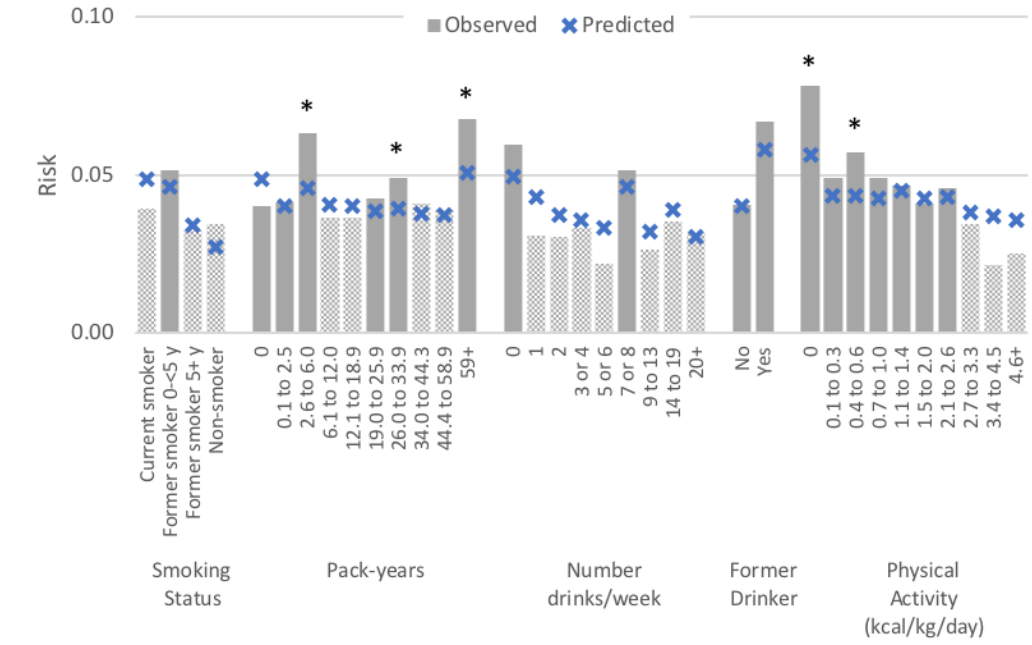
$$(X - knot)_+ = X - knot, \; X - knot > 0$$
$$= 0, \qquad X - knot \leq 0$$

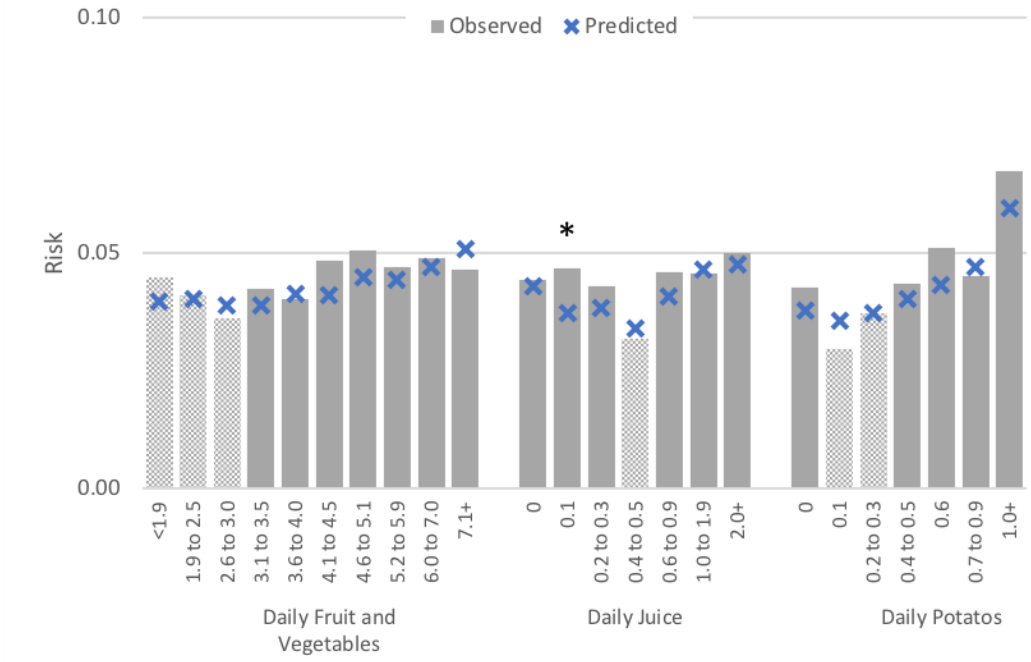
(ref: http://biostat.mc.vanderbilt.edu/wiki/pub/Main/SasMacros/survrisk.txt)

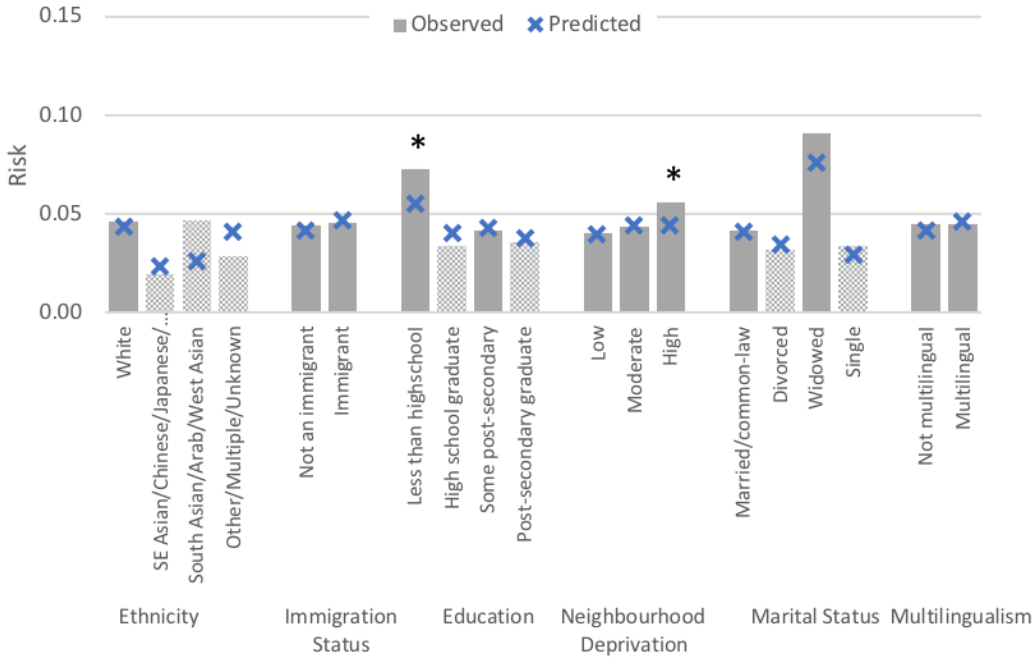
**Appendix 3-6.** Observed versus predicted for subgroups in the validation model – males

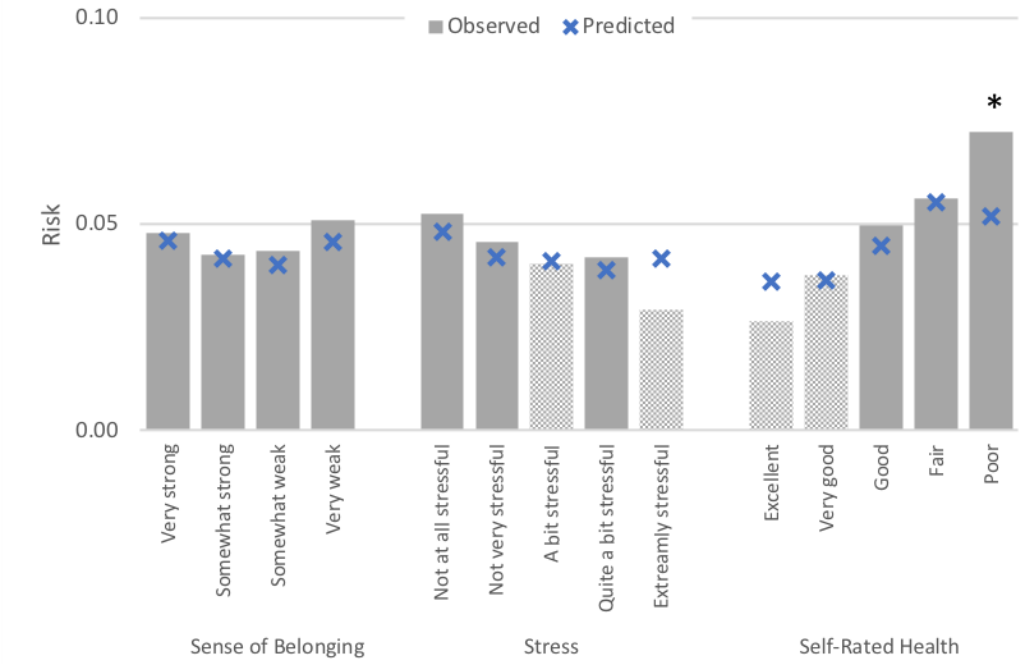
- Light patterned bars indicate less than 5% dementia prevalence within the subgroup
- Asterix indicates that observed and predicted risks are more than 20% different; only evaluated within subgroups with at least 5% dementia prevalence

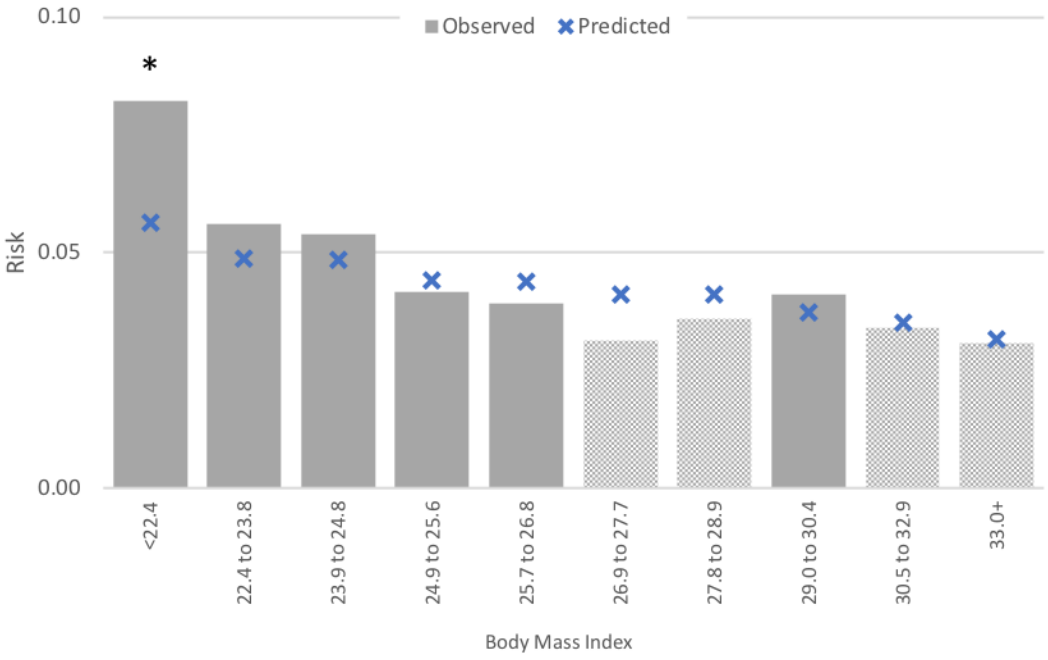


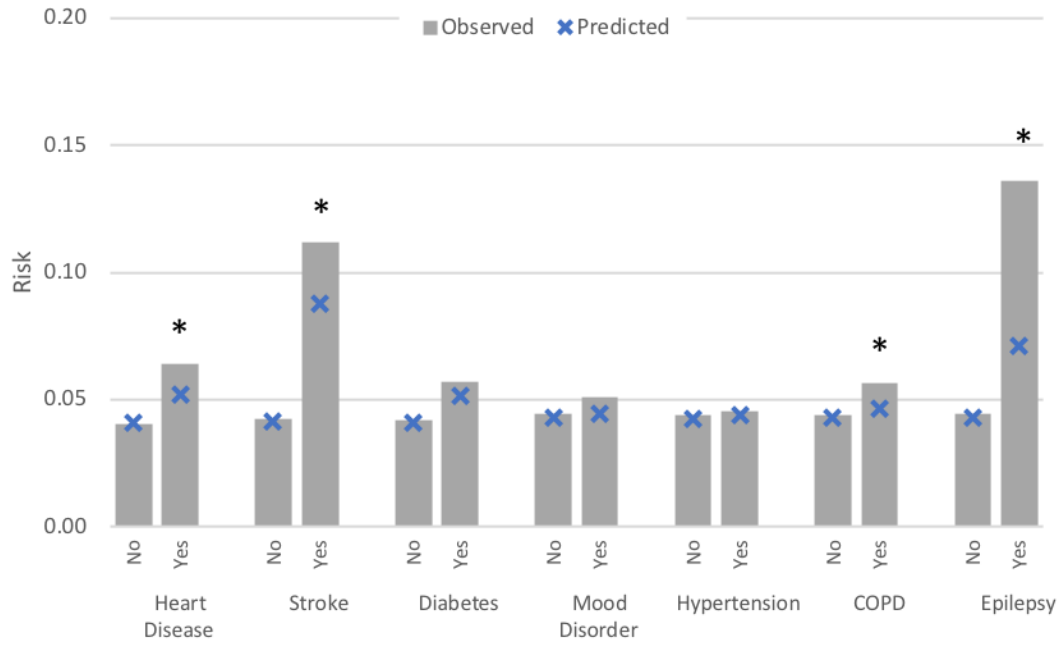



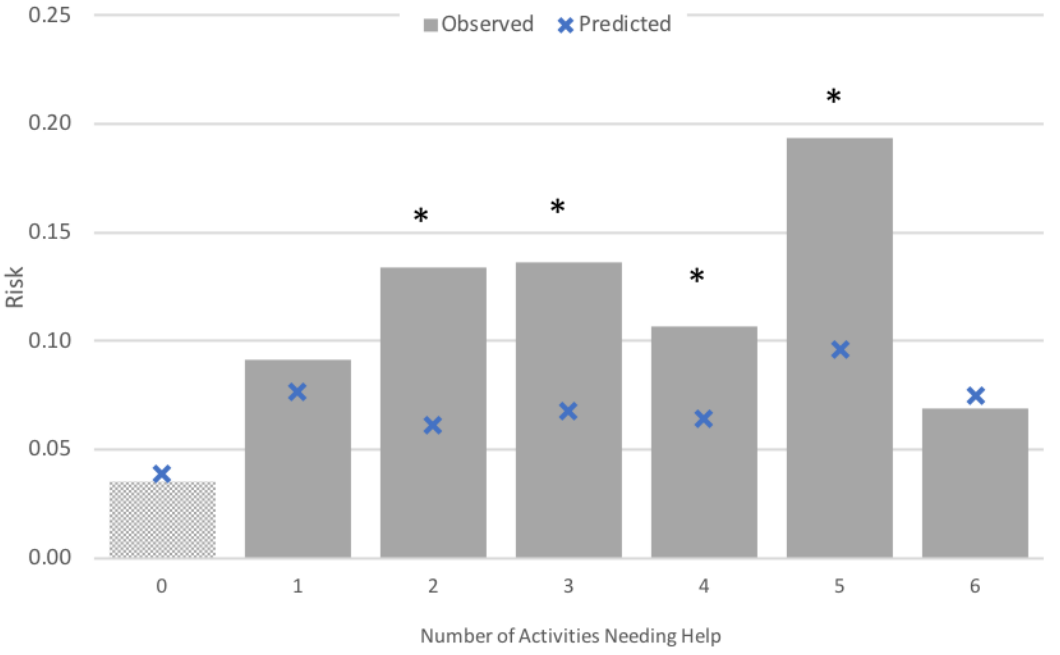

0.10
0.05
0.00
Risk
Observed
Predicted
*
<1.9
1.9 to 2.5
2.6 to 3.0
3.1 to 3.5
3.6 to 4.0
4.1 to 4.5
4.6 to 5.1
5.2 to 5.9
6.0 to 7.0
7.1+
0
0.1
0.2 to 0.3
0.4 to 0.5
0.6 to 0.9
1.0 to 1.9
2.0+
0
0.1
0.2 to 0.3
0.4 to 0.5
0.6
0.7 to 0.9
1.0+
Daily Fruit and Vegetables
Daily Juice
Daily Potatos


0.15
0.10
0.05
0.00
Risk
Observed
Predicted
*
*
White
SE Asian/Chinese/Japanese/...
South Asian/Arab/West Asian
Other/Multiple/Unknown
Not an immigrant
Immigrant
Less than highschool
High school graduate
Some post-secondary
Post-secondary graduate
Low
Moderate
High
Married/common-law
Divorced
Widowed
Single
Not multilingual
Multilingual
Ethnicity
Immigration Status
Education
Neighbourhood Deprivation
Marital Status
Multilingualism


0.10
Observed
Predicted
Risk
0.05
0.00
*
Very strong
Somewhat strong
Somewhat weak
Very weak
Not at all stressful
Not very stressful
A bit stressful
Quite a bit stressful
Extreamly stressful
Excellent
Very good
Good
Fair
Poor
Sense of Belonging
Stress
Self-Rated Health

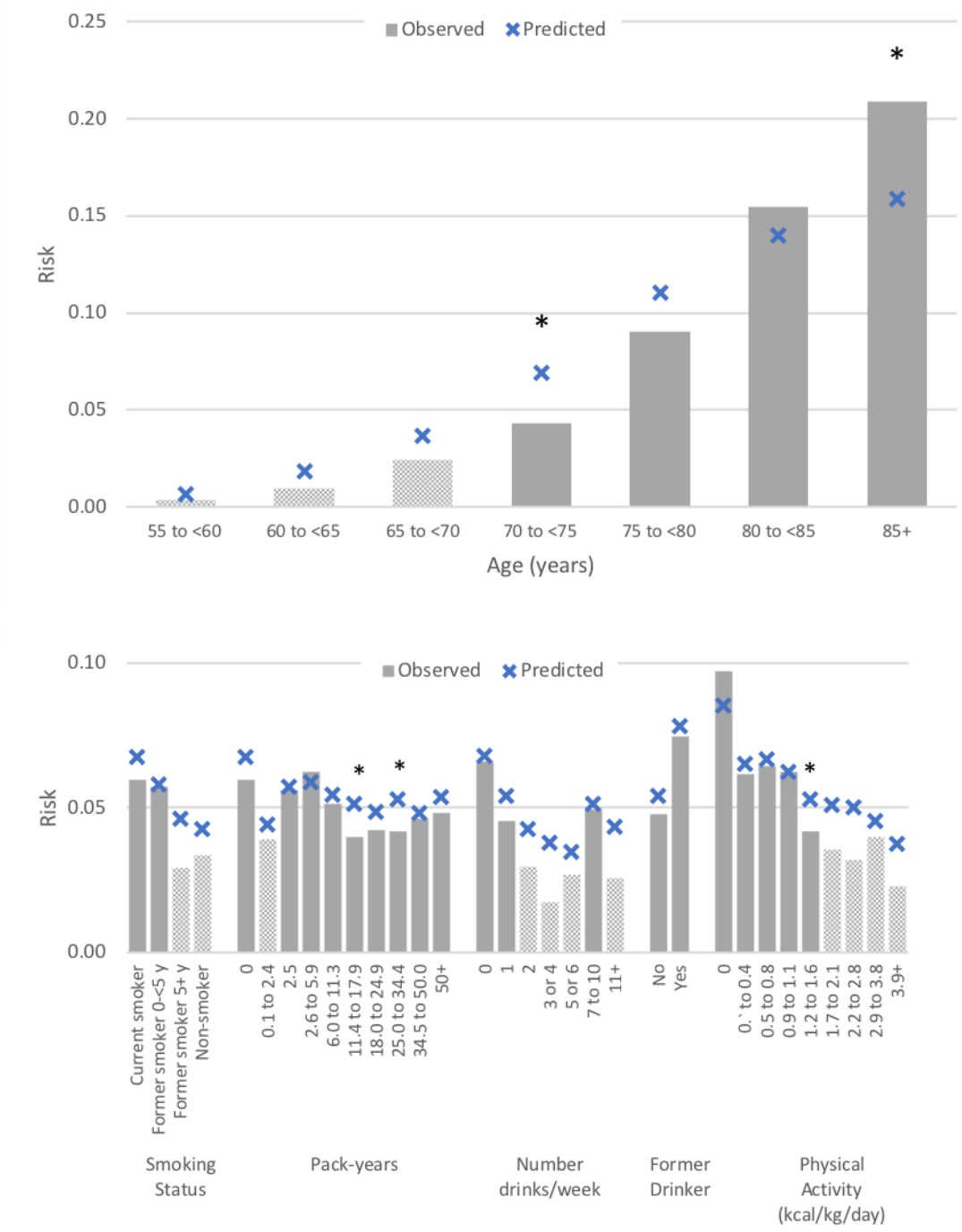

0.10
Observed
Predicted
*
Risk
0.05
0.00
<22.4
22.4 to 23.8
23.9 to 24.8
24.9 to 25.6
25.7 to 26.8
26.9 to 27.7
27.8 to 28.9
29.0 to 30.4
30.5 to 32.9
33.0+
Body Mass Index

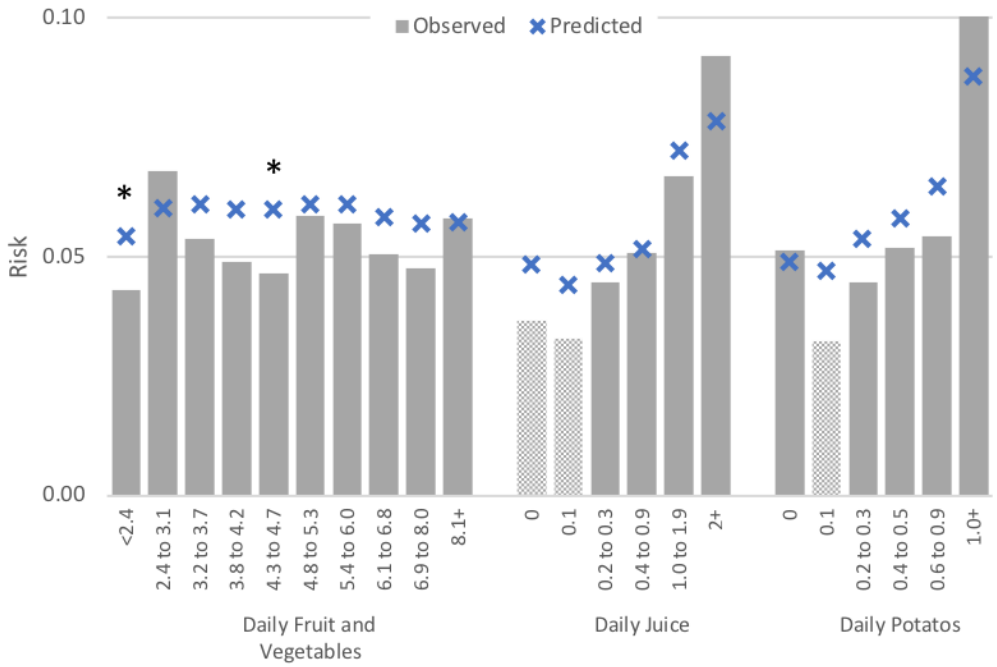

Observed
Predicted
0.20
0.15
0.10
0.05
0.00
Risk
*
No
Yes
Heart Disease
Stroke
Diabetes
Mood Disorder
Hypertension
COPD
Epilepsy

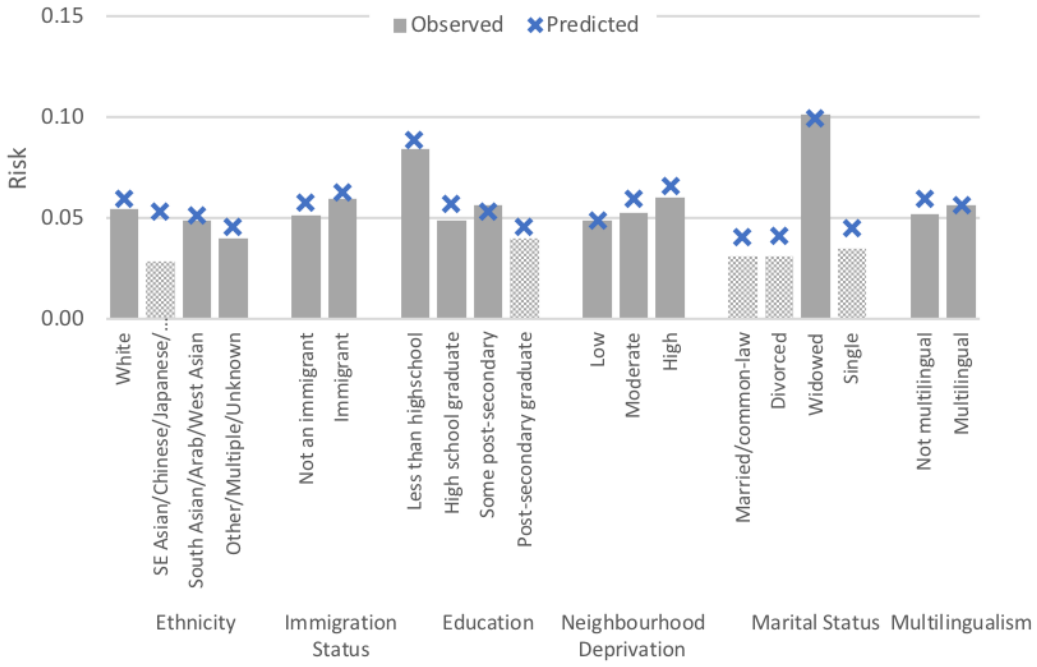

Observed
Predicted
0.25
0.20
0.15
0.10
0.05
0.00
Risk
*
0
1
2
3
4
5
6
Number of Activities Needing Help

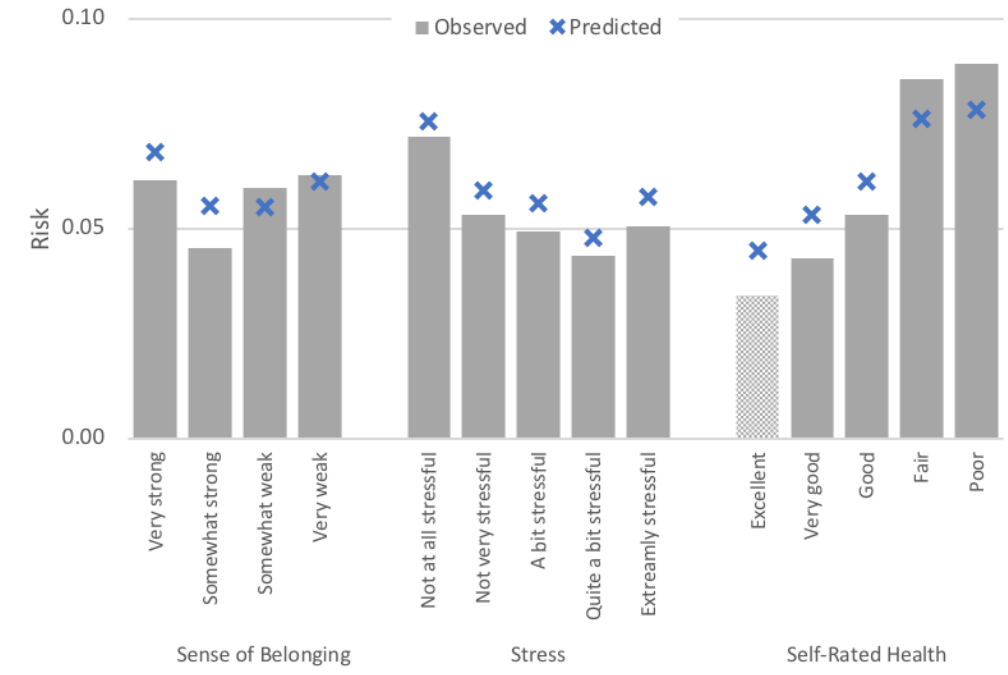
**Appendix 3-7.** Observed versus predicted for subgroups in the validation model – females

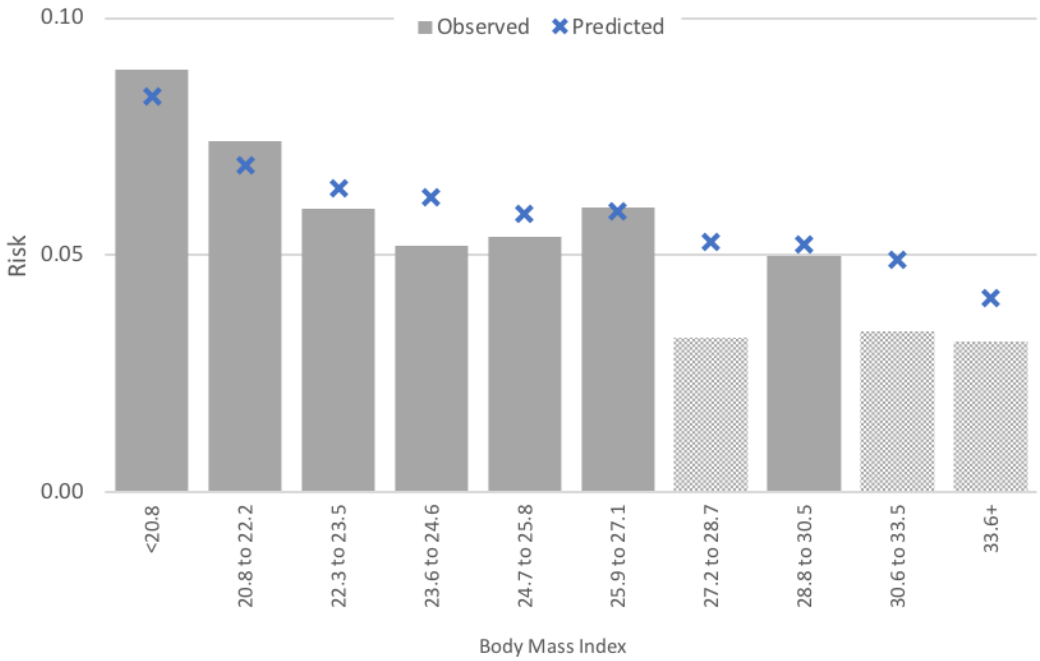
- Light patterned bars indicate less than 5% dementia prevalence within the subgroup
- Asterix indicates that observed and predicted risks are more than 20% different; only evaluated within subgroups with at least 5% dementia prevalence

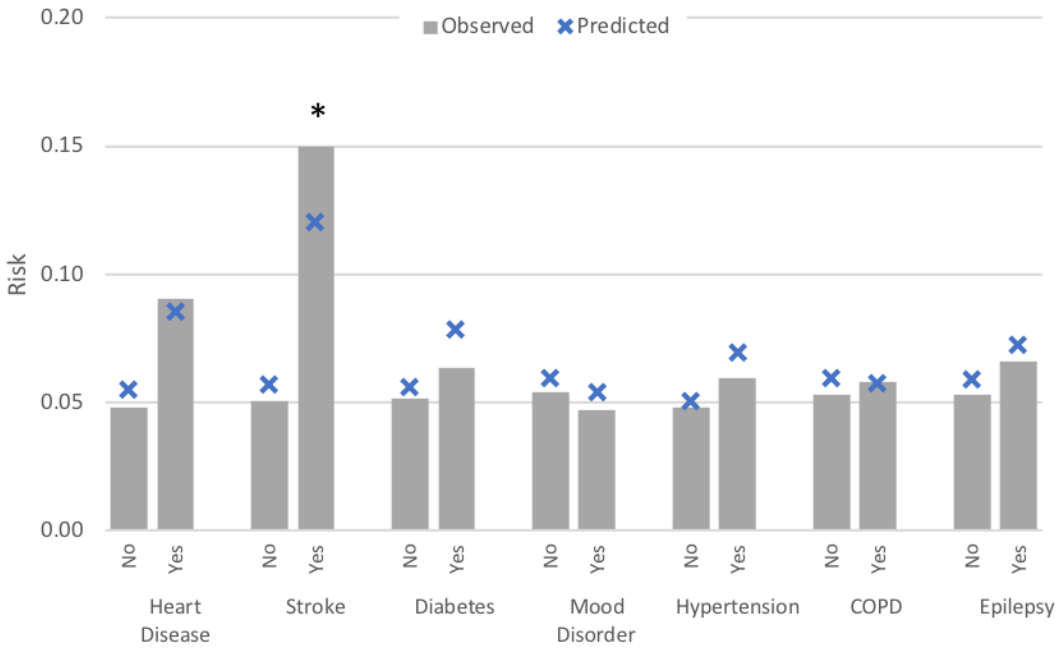


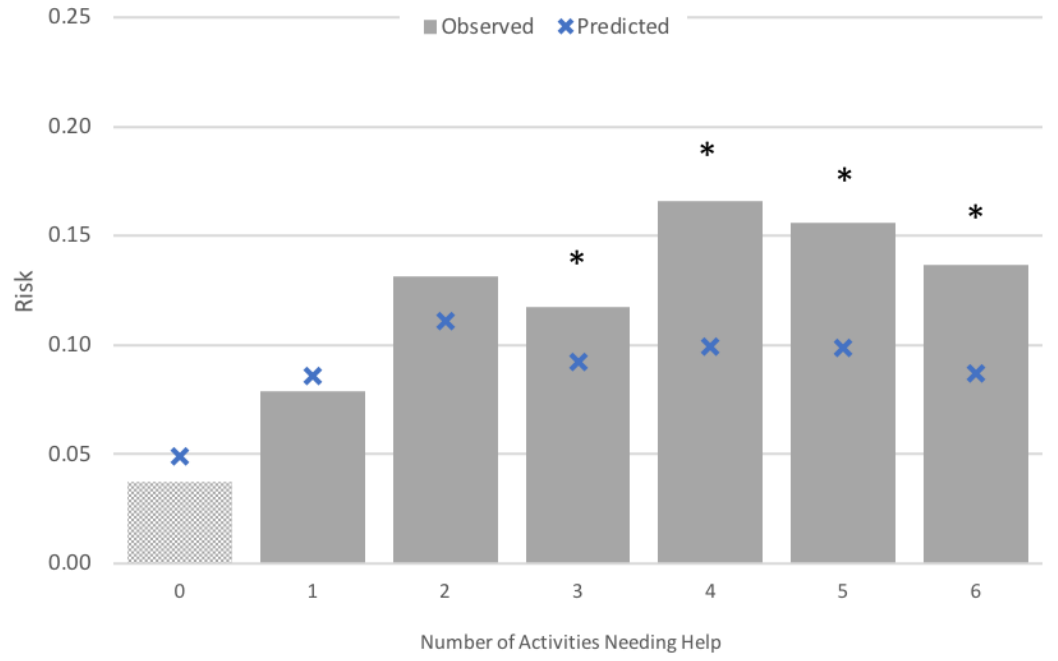

0.10
0.05
0.00
Risk
Observed
Predicted
*
*
<2.4
2.4 to 3.1
3.2 to 3.7
3.8 to 4.2
4.3 to 4.7
4.8 to 5.3
5.4 to 6.0
6.1 to 6.8
6.9 to 8.0
8.1+
Daily Fruit and Vegetables
0
0.1
0.2 to 0.3
0.4 to 0.9
1.0 to 1.9
2+
Daily Juice
0
0.1
0.2 to 0.3
0.4 to 0.5
0.6 to 0.9
1.0+
Daily Potatos


0.15
0.10
0.05
0.00
Risk
Observed
Predicted
White
SE Asian/Chinese/Japanese/...
South Asian/Arab/West Asian
Other/Multiple/Unknown
Ethnicity
Not an immigrant
Immigrant
Immigration Status
Less than highschool
High school graduate
Some post-secondary
Post-secondary graduate
Education
Low
Moderate
High
Neighbourhood Deprivation
Married/common-law
Divorced
Widowed
Single
Marital Status
Not multilingual
Multilingual
Multilingualism


0.10
0.05
0.00
Risk
Observed
Predicted
Very strong
Somewhat strong
Somewhat weak
Very weak
Not at all stressful
Not very stressful
A bit stressful
Quite a bit stressful
Extreamly stressful
Excellent
Very good
Good
Fair
Poor
Sense of Belonging
Stress
Self-Rated Health


0.10
0.05
0.00
Risk
Observed
Predicted
<20.8
20.8 to 22.2
22.3 to 23.5
23.6 to 24.6
24.7 to 25.8
25.9 to 27.1
27.2 to 28.7
28.8 to 30.5
30.6 to 33.5
33.6+
Body Mass Index


Observed
Predicted
0.20
0.15
0.10
0.05
0.00
Risk
*
No
Yes
Heart Disease
Stroke
Diabetes
Mood Disorder
Hypertension
COPD
Epilepsy


Observed
Predicted
0.25
0.20
0.15
0.10
0.05
0.00
Risk
*
0
1
2
3
4
5
6
Number of Activities Needing Help

## 7.2 CHAPTER 4 APPENDICIES

**Appendix 4-1.** Model and variable specification of models designed to estimate the effects smoking, alcohol, diet and physical activity on 5-year risk of dementia

*Model Specification*

- Model 1: Age + health behaviours (smoking, alcohol, diet, physical activity)
- Model 2: Model 1 + demographics (ethnicity + immigration + multilingual)
- **Model 3: Model 2 + social variables (education + marital status + deprivation)**
- Model 4: Model 3 + chronic conditions (heart disease, stroke, mood disorder, diabetes, high blood pressure, COPD, epilepsy, BMI)
- Model 5: Model 4 + upstream factors (self-rated health, self-perceived stress, sense of belonging and functional measures)

*Variable Specification*

- Variable specification is the same as found in the reduced DemPoRT models; see Table 3-1 in Chapter 3. All variables were centered on their means

**Appendix 4-2.** Method for calculating cause-deleted risk using a predictive algorithm

# Calculation of cause-deleted risk using a predictive algorithm

2019-06-24

Manuel DG, Bennett C, Fisher S, Tuna M

## Example

Zara is a current smoker. Her smoking status places her at a higher risk of dying compared to a never smoker. What is the effect of smoking on Zara's risk of dying? In other words, how much does smoking impact her risk of dying (as assessed using the Mortality Population Risk Tool (MPoRT) algorithm)? How much does smoking reduce her life expectancy?

## Two methodological approaches

This document describes two approaches to calculate the cause-deleted effect of an exposure[1] using a multivariable risk algorithm. For Zara, we will use the MPoRT all-cause mortality risk algorithm, which can be used to calculate 5-year risk of death and other mortality-based measures like life expectancy.

## Method 1 - using coefficients that are "within" the algorithm

The "within" algorithm approach uses all regression coefficients[2] for all exposures from "within" a single model multivariable model. This is the most common approach used in the literature.

There are three steps to calculate the effect of an exposure using the "within" approach:

Step 1.

Predictive risk is calculated for a respondent's profile. A respondent's profile is their exposure to risk factors (age, sex, smoking status, etc.). This calculation is called the "original risk".

Step 2.

"Cause-deleted risk" (aka effect-deleted risk) is calculated by setting an exposure to a reference (non-exposed) value (all other risk exposures remain unchanged). E.g., *smoking status* has potential exposures of *current smoker*, *former smoker* and *never smoker*. The reference for *smoking status* is usually *never smoker*.

Step 3.

The "cause-deleted effect" of an exposure is calculated by subtracting the "cause-deleted risk" (Step 2) from the "original risk" (Step 1).

---

[1] An exposures is also known as an input, risk factor, predictor, independent variable, covariate, or feature. '*Current smoking*' is Zara's exposure to the covariate '*smoking status*'

[2] See Appendix for a definition of a regression coefficient

*Zara:* Calculate Zara's risk of dying using her current profile of exposures, including her current smoking status ("original risk"). Second, calculate her risk of dying using the counterfactual scenario where she had never smoked ("cause-deleted risk"). The effect of smoking on Zara's risk of dying is the original risk estimate minus the cause-deleted risk estimate.

### Method 2 - using beta-coefficients from external sources

The "external" algorithm approach calculates the effect of a risk factor using a risk coefficient that is external to original multivariable model. The "external" coefficient of the risk is typically estimated from either reviews (systematic review or meta-analysis) or from a model specifically designed to estimate the causal relationship between the risk and the outcome.

The use of an external risk coefficient addresses three concerns. First, is a concern that coefficients in a multivariable risk algorithm may not reflect the causal association between the exposure and the outcome. Predictive algorithms frequently include variables that are on causal pathway between smoking and mortality to improve the predictive accuracy of the algorithm. The predictors may mediate or attenuate the effect of smoking, which would generally result in a lower calculated effect of smoking on mortality. Furthermore, predictive algorithms with many variables (such as machine learning algorithms) have an increased likelihood that coefficients are correlated. Including correlated coefficients reduces the effect size for coefficients representing the causal exposure.

Second, there are an increasing number of meta-analyses and systematic reviews that have assessed the causal relationship for risk factors. These reviews may reflect a consensus of the causal relationship.

Third, a predictive model may not have been developed to include all exposures. Furthermore, exposures may not be included in an algorithm because they have not been shown to improve predictive accuracy. However, a lack of predictive ability during algorithm development does not preclude a causal relationship between the exposure and outcome.

*Zara:* Zara also has cancer. Both smoking and cancer increase the risk of death, and smoking increases the risk of cancer. Because the MPoRT risk algorithm includes cancer as one of the predictors, the effect of smoking on Zara's risk of death will be split between the coefficients for smoking and cancer.

The use of "external" coefficients addresses the concern of mediation or risk attenuation in highly specified models such as MPoRT.

The "cause-deleted effect$_{external}$" is calculated with an external coefficient for the exposure of interest that is included in the model. The external coefficients are the causal coefficients (e.g., coefficients from a literature review or causal model.) There are 6 steps to calculate the effect of an exposure using the "external" approach.

Step 1.

Modify the original algorithm to include the external coefficient(s).

a. Remove the original regression coefficient(s) for the exposure. This step can be performed by assigning the sum of coefficients X exposure to zero (ßX = 0), or exposure to zero (X = 0). All coefficients for the exposure should be removed or set to zero.[3]
b. Add the new external coefficient(s) to the algorithm[4,5]

Step 2.

Risk is calculated using the modified algorithm created in Step 1 and the respondent's original profile. Note: this step is a replication of Method 1: Step 1, but with the new external coefficients. This is the "external coefficient risk".

Step 3.

"Cause-deleted risk ' " is calculated by setting an exposure to a reference (non-exposed) value (all other risk exposures remain unchanged). This step is a similar to Method 1, Step 2.

Step 4.

The "cause-deleted effect$_{external}$" is calculated as "external coefficient risk" (Step 2) minus the "cause-deleted risk' "(Step 3).

Step 5.

Risk is calculated using the original algorithm and the original respondent's profile. This is the same step as in Method 1, Step 1 ("original risk").

Step 6.

The "cause-deleted risk$_{external}$" is calculated by subtracting the "cause-deleted effect$_{external}$" (Step 4) from the "original risk" (Step 5).

*Zara:* **To calculate the impact of Zara's current smoking status on her mortality risk using method 2:**

1) Calculate Zara's mortality risk using the modified algorithm with external coefficients for smoking added and original coefficients for smoking removed ("external coefficient risk").
2) Change her smoking status to *never smoker*.
3) Calculate her mortality risk a second time using the modified algorithm ("cause-deleted risk' "). Note that we will only use to calculate cause-deleted effect'$_{external}$ (next Step 5). The preferred estimate "cause-deleted risk$_{external}$"(without ') is calculated in step 6.

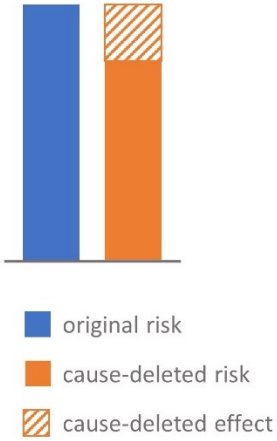
[3] Remember to not only remove the main effects but also, for example, interaction terms. Also be careful to remove risk factors that are a combination of several coefficients such as smoking status that may include the amount of smoking (pack-years of smoking), in addition to smoking status (current, former, never).

[4] It is a good practice is to add the coefficient using the same reference as in the original algorithm. For example, center the new coefficient, if the original coefficient was centered, use never-smokers as the reference is never-smokers were used as the reference in the original algorithm.

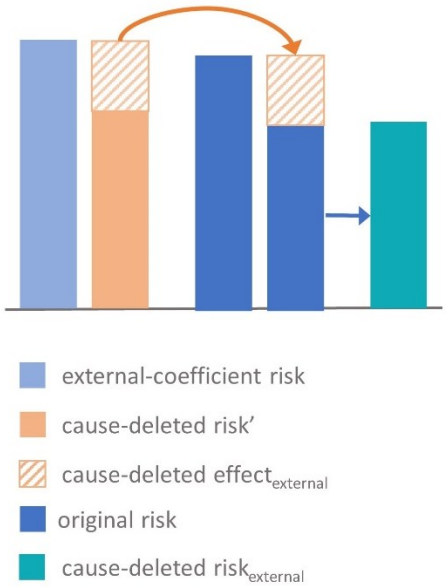
[5] The coefficient must be of the same transformed type as the model. For example, if the algorithm is a Cox proportional hazard model, then coefficient must be a relative hazard or an approximation of relative hazard.

4) Subtract "cause-deleted $risk_{external}$" (step 4) from "external coefficient risk" (step 3). This is the "cause-deleted $effect_{external}$".
5) Calculate Zara's mortality risk, given her current health profile, including *current smoker*, using the original algorithm ("original risk").
6) Subtract "cause-deleted $effect_{external}$" from "original risk". This is the cause-deleted $effect_{external}$

Method 1 - with algorithm coefficients




Method 2 – with external coefficients




## Additional considerations

The cause-effect and cause-deleted risk estimate can be calculated for complex risk factors and coefficients, including multiple risk factors, "J" shape or non-monotonic risk relationships, risk factors with multiple coefficients, or when risk factors are not in the original algorithm. Calculations can be performed using either method 1 or 2.

## Multiple risk factors

The cause-effect and cause-deleted risk estimate can be calculated for multiple risk factors separately or in combination using either method 1 or 2. When calculated in combination, the respondent's profile is simultaneously changed to be unexposed for multiple exposures. For example, to calculate Zara's cause-effect of smoking AND physical activity the unexposed profile would include both *never smoking* and *active physical activity* (e.g. 3 METS/day).
Note that the combined effect of multiple risk factors calculated simultaneously will general be less than the sum of individual effects.

**Appendix 4-3.** Beta coefficients for the association between unhealthy behaviours and dementia incidence, increasingly adjusted using models specified in Appendix 4-1[1]

| Variable | Description | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| PackYearsC_rcs1 | Pack-years of smoking; first RCS component | -0.00582 | -0.00602 | -0.00646 | -0.00664 | -0.00662 |
| PackYearsC_rcs2 | Pack-years of smoking; second RCS component | 0.014948 | 0.015083 | 0.015595 | 0.014959 | 0.01419 |
| SmokeFormer5PlusC_cat | Smoking: Former smoker quit ≥5 years ago | 0.084967 | 0.073578 | 0.069238 | 0.058049 | 0.056508 |
| SmokeFormer0to5C_cat | Smoking: Former smoker quit <5 years ago | 0.086319 | 0.076195 | 0.071511 | 0.042939 | 0.019027 |
| SmokeCurrentC_cat | Smoking: Current smoker | 0.368674 | 0.360797 | 0.335291 | 0.276314 | 0.228497 |
| DrinksLastWeekC_rcs1 | Drinks last week; first RCS component | -0.03435 | -0.03637 | -0.03307 | -0.02737 | -0.01954 |
| DrinksLastWeekC_rcs2 | Drinks last week; second RCS component | 0.155967 | 0.165692 | 0.147735 | 0.122634 | 0.086508 |
| FormerDrinkerC_cat | Former drinker | 0.223751 | 0.223667 | 0.209083 | 0.167692 | 0.128856 |
| FruitVegC_rcs1 | Daily fruit and veg; first RCS component | -0.09468 | -0.09156 | -0.0826 | -0.08008 | -0.07619 |
| FruitVegC_rcs2 | Daily fruit and veg; second RCS component | 0.10572 | 0.104114 | 0.098716 | 0.090126 | 0.084842 |
| PotatoC_rcs1 | Daily potato; first RCS component | 0.062118 | 0.00535 | 0.031071 | 0.057293 | 0.063947 |
| PotatoC_rcs2 | Daily potato; second RCS component | 0.217499 | 0.271011 | 0.219815 | 0.178374 | 0.144313 |
| JuiceC_rcs1 | Daily juice; first RCS component | 0.046368 | 0.033952 | 0.03246 | 0.06931 | 0.055872 |
| JuiceC_rcs2 | Daily juice; second RCS component | 0.057781 | 0.066545 | 0.063154 | 0.037852 | 0.044655 |
| PhysicalActivityC_rcs1 | Daily physical activity; first RCS component | -0.17866 | -0.18058 | -0.17304 | -0.15126 | -0.09235 |
| PhysicalActivityC_rcs2 | Daily physical activity; second RCS component | 0.152625 | 0.153969 | 0.146899 | 0.127723 | 0.078595 |
| AgeCXPackYearsC_int | Interaction: age and pack years of smoking | -3.9E-05 | -2.9E-05 | -2.7E-05 | 1.62E-05 | 4.02E-05 |
| AgeCXSmokeFormer5PlusC_int | Interaction: age and former smoker quit ≥5 years ago | -0.00378 | -0.00363 | -0.00383 | -0.00394 | -0.00472 |
| AgeCXSmokeFormer0to5C_int | Interaction: age and former smoker quit <5 years ago | -0.00402 | -0.00411 | -0.00384 | -0.00302 | -0.00251 |
| AgeCXSmokeCurrentC_int | Interaction: age and current smoker | -0.00167 | -0.00199 | -0.00214 | -0.00067 | -0.0027 |
| AgeCXDrinksLastWeekC_int | Interaction: age and drinks last week | 0.000701 | 0.000696 | 0.000731 | 0.000701 | 0.000587 |
| AgeCXFormerDrinkerC_int | Interaction: age and former drinker | -5.6E-05 | -0.00022 | 0.000363 | 0.001873 | 0.002079 |
| AgeCXFruitVegC_int | Interaction: age and daily fruit and vegetable | -0.00017 | -0.00025 | -0.00031 | -0.00018 | 0.000103 |
| AgeCXJuiceC_int | Interaction: age and daily juice | -0.00696 | -0.00677 | -0.00652 | -0.00731 | -0.00748 |
| AgeCXPotatoC_int | Interaction: age and daily potato | -0.00681 | -0.00684 | -0.00683 | -0.00803 | -0.00659 |
| AgeCXPhysicalActivityC_int | Interaction: age and physical activity | 0.001198 | 0.00124 | 0.001186 | 0.000812 | 0.000781 |

[1] All variables were centered on their means

**Appendix 4-4.** Beta coefficients for the association between unhealthy behaviours and dementia incidence, increasingly adjusted using models specified in Appendix 4-1[1]

| Variable | Description | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| PackYearsC_rcs1 | Pack-years of smoking; first RCS component | 0.010887 | 0.010921 | 0.009924 | 0.010306 | 0.009976 |
| PackYearsC_rcs2 | Pack-years of smoking; second RCS component | -0.0553 | -0.05566 | -0.05128 | -0.06031 | -0.06119 |
| SmokeFormer5PlusC_cat | Smoking: Former smoker quit ≥5 years ago | 0.022246 | 0.022802 | 0.030925 | 0.024603 | 0.03432 |
| SmokeFormer0to5C_cat | Smoking: Former smoker quit <5 years ago | 0.236937 | 0.236652 | 0.23816 | 0.205288 | 0.187921 |
| SmokeCurrentC_cat | Smoking: Current smoker | 0.192243 | 0.193289 | 0.190501 | 0.115902 | 0.112241 |
| DrinksLastWeekC_rcs1 | Drinks last week; first RCS component | -0.06341 | -0.0631 | -0.0555 | -0.04401 | -0.03335 |
| DrinksLastWeekC_rcs2 | Drinks last week; second RCS component | 0.070538 | 0.070206 | 0.062105 | 0.048146 | 0.036415 |
| FormerDrinkerC_cat | Former drinker | 0.256236 | 0.256673 | 0.249442 | 0.192415 | 0.15339 |
| FruitVegC_rcs1 | Daily fruit and veg; first RCS component | -0.09932 | -0.09921 | -0.09006 | -0.08359 | -0.07905 |
| FruitVegC_rcs2 | Daily fruit and veg; second RCS component | 0.073798 | 0.073291 | 0.068481 | 0.062343 | 0.058129 |
| PotatoC_rcs1 | Daily potato; first RCS component | 0.101306 | 0.107123 | 0.120918 | 0.107386 | 0.111653 |
| PotatoC_rcs2 | Daily potato; second RCS component | 0.155498 | 0.150236 | 0.127188 | 0.114028 | 0.092158 |
| JuiceC_rcs1 | Daily juice; first RCS component | 0.188039 | 0.188708 | 0.182823 | 0.214781 | 0.218019 |
| JuiceC_rcs2 | Daily juice; second RCS component | -0.03832 | -0.03901 | -0.03913 | -0.06466 | -0.07825 |
| PhysicalActivityC_rcs1 | Daily physical activity; first RCS component | -0.16207 | -0.16204 | -0.15773 | -0.15621 | -0.08174 |
| PhysicalActivityC_rcs2 | Daily physical activity; second RCS component | 0.185962 | 0.18589 | 0.181845 | 0.183062 | 0.107841 |
| AgeCXPackYearsC_int | Interaction: age and pack years of smoking | -0.00012 | -0.00012 | -0.00011 | -4.9E-05 | -4.9E-05 |
| AgeCXSmokeFormer5PlusC_int | Interaction: age and former smoker quit ≥5 years ago | 0.002285 | 0.002304 | 0.002374 | 0.002713 | 0.002725 |
| AgeCXSmokeFormer0to5C_int | Interaction: age and former smoker quit <5 years ago | -0.00573 | -0.0058 | -0.00557 | -0.00491 | -0.00492 |
| AgeCXSmokeCurrentC_int | Interaction: age and current smoker | -0.00918 | -0.00926 | -0.00882 | -0.00892 | -0.00802 |
| AgeCXDrinksLastWeekC_int | Interaction: age and drinks last week | 0.000264 | 0.000265 | 0.000304 | -0.00014 | -0.0002 |
| AgeCXFormerDrinkerC_int | Interaction: age and former drinker | -0.00764 | -0.00766 | -0.00767 | -0.00464 | -0.00354 |
| AgeCXFruitVegC_int | Interaction: age and daily fruit and vegetable | 0.001422 | 0.001426 | 0.001384 | 0.001131 | 0.001244 |
| AgeCXJuiceC_int | Interaction: age and daily juice | -0.00532 | -0.0053 | -0.00513 | -0.00534 | -0.005 |
| AgeCXPotatoC_int | Interaction: age and daily potato | -0.00374 | -0.00382 | -0.00424 | -0.00426 | -0.00353 |
| AgeCXPhysicalActivityC_int | Interaction: age and physical activity | -0.00057 | -0.00056 | -0.00061 | -0.00157 | -0.00216 |

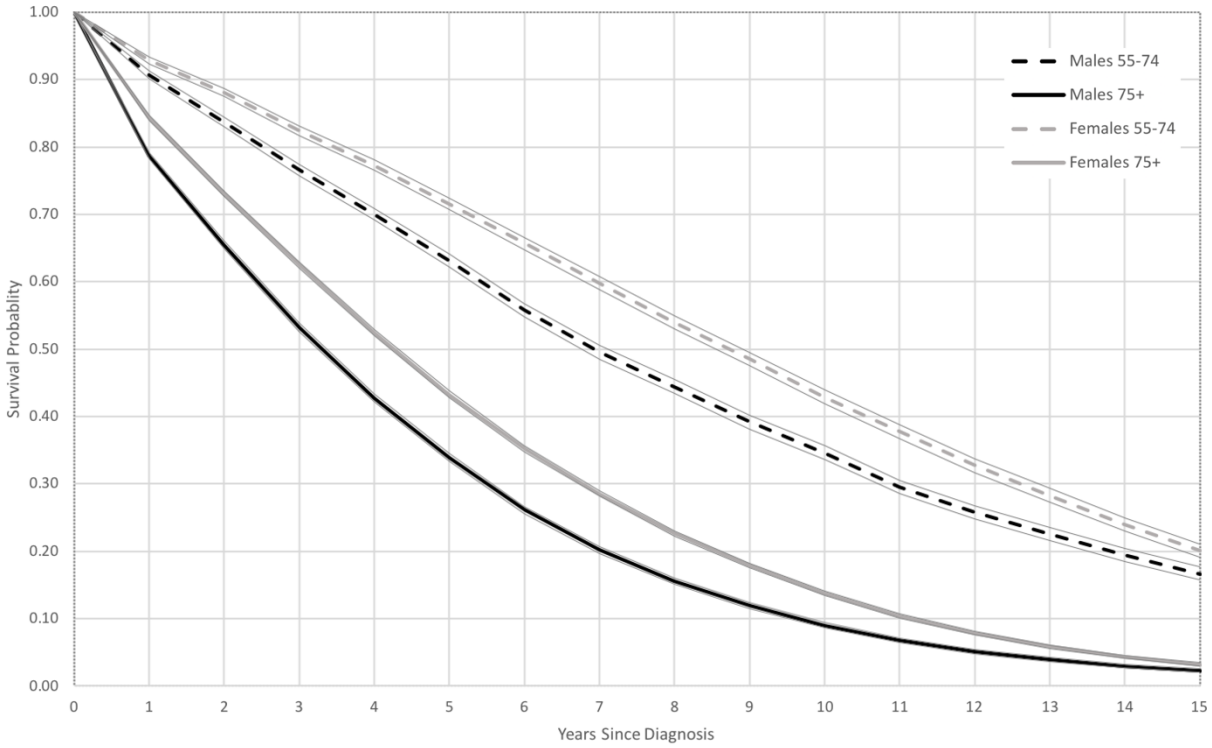
[1] All variables were centered on their means

**Appendix 4-5.** Five-year dementia risk attributed to unhealthy behaviours, evaluated using five increasing specified models described in Appendix 4-1

| | Smoking-deleted | | Poor diet-deleted | | Inactivity-deleted | | All three-deleted | |
|---|---|---|---|---|---|---|---|---|
| | Attributable Risk (%) | Attributable Cases | Attributable Risk (%) | Attributable Cases | Attributable Risk (%) | Attributable Cases | Attributable Risk (%) | Attributable Cases |
| **Combined** | | | | | | | | |
| Model 1 | 16.6 | 75,100 | 16.3 | 73,800 | 12.5 | 56,800 | 45.2 | 205,000 |
| Model 2 | 15.8 | 71,400 | 16.1 | 72,700 | 12.6 | 57,000 | 44.3 | 200,000 |
| Model 3 | 14.4 | 65,100 | 14.5 | 65,400 | 12.2 | 55,000 | 40.5 | 183,000 |
| Model 4 | 11.0 | 49,700 | 14.6 | 65,900 | 12.0 | 54,500 | 36.3 | 164,000 |
| Model 5 | 9.4 | 42,400 | 13.5 | 60,900 | 6.7 | 30,500 | 28.0 | 127,000 |
| **Males** | | | | | | | | |
| Model 1 | 11.0 | 19,900 | 13.2 | 23,900 | 13.7 | 24,900 | 39.1 | 70,800 |
| Model 2 | 8.8 | 16,000 | 12.6 | 22,800 | 13.9 | 25,200 | 36.6 | 66,400 |
| Model 3 | 6.1 | 11,100 | 10.9 | 19,700 | 13.3 | 24,000 | 31.1 | 56,500 |
| Model 4 | 2.1 | 3,740 | 11.2 | 20,300 | 11.7 | 21,200 | 25.1 | 45,600 |
| Model 5 | -1.6 | -2,832 | 9.7 | 17,600 | 6.6 | 11,900 | 14.6 | 26,600 |
| **Females** | | | | | | | | |
| Model 1 | 20.4 | 55,300 | 18.4 | 49,900 | 11.8 | 31,900 | 49.4 | 134,000 |
| Model 2 | 20.5 | 55,500 | 18.4 | 49,900 | 11.7 | 31,800 | 49.4 | 134,000 |
| Model 3 | 19.9 | 54,000 | 16.9 | 45,700 | 11.4 | 31,000 | 46.8 | 127,000 |
| Model 4 | 17.0 | 46,000 | 16.8 | 45,600 | 12.3 | 33,300 | 43.8 | 119,000 |
| Model 5 | 16.7 | 45,200 | 16.0 | 43,300 | 6.9 | 18,600 | 36.9 | 100,000 |

## 7.3 CHAPTER 5 APPENDICIES

**Appendix 5-1.** Age at diagnosis and sex-specific survival curves from dementia diagnosis, calculated using the life table standing cohort population




| | Survival Probability (95% CI) | | |
|---|---|---|---|
| | **1-year** | **5-year** | **10-year** |
| Males 55-74 years | 0.91 | 0.63 | 0.35 |
| Males 75+ years | 0.79 | 0.34 | 0.09 |
| Females 55-74 years | 0.93 | 0.72 | 0.43 |
| Females 75+ years | 0.84 | 0.43 | 0.14 |

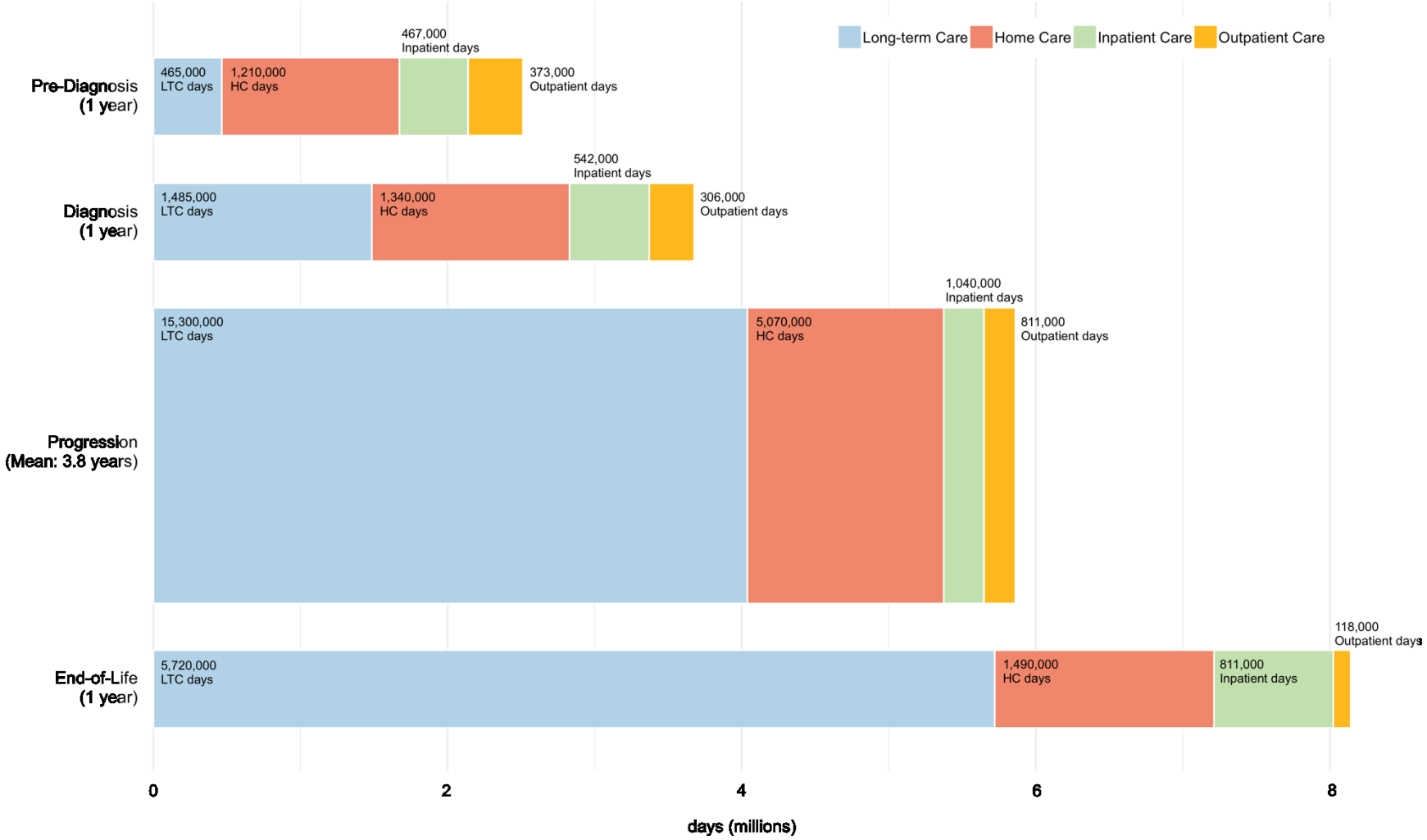
**Appendix 5-2.** Life expectancy and average number of days of long term care, home care, inpatient care and outpatient care received per year by individuals with dementia in Ontario 2014 to 2017, by years since diagnosis, and stratified by sex and age at diagnosis

| | Time interval (Years since diagnosis) | Interval-specific death rate (per 1,000) | Life expectancy in years[1] (95% CI) | Long-Term Care[2] (days) | Home Care[2] (days) | Inpatient Care[2] (days) | Outpatient Care[2] (days) | No Care[2] (days) |
|---|---|---|---|---|---|---|---|---|
| **Overall** | 0 to 1 | 172 | 5.8 (5.7, 5.8) | 749 | 236 | 93 | 41 | 1092 |
| | 1 to 2 | 136 | 5.8 (5.7, 5.8) | 825 | 229 | 72 | 36 | 1055 |
| | 2 to 3 | 149 | 5.5 (5.5, 5.6) | 848 | 212 | 66 | 32 | 974 |
| | 3 to 4 | 159 | 5.4 (5.3, 5.4) | 863 | 200 | 61 | 29 | 909 |
| | 4 to 5 | 171 | 5.2 (5.2, 5.2) | 873 | 188 | 58 | 27 | 857 |
| | 5 to 6 | 185 | 5.1 (5.0, 5.1) | 879 | 179 | 55 | 26 | 817 |
| | 6 to 7 | 185 | 5.0 (5.0, 5.1) | 887 | 172 | 53 | 24 | 792 |
| | 7 to 8 | 191 | 4.9 (4.9, 5.0) | 887 | 165 | 51 | 23 | 770 |
| | 8 to 9 | 192 | 4.8 (4.8, 4.9) | 884 | 158 | 49 | 22 | 755 |
| | 9 to 10 | 206 | 4.8 (4.7, 4.8) | 874 | 151 | 48 | 21 | 744 |
| | 10 to 11 | 213 | 4.7 (4.7, 4.8) | 869 | 147 | 48 | 21 | 747 |
| | 11 to 12 | 211 | 4.8 (4.7, 4.8) | 866 | 145 | 47 | 21 | 758 |
| | 12 to 13 | 210 | 4.8 (4.7, 4.9) | 859 | 144 | 47 | 21 | 768 |
| | 13 to 14 | 214 | 4.8 (4.7, 4.9) | 849 | 141 | 48 | 21 | 778 |
| | 14 to 15 | 213 | 4.8 (4.7, 4.9) | 844 | 140 | 48 | 21 | 791 |
| | 15+ | 207 | 4.8 (0, 0) | 842 | 139 | 48 | 21 | 798 |
| **Males 55-74** | 0 to 1 | 98 | 8.7 (8.6, 8.8) | 854 | 260 | 150 | 74 | 1835 |
| | 1 to 2 | 81 | 8.5 (8.4, 8.7) | 897 | 253 | 127 | 67 | 1769 |
| | 2 to 3 | 89 | 8.2 (8.1, 8.4) | 907 | 243 | 118 | 62 | 1665 |
| | 3 to 4 | 90 | 7.9 (7.8, 8.1) | 911 | 234 | 111 | 59 | 1576 |
| | 4 to 5 | 103 | 7.6 (7.5, 7.8) | 904 | 225 | 105 | 55 | 1491 |
| | 5 to 6 | 124 | 7.4 (7.2, 7.6) | 897 | 219 | 101 | 53 | 1428 |
| | 6 to 7 | 119 | 7.3 (7.1, 7.5) | 898 | 218 | 98 | 52 | 1398 |
| | 7 to 8 | 109 | 7.2 (7.0, 7.3) | 889 | 214 | 97 | 51 | 1362 |
| | 8 to 9 | 126 | 6.9 (6.7, 7.1) | 864 | 208 | 95 | 49 | 1312 |
| | 9 to 10 | 123 | 6.8 (6.6, 7.0) | 847 | 205 | 94 | 48 | 1283 |
| | 10 to 11 | 160 | 6.6 (6.4, 6.8) | 822 | 201 | 93 | 47 | 1248 |
| | 11 to 12 | 135 | 6.7 (6.5, 6.9) | 821 | 206 | 95 | 48 | 1263 |
| | 12 to 13 | 134 | 6.6 (6.4, 6.8) | 802 | 208 | 95 | 47 | 1242 |
| | 13 to 14 | 146 | 6.4 (6.2, 6.6) | 774 | 206 | 97 | 47 | 1224 |
| | 14 to 15 | 155 | 6.4 (6.2, 6.5) | 754 | 207 | 99 | 47 | 1218 |
| | 15+ | 158 | 6.4 (-) | 747 | 207 | 101 | 47 | 1216 |
| **Males 75+** | 0 to 1 | 239 | 4.4 (4.3, 4.4) | 455 | 200 | 100 | 33 | 803 |
| | 1 to 2 | 184 | 4.4 (4.4, 4.5) | 521 | 196 | 77 | 29 | 785 |
| | 2 to 3 | 206 | 4.2 (4.1, 4.2) | 536 | 181 | 70 | 26 | 718 |
| | 3 to 4 | 218 | 4.0 (4.0, 4.1) | 545 | 170 | 66 | 23 | 670 |
| | 4 to 5 | 231 | 3.9 (3.8, 4.0) | 548 | 160 | 63 | 21 | 634 |
| | 5 to 6 | 259 | 3.8 (3.7, 3.9) | 544 | 152 | 61 | 20 | 608 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 6 to 7 | 258 | 3.8 (3.7, 3.9) | 549 | 148 | 59 | 19 | 603 |
| | 7 to 8 | 256 | 3.7 (3.6, 3.8) | 549 | 144 | 57 | 18 | 598 |
| | 8 to 9 | 263 | 3.7 (3.6, 3.8) | 541 | 138 | 55 | 17 | 599 |
| | 9 to 10 | 279 | 3.7 (3.5, 3.8) | 528 | 133 | 53 | 16 | 608 |
| | 10 to 11 | 289 | 3.7 (3.5, 3.9) | 517 | 130 | 51 | 15 | 633 |
| | 11 to 12 | 283 | 3.8 (3.6, 4.0) | 507 | 132 | 50 | 15 | 672 |
| | 12 to 13 | 264 | 3.8 (3.6, 4.1) | 495 | 132 | 48 | 14 | 713 |
| | 13 to 14 | 273 | 3.9 (3.6, 4.1) | 481 | 127 | 46 | 14 | 740 |
| | 14 to 15 | 255 | 3.9 (3.7, 4.1) | 472 | 125 | 43 | 14 | 774 |
| | 15+ | 256 | 3.9 (-) | 470 | 124 | 40 | 14 | 780 |
| **Females 55-74** | 0 to 1 | 75 | 9.8 (9.6, 9.9) | 1252 | 299 | 109 | 71 | 1831 |
| | 1 to 2 | 53 | 9.5 (9.3, 9.6) | 1303 | 290 | 90 | 63 | 1714 |
| | 2 to 3 | 66 | 9.0 (8.8, 9.1) | 1306 | 274 | 82 | 56 | 1554 |
| | 3 to 4 | 65 | 8.5 (8.4, 8.7) | 1311 | 260 | 76 | 50 | 1421 |
| | 4 to 5 | 77 | 8.1 (7.9, 8.2) | 1299 | 243 | 70 | 45 | 1289 |
| | 5 to 6 | 86 | 7.7 (7.5, 7.8) | 1287 | 228 | 66 | 41 | 1179 |
| | 6 to 7 | 94 | 7.3 (7.2, 7.5) | 1274 | 215 | 61 | 37 | 1083 |
| | 7 to 8 | 104 | 7.0 (6.8, 7.1) | 1255 | 204 | 58 | 34 | 999 |
| | 8 to 9 | 106 | 6.7 (6.5, 6.8) | 1235 | 193 | 54 | 32 | 930 |
| | 9 to 10 | 123 | 6.4 (6.2, 6.5) | 1202 | 183 | 52 | 29 | 865 |
| | 10 to 11 | 127 | 6.2 (6.0, 6.3) | 1176 | 176 | 51 | 27 | 816 |
| | 11 to 12 | 144 | 5.9 (5.8, 6.1) | 1142 | 171 | 50 | 26 | 772 |
| | 12 to 13 | 145 | 5.8 (5.6, 5.9) | 1114 | 170 | 50 | 25 | 746 |
| | 13 to 14 | 164 | 5.6 (5.4, 5.7) | 1079 | 166 | 49 | 24 | 720 |
| | 14 to 15 | 176 | 5.5 (5.4, 5.6) | 1063 | 165 | 48 | 23 | 705 |
| | 15+ | 183 | 5.5 (-) | 1056 | 165 | 48 | 23 | 700 |
| **Females 75+** | 0 to 1 | 171 | 5.2 (5.2, 5.3) | 569 | 238 | 115 | 39 | 953 |
| | 1 to 2 | 143 | 5.1 (5.1, 5.2) | 619 | 226 | 88 | 33 | 904 |
| | 2 to 3 | 157 | 4.8 (4.8, 4.9) | 627 | 207 | 81 | 29 | 822 |
| | 3 to 4 | 175 | 4.6 (4.5, 4.6) | 623 | 192 | 75 | 26 | 755 |
| | 4 to 5 | 191 | 4.4 (4.3, 4.4) | 614 | 178 | 70 | 24 | 704 |
| | 5 to 6 | 205 | 4.2 (4.1, 4.2) | 600 | 167 | 67 | 22 | 665 |
| | 6 to 7 | 210 | 4.0 (3.9, 4.1) | 584 | 158 | 63 | 20 | 636 |
| | 7 to 8 | 231 | 3.8 (3.8, 3.9) | 564 | 148 | 59 | 18 | 607 |
| | 8 to 9 | 233 | 3.7 (3.6, 3.8) | 544 | 139 | 55 | 17 | 594 |
| | 9 to 10 | 265 | 3.5 (3.5, 3.6) | 515 | 129 | 52 | 15 | 581 |
| | 10 to 11 | 272 | 3.5 (3.4, 3.5) | 491 | 123 | 49 | 15 | 587 |
| | 11 to 12 | 281 | 3.4 (3.3, 3.5) | 462 | 120 | 45 | 13 | 599 |
| | 12 to 13 | 298 | 3.3 (3.2, 3.4) | 434 | 116 | 42 | 13 | 615 |
| | 13 to 14 | 300 | 3.3 (3.2, 3.4) | 418 | 110 | 40 | 12 | 636 |
| | 14 to 15 | 299 | 3.3 (3.2, 3.4) | 402 | 107 | 37 | 12 | 658 |
| | 15+ | 301 | 3.3 (-) | 399 | 105 | 34 | 12 | 663 |

[1] At the beginning of the time interval
[2] Average days from the beginning of the interval until death

**Appendix 5-3.** Average number of days of long-term care, home care, inpatient care and outpatient care received by individuals with dementia in Ontario in an average year (2014-2017), by stage of disease




Abbreviations: HC, home care; LTC, long-term care