The Burden of Unhealthy Behaviours: A Lifetime Approach Using Linked Population-Level Health Surveys

Richard Perez

Thesis submitted to 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

© Richard Perez, Ottawa, Canada, 2022

ACKNOWLEDGEMENTS

I would like to first thank my supervisor, Dr. Doug Manuel, who has inspired me to develop innovative approaches to important questions that have historically challenged population health. Over the past years, I have gained valuable insight and perspective from working with Dr. Doug Manuel who routinely combines conventional research with novel applications, such as the development of online health risk calculators. This rather unique experience has provided me with a foundation unlike most graduate students, and has better prepared me for a future where traditional academic research is increasingly translated into tools for the public and healthcare providers. I would also like to thank my other committee members, Dr. Monica Taljaard, Dr. Claudia Sanmartin, and Dr. Kevin Brand for providing valuable feedback on this challenging topic. And lastly, I would like to thank the Director of ICES McMaster, Dr. Hsien Seow, for providing me the flexibility of completing my PhD while working.

THESIS SUMMARY

The purpose of this thesis was to develop an approach that could evaluate the burden of unhealthy behaviours over a lifetime through linked population-based health surveys. The Canadian Community Health Survey (CCHS) is one such cross-sectional survey that is routinely administered to the household population and has been linked to a multitude of administrative healthcare databases. Utilizing the linked CCHS to evaluate the burden of unhealthy behaviours over a lifetime is challenging. Health behaviours naturally change over a lifetime due to many factors, and the burden of unhealthy behaviours has many different dimensions (e.g., mortality, disability, and healthcare costs) that are interconnected with each other. The degree to which lifetime disability and healthcare costs vary in relation to differences in life expectancy remains an area of debate. It is unclear whether individuals with healthy behaviours actually experience less lifetime disability and healthcare costs than individuals with unhealthy behaviours since they typically live much longer. Through several studies, this thesis developed various components that can be potentially combined into a lifetime approach which incorporates multivariable transitions.

The first two studies assessed the burden of unhealthy behaviours on period life expectancy and period lifetime healthcare costs. In the first study, CCHS-based multivariable risk algorithms were constructed to provide estimates of the causal associations between each unhealthy behavior (smoking history, leisure physical inactivity, non-active transport, leisure sedentary activity, and poor diet) and mortality. The burden of unhealthy behaviours on period life expectancy was estimated to be 7.5 (6.5-8.3) life years in 2000-2004 and 6.7 (5.8-7.4) life years in 2010-2014. The largest burdens were attributed to non-active transport and smoking. In the second study, CCHS-based multivariable risk algorithms were constructed to provide estimates of the causal associations between each unhealthy behavior and healthcare costs within different phases of life (i.e., defined by proximity to death). Unhealthy behaviours were attributed with 10.2% (2.5%-17.7%) of the period lifetime healthcare costs in 2000-2004, and 12.9% (5.6%-19.8%) in 2010-2014. Leisure sedentary activity and non-active transport were responsible for almost this entire burden, while the other unhealthy behaviours appeared to actually reduce period lifetime healthcare costs. The degree to which these estimates are accurate

is unclear given the limitations of period life tables and the potential for unhealthy behaviours relating to physical activity to be a product of aging and prior illness.

The third study focused on developing methods by which to derive CCHS-based multivariable transition risk algorithms, which would allow for the creation of cohort life tables rather than period life tables. Novel methods involving multiple imputation models were utilized to create quasi-longitudinal CCHS cohorts from multiple cycles of the CCHS. These quasi-longitudinal cohorts were leveraged to develop multivariable risk algorithms for transitions towards different levels of immobility, an exposure that had been included in the prior algorithms for mortality and healthcare costs. Transitions towards moderate immobility were predicted by all unhealthy behaviours except poor diet, and transitions towards severe immobility were predicted by all unhealthy behaviours except sedentary activity. This approach can also be utilized to develop multivariable transitions for the unhealthy behaviours, which were simultaneously allowed to transition in the quasi-longitudinal CCHS cohorts. Such multivariable transition algorithms could potentially be combined with the previously derived algorithms for mortality and healthcare costs to generate more realistic estimates of life expectancy and lifetime healthcare costs. Large variability in the imputed quasi-longitudinal CCHS cohorts requires further examination, and may be reduced by including comorbidities, healthcare costs, and other information from linked administrative healthcare databases.

The last two studies evaluated the representativeness of linked CCHS respondents for population-based studies. Response and consent (to linkage) rates in the CCHS have been declining since its introduction raising concerns surrounding the comparability of CCHS samples over time. Similar to other population-based surveys, survey weights are provided that are designed address biases that may arise from non-response and non-consent to linkage. Unfortunately, these survey weights are not necessarily appropriate for many linked health outcomes that are rare. As a result, CCHS-based multivariable health risk algorithms are frequently derived from pooled unweighted CCHS samples. Fortunately, relative to wider sampling frames, unweighted linked CCHS samples were observed to be comparable over time. Nevertheless, linked CCHS respondents were observed to be healthier than comparable individuals in the community-dwelling and general populations at older ages, where they demonstrated lower risks of mortality, long-term care admission, and healthcare costs. This was

not unexpected given that important segments of the population (e.g., residents of retirement homes and long-term care care) are excluded from the CCHS sampling frame. These studies highlighted the difficulties of estimating life expectancy and corresponding lifetime healthcare costs from the household population, and the necessity to ensure that such estimates realistically incorporate the time individuals may live outside of the household population over a lifetime.

These series of studies therefore resulted in mortality, healthcare cost, and transition risk algorithms that could potentially be combined to generate lifetime estimates of life expectancy, disability, and healthcare costs for a CCHS respondent. The development of transition risk algorithms requires further research. Once these methods are optimized and transition risk algorithms for all exposures of interest are generated, all the components required for this framework will be complete. At that point, explicit methods by which to combine the algorithms and validate projections will be required. This framework will enable a cause-deleted approach to be applied that simultaneously considers the impact of unhealthy behaviours on mortality, disability, transitions, and healthcare costs. This thesis represents an initial first step towards creating a framework that has the potential to generate lifetime estimates, as well as counterfactual estimates, which better reflect the complex nature of lifetime trajectories.

TABLE OF CONTENTS

1. **Background**
1.1 Debate Surrounding Longer Life Expectancies 1
1.2 Historical Methods 3
1.3 Impact of Unhealthy Behaviours 5
1.4 Advent of Linked National Health Surveys 9
1.5 Period Life Table Approaches 13
1.6 Evolution towards Cohort Life Tables 15
1.7 Development of Transition Models 18
1.8 Objectives 21

2. **Data and Associated Derived Measures**
2.1 Source of Data Holdings 23
2.2 Population Frames 23
2.3 Healthcare Service & Utilization 24
2.4 Comorbidity Measures 25
2.5 Healthcare Costs 26

3. **Burden of Unhealthy Behaviours on Ontario Period Life Expectancy in 2000-2004, 2005-2009, and 2010-2014**
3.1 Abstract 28
3.2 Introduction 29
3.3 Data & Methods 30
3.4 Results 37
3.5 Discussion 38
3.6 Tables 40

4. **Burden of Unhealthy Behaviours on Period Lifetime Healthcare Costs of Ontarians in 2000-2004, 2005-2009, and 2010-2014**
4.1 Abstract 53
4.2 Introduction 54
4.3 Data & Methods 56

4.4 Results 60
4.5 Discussion 61
4.6 Tables 63

5. Assessment of Impact of Unhealthy Behaviours on Transitions towards Immobility using Linked Cross-sectional National Health Surveys

5.1 Abstract 80
5.2 Introduction 81
5.3 Data & Methods 83
5.4 Results 89
5.5 Discussion 91
5.6 Figures 93
5.7 Tables 95

6. Representation of Canadian Community Health Survey respondents for health outcome linkage studies: 2003 to 2014

6.1 Abstract 123
6.2 Introduction 124
6.3 Data & Methods 127
6.4 Results 132
6.5 Discussion 133
6.6 Tables 136

7. Representation of Canadian Community Health Survey respondents for healthcare utilization and cost outcome linkage studies: 2007 to 2014

7.1 Abstract 147
7.2 Introduction 148
7.3 Data & Methods 151
7.4 Results 157
7.5 Discussion 159
7.6 Tables 162

8. Conclusions, Implications, and Future Directions 169

9. References 175

Background

1.1 DEBATE SURROUNDING LONGER LIFE EXPECTANCIES

Life expectancy has dramatically increased around the world over the last 150 years.[1] It was less than 30 years in 1870, but has reached almost 73 years in 2019.[2] Initial gains in life expectancy were primarily the product of declines in mortality from infectious diseases, such as tuberculosis, influenza, and pneumonia.[3,4] These types of disease impacted children to a greater extent, and prevented the majority of the population from reaching old age.[5] The bulk of life years being added were therefore healthy years. These changes provided clear benefits to the individual, who experienced additional years of high-quality life, and society, which gained more working-age years per individual. Reductions in mortality among the non-elderly population has continued up until the present. In 1960, only 43% of males and 53% of females were expected to survive until old age, but by 2016, 73% of males and 82% of females were expected to reach this threshold.[1] In recent decades, gains in life expectancy have increasingly stemmed from declines in mortality among the elderly. Not only are more individuals reaching old age, but they are spending an increasingly larger proportion of their life in old age. The benefit of these additional years is less clear since the elderly typically experience much higher levels of chronic disease.[6] This has raised concerns that living longer may come with additional years of disease, as well as additional diseases.[7] The situation appears even worse when one considers that a considerable amount of disability in the world can be attributed to chronic diseases, such as heart disease, stroke, and chronic obstructive pulmonary disease.[8] If living longer is accompanied by more years of disease and more diseases, it is natural to surmise that this may also include more years of disability, and more severe disability.[9] Given that healthcare costs are closely associated with both disease and disability, these additional years may be simultaneously very costly for society.[10] The amount of disease, disability, and healthcare costs associated with longer life expectancies has naturally been an area of immense interest and discussion.[11]

The debate surrounding the impact of longer life expectancies produced several competing hypotheses early on. The most hopeful hypothesis, known as the 'compression of morbidity', suggests that increases in life expectancies will be accompanied by less years of morbidity.[12] This idea stems from the assumption that life expectancy and morbidity share many

of the same underlying causes, but only life expectancy is constrained by a biological limit. Using this premise, it was reasoned that improvements in health would lead morbidity to become increasingly compressed into shorter periods of time before death. Given that life expectancy continues to increase in many developed nations, it appears unlikely that a biological limit will be reached any time soon.[13] Nevertheless, this does not necessarily preclude a compression of morbidity from occurring if the onset of morbidity is being delayed for a period of time that is greater than the years of life being added. The undesirable counter hypothesis, known as the 'expansion of morbidity', suggests that increases in life expectancies will encompass more years of morbidity.[14,15] This idea stems from the assumption that increases in life expectancy are mainly the result of advancements in healthcare and technology that are enabling individuals with disease and disability to live longer. A highly plausible scenario when one considers that improvements in various treatments have contributed to dramatic decreases in the mortality rates of cancer and cardiovascular disease in the last few decades.[16,17] Drug innovation alone has been attributed with half of the recent gains in life expectancies.[18] The last major hypothesis, known as 'dynamic equilibrium', postulates that the proportions of life associated with different levels of morbidity will remain relatively stable regardless of life expectancy.[19] Its author assumed that longer life expectancies would be characterized by later onsets of morbidity, as well as slower progressions across the different levels of morbidity. This hypothesis positions itself between the optimistic compression of morbidity hypothesis and the pessimistic expansion of morbidity hypothesis. These hypotheses provide a starting framework, but the impact of life expectancy on morbidity is unlikely to be as simple as any of these hypotheses suggest.

The extent to which any of these hypotheses are correct, has important implications for nations and their healthcare systems. In 2016, only 8.5% of the world were elderly, but projections suggest that 17% of the world's population could be elderly by 2050.[20] For developed nations, the impact will be felt even sooner. In Canada, the elderly are expected to represent approximately a quarter of the population by 2050.[21] The working-age population of each nation will in turn become an increasingly smaller proportion of the overall population. This shrinking working-age population will need to support government programs that will grow in cost to accommodate the needs of the growing elderly population. One of the largest drivers of government expenditures that is associated with the elderly is healthcare. Among OECD nations, healthcare expenditures currently represent 8.8% of GDP, a significant portion of these

economies.[22] In 2018, healthcare spending already amounted to 16.9% of the GDP in the United States, and 10.7% of the GDP in Canada.[23] Relative to their proportion of the overall population, the elderly receive a much larger proportion of healthcare than the non-elderly. For example, the elderly represent approximately 16% of the American and Canadian populations, but are attributed with 36% of healthcare spending in the United States and 46% of healthcare spending in Canada.[23,24] There is concern that healthcare systems will be overburdened by rapidly aging populations, and require substantial increases in funding to maintain their current level of care.[25] It is also widely believed that current healthcare systems which focus principally on acute hospital-based care are ill-equipped to service aging populations that will likely require more homecare services and long-term care facilities.[26,27] Understanding the amount of morbidity associated with longer life expectancies, and the corresponding impact on healthcare systems would improve the ability of governments to anticipate the future healthcare needs of their aging populations.

1.2 HISTORICAL METHODS

Researchers have primarily relied on publicly available mortality statistics to examine life expectancies over time. Many nations have reported national and regional age-specific mortality rates on an annual basis for decades. These rates can be converted into corresponding age-specific probabilities of death. Government agencies and public health organizations routinely utilize this information to construct period life tables from which they can derive period life expectancies. This methodology typically involves applying the age-specific probabilities of death to a hypothetical cohort of 100,000 individuals. As a result, we are provided with the number of individuals alive at the beginning of each age interval, and the number of deaths that occurs within each age interval. The total person-years lived at each age interval can be calculated from the number of individuals that survive an age interval and the number of deaths in the age interval. Each survivor contributes a number of person-years equal in length to the age interval, and each decedent contributes a portion of time that is less than the age interval. Dividing the total person-years across all ages by the original size of the hypothetical cohort provides a period life expectancy. This metric is very appealing given the ease with which it can be calculated from available mortality statistics. However, it should be noted that this is actually not a realistic estimate of life expectancy since the age-specific mortality rates are being selected

from many different birth cohorts. In reality, birth cohorts experience different age-specific mortality rates due to differences in their environment and personal characteristics. For example, technological advancements in healthcare suggests that younger generations will experience better healthcare in their old age than older generations do presently. Alternatively, rising trends in obesity suggests that younger generations will experience more obesity-related diseases in their old age than older generations do presently. Despite these drawbacks, period life expectancies are still the predominant form of reported life expectancy.

The widespread use of period life tables can be partially attributed to the development of health expectancy metrics that use the period life table as their foundation. In order to qualify the quality-of-life years associated with a life expectancy, researchers have developed metrics that combine the mortality information from these tables with indicators of morbidity. Unlike readily available mortality statistics, indicators of morbidity relating to disease or disability have been generally difficult to obtain. The use of administrative healthcare databases to measure disease and disability across populations has been a recent development that is still rather challenging and limited in many jurisdictions.[28,29] Historically, national and regional surveys have been the primary source of population-based morbidity measures for researchers. Many of these cross-sectional surveys contain health-related modules that can be used to estimate age-specific rates of disease and disability. Almost a half-century ago, Sullivan developed a method that could combine these survey-based morbidity measures with corresponding period life tables.[30] This methodology divides the life years lived at each age in the period life table into healthy and unhealthy proportions based on the age-specific prevalence of morbidity derived from the cross-sectional survey. Healthy and unhealthy life expectancies are then derived from the total healthy and unhealthy person-years. The healthy and unhealthy life expectancies sum to the overall period life expectancy. It was typical of the first studies exploring the impact of life expectancy on morbidity to apply this methodology to two or more surveys from different time periods.[31–34] If the increase in life expectancy was primarily a product of increases in healthy life expectancy, researchers would conclude that morbidity was being compressed. Alternatively, if the increase in life expectancy was primarily a product of increases in unhealthy life expectancy, researchers would conclude that morbidity was being expanded. This approach provided a rather simplified examination of a complicated question.

Studies that have used this methodology to evaluate the impact of life expectancy on morbidity have produced mixed results. Researchers struggled initially to find appropriate measures of morbidity. National and regional surveys were not as common as they are today, nor was it common for the same survey to be administered at different points in time. As a result, it was not unusual for researchers to examine changes in life expectancy and morbidity between two time points using different surveys. This meant studies lacked internal consistency, and comparisons over time were performed using surveys that often differed in their measures of morbidity, methodology, or target population. As national and regional surveys have become more common, researchers have had the opportunity to use iterations of the same survey at different time points.[35,36] Improvements in the quality of the studies, however, have not made the picture any clearer. Some nations have demonstrated compression of morbidity, others have suggested expansion of morbidity, and a few have aligned most closely with dynamic equilibrium.[37–42] Differences between studies make it difficult to draw conclusions. For example, some studies measure morbidity by number of chronic diseases, and others use different forms of disability. In addition, not all studies include institutionalized individuals, ignoring the most diseased and disabled segment of the general population. Nonetheless, the research has highlighted that results can drastically differ within the same populations, as substantial variation has been observed across different socioeconomic groups (e.g. gender, ethnicity, and education).[43] Whether individuals experience compression or expansion of morbidity may depend on their underlying risk factors. Analysis at the population level, which overlooks the heterogeneity of risk factors at the individual level, could be producing misleading results.

1.3 IMPACT OF UNHEALTHY BEHAVIOURS

Unhealthy behaviours are modifiable risk factors that could help explain whether the compression or expansion of morbidity occurs in different populations. It is widely known that unhealthy behaviours are important drivers of mortality. Some of the most frequently studied unhealthy behaviours include smoking, leisure physical inactivity, sedentary activity, and poor diet. Historically, smoking has been the unhealthy behaviour with the largest recognized impact on mortality. Despite dramatic decreases in the prevalence of smoking, it remains the leading cause of premature death around the world.[44,45] Smokers have been observed to have mortality

rates over two times higher than non-smokers.[46–48] Several studies have estimated that smokers live at least 10 years less than non-smokers.[47,49] Physical inactivity and sedentary activity are unhealthy behaviours that have been receiving ever-increasing attention because of their worrisome trajectory. They have been on the rise for decades, and projections suggest the situation will only get worse as societies continue to become less physically active and more sedentary.[50,51] Physically inactive individuals have been observed to have mortality rates approximately twice that of physically active individuals.[52] A review of studies examining the impact of physical inactivity on life expectancy estimated that a physically inactive individual has a life expectancy almost 3.5 years shorter than a physically active individual.[53] Sedentary activity has been observed to have a similar impact on mortality and life expectancy.[54,55] Among the aforementioned unhealthy behaviours, poor diet is the most difficult to understand. Dietary patterns are complex, as low consumption of particular foods are often associated with high consumption of other types of food. Despite these intricacies, research has in general been able to demonstrate that various dietary components (e.g. low fruit and vegetable consumption, high sugar consumption, and high meat consumption), are associated with poor health outcomes, and reduced life expectancies.[56–58] Various studies have demonstrated that having multiple unhealthy behaviours has a cumulative detrimental impact on life expectancy.[59–61] A substantial amount of the variation observed in life expectancy can therefore be explained by differences in health behaviours.

Unhealthy behaviours are also key contributors to many of the diseases that are commonly associated with ageing. Cardiovascular diseases account for the greatest number of deaths in the world.[62] Case-control studies involving dozens of nations have demonstrated unhealthy behaviours to be responsible for the majority of cardiovascular disease.[63,64] Over the last few decades, mortality from cardiovascular disease has been declining dramatically across many developed nations.[17] A significant portion of this improvement can be attributed to the success of public health programs in dramatically reducing the prevalence of smoking.[65,66] The risk for cardiovascular disease is at least 50% higher among smokers than non-smokers.[67] This elevated risk begins to rapidly dissipate for smokers once they quit, until eventually they have the same risk as non-smokers.[68] In contrast, diabetes is a disease that has been steadily increasing over the last few decades as a result of declining levels of physical activity, increases in sedentary activities, and worsening diets.[50,69–71] Prospective cohort studies have demonstrated

that individuals who exhibit these unhealthy behaviours develop diabetes at a higher rate, while improvements in these health behaviours can also reduce the risk of diabetes.[72–74] In the last few decades, unhealthy behaviours have contributed to a doubling of obesity and a worldwide diabetes epidemic.[75,76] Even diseases like cancer that are typically viewed as a product of ageing and genetics can be dramatically influenced by unhealthy behaviours. The majority of cancers occur in the elderly, which is not surprising since cancer involves a deterioration of cellular processes that coincides with ageing. Nevertheless, individuals with unhealthy behaviours have still been observed to develop cancer at a rate at least 50% higher than those with healthy behaviours.[77,78] Smoking is the predominant unhealthy behaviour associated with cancer, and viewed as the main cause of many cancers.[79] The distribution of unhealthy behaviours within a population therefore plays an integral role in the morbidity experienced by a population.

From a quality-of-life perspective, the extent of impairments and disabilities an individual experiences are more important than the number and combination of chronic diseases. As individuals age, they accumulate health problems from acute and chronic conditions.[80,81] This leads to the development of impairments, such as the loss of hearing, dexterity, or memory. An individual is typically regarded as functionally disabled once impairments prevent them from taking care of themselves (e.g., bathing and self-feeding) or from performing tasks necessary to live independently in the community (e.g., housework and preparing meals). Functional disability is viewed as a severe form of disability; a strong predictor of institutionalization, and more common in the last years of life.[82,83] Elderly individuals with chronic diseases have been observed to have substantially higher rates of functional disability.[80,84,85] It is not necessarily the higher levels of chronic disease that are projected with aging populations which has raised concerns, but rather the higher levels of severe disability that are expected to accompany the chronic diseases.[86] In the United States and Canada, at least a third of the elderly report to have at least one disability, while not even a tenth of the non-elderly report to have a disability.[87,88] Even the most optimistic projections would suggest that disability will dramatically increase because of aging populations. Nevertheless, research has demonstrated that disability does not necessarily always increase at rates that would be expected given the increases observed in chronic diseases.[89] In some cases, disability among the elderly has actually decreased.[89] Without accounting for behavioural and socioeconomic factors, chronic diseases may be a poor proxy for disability. Unhealthy behaviours have been shown to predict earlier onsets of functional

disability, and also demonstrate a dose-response relationship where the likelihood of onset increases with each added unhealthy behaviour.[90,91] Evaluating disability with respect to unhealthy behaviours may provide a better understanding of population trends in disability.

For governments that manage healthcare systems, the utilization and cost associated with an unhealthy behaviour are key pieces of information that can inform future policy and the development of interventions. Linkage of cohorts and health surveys to administrative healthcare databases has enabled researchers to directly estimate healthcare utilization and costs associated with specific unhealthy behaviours. The results are not surprising; individuals with unhealthy behaviours typically use healthcare services at a much higher rate than those with healthy behaviours. The earlier onset of disease and disability among those with unhealthy behaviours naturally leads to higher healthcare utilization, and a more rapid accumulation of healthcare costs. Smokers and former smokers have demonstrated higher healthcare utilization and costs than non-smokers across numerous types of care such as hospitalizations, emergency services, and home care.[92–96] Smoking bans and cessation programs have been observed to lead to lower hospitalization rates for smoking-related diseases.[97,98] Studies examining hospital use and costs have demonstrated dose-response relationships corresponding to smoking intensity, and the recency of the smoking behaviour.[99,100] Similarly, physically inactive individuals have also been observed to have higher utilization than physically active individuals across numerous types of care such as hospitalizations, emergency services, home care, and prescription drugs.[101–106] Several studies have demonstrated that physical activity interventions can reduce healthcare costs.[107,108] The evidence for diet is more sparse.[109,110] Low fruit and vegetable consumption, which is the most common measure of poor diet quality, has demonstrated an association with higher healthcare utilization and costs.[99,111] Actual trials on the impact of diet interventions on healthcare utilization and costs are currently severely lacking.[112] The combined impact of multiple unhealthy behaviours can have a very serious impact on healthcare systems. In Ontario, 22% of healthcare costs have been attributed to several unhealthy behaviours.[111] This suggests that a substantial amount of the variation in healthcare costs can be explained by differences in health behaviours.

These cross-sectional or prevalence-based approaches of evaluating the burden of unhealthy behaviours on disease, disability, and healthcare costs have come under criticism

because they overlook differences in the life expectancies between individuals with healthy and unhealthy behaviours.[113,114] Similar to the debate around whether longer life expectancies incur more morbidity and healthcare costs, there is also a debate around whether individuals with healthy behaviours actually experience more morbidity and accumulate more healthcare costs because of their longer life expectancies. Unfortunately, long-term longitudinal studies that follow individuals until death and simultaneously capture measures of health behaviours, disease, disability, and healthcare costs are basically non-existent. Longitudinal studies that do exist are often limited in sample size, have insufficient follow-up, or are not generalizable. Many researchers attempting to understand whether healthier behaviours lead to a compression or expansion of morbidity have relied on period life table approaches. Separate period life tables are habitually constructed for the healthy and unhealthy behavioural groups of interest, and each table is utilized to generate a corresponding life expectancy. Age-specific prevalence measures of disease, disability or healthcare costs for each group of interest are then superimposed on their corresponding lifetable to create corresponding lifetime measures. Given the lack of appropriate data, it is not uncommon for researchers to take behavioural exposures, socioeconomic factors, measures of disability, healthcare costs, and mortality statistics from different sources.[115,116] This makes it difficult to decompose the impact of different unhealthy behaviours that often occur in tandem, and more frequently within low socioeconomic groups.[117–120] Studies that have attempted to incorporate differences in life expectancies when evaluating the lifetime burden of smoking on disability and healthcare costs have produced mixed results.[113,114,121,122]

1.4 ADVENT OF LINKED NATIONAL HEALTH SURVEYS

National health surveys have been our main source of health behaviour surveillance. The Canadian Community Health Survey (CCHS) is a population-based survey routinely administered by Statistics Canada.[123] Similar to other population health surveys worldwide, the CCHS was originally designed to provide cross-sectional measures of exposures (e.g. lifestyle and sociodemographic information) and health status not generally captured within the healthcare system, nor readily available through other sources. The linking of these surveys to administrative healthcare databases has enhanced their utility beyond this original purpose. These databases provide access to longitudinal outcomes and other complementary content that can enrich the information already captured through the surveys. This provides us with

opportunities to directly evaluate the impact of unhealthy behaviours on disease, disability, and mortality simultaneously. This is a significant advancement for researchers striving to gain a comprehensive understanding of how unhealthy behaviours effect individuals over a lifetime. Research stemming from analysis of these linked surveys could provide insight into the conditions under which contraction or expansion of morbidity occurs. The extent of linkage has also reached a point in certain jurisdictions where the cost of almost all the different types of publicly funded healthcare associated with a survey respondent can be directly estimated. This is a major development for researchers that traditionally have had to piece together healthcare costs from sources that were not necessarily consistent with the sources providing the measures of disease, disability, and mortality. This provides opportunities to more accurately assess the economic impact of unhealthy behaviours, and potentially determine whether individuals with healthy behaviours actually do have higher lifetime healthcare costs than individuals with unhealthy behaviours.

Deriving lifetime estimates from the linked CCHS remains no simple task despite its advantages over traditional data sources. For starters, its target population is not the general population. The CCHS is designed to only provide representative measures of the household population, which is limited to individuals living within private households. The survey employs a complex multistage sampling strategy that results in a weight being assigned to each one of its participants. This weight indicates the number of individuals in the household population that is represented by the participant, after adjusting for unequal selection probabilities and non-response. Estimates for the household population are derived by incorporating these weights in analysis of the CCHS. It is generally assumed that estimates from the CCHS are healthier than estimates that would be derived from the general population because the household population does not include individuals residing in institutions. Although there is only a small proportion of the population living in the institutional setting, they are predominantly elderly long-term care residents who have poor health and high healthcare use. The linkage of the CCHS and population-based registries to administrative healthcare databases has provided measures common to both CCHS respondents and the general population. Comparisons of the linked CCHS and the general population can help inform research aimed at deriving lifetime measures from the linked CCHS.

Excluding long-term care residents and other institutionalized populations from the general population results in the community-dwelling population. This population comprises of individuals living in private households, the target population of the CCHS, but also individuals living in collective dwellings such as retirement homes. Unlike long-term care facilities, retirement homes do not provide round-the-clock health monitoring by health professionals. It is less clear whether estimates derived from the CCHS are healthier than estimates that would be derived from the community-dwelling population. Individuals in retirement homes may not be significantly more ill than their counterparts in the household population. Linkages to administrative healthcare databases has made it possible to identify and exclude long-term care residents from population-based registries. Comparisons of the linked CCHS and the community-dwelling population could better our understanding of whether expected health differences between the CCHS and the general population is solely the product of long-term care residents. Given current limitations in identifying residents of retirement homes within population-based registries, the community-dwelling population is also the closest approximation of the household population that is available. If no significant health differences are observed between the linked CCHS and community-dwelling population, it could be inferred that the community-dwelling population is a reasonable approximation of the household population, and that the CCHS is a reasonable approximation of both populations. If this not the case, deriving lifetime estimates from the linked CCHS becomes even more challenging since the absence of individuals in retirement homes and long-term care then both need to be addressed.

Health differences between the CCHS and either the community-dwelling and general population may not be constant over time. If long-term care is not sufficiently expanding to meet the demand of an aging population, ill and frail seniors will have to increasingly remain in the community-dwelling population. In this scenario, health differences between the CCHS and the general population would decrease over time, since the community-dwelling population would increasingly resemble the general population. At the same time, national health surveys like CCHS are experiencing notable declines in response rates. From 2000 to 2013, response rates to the CCHS dropped from 88% to 67%.[124] This has raised concerns that the level of non-response has reached critical levels where weighting procedures are no longer an adequate method to adjust for differences between respondents and non-respondents. For example, respondents of health surveys can be healthier than non-respondents, what is commonly known as healthy

response bias. These concerns are compounded with the linked CCHS. Linked CCHS respondents represent a subset of respondents since not all respondents also consent to link their survey information. For example, the CCHS cycle administered in 2011 had a response rate of 70%, but only 58% actually responded and consented to linkage. [125] Research has demonstrated that consenting respondents can also be healthier than non-consenting respondents.[126] It is plausible that the linked CCHS is becoming increasingly healthier than its target population because of increases in survey non-response and non-consent to linkage of the survey information. The nature and extent to which health differences exist between the linked CCHS and each population (i.e., community-dwelling, and general populations) may therefore be complex since the linked CCHS, community-dwelling, and general populations could all be evolving over time. Generating accurate lifetime measures from the linked CCHS becomes more difficult if these differences are not stable over time.

Novel areas of research utilizing the linked CCHS do not always incorporate the survey weights. Many of the longitudinal outcomes of interest being linked to the CCHS are extremely rare. This creates challenges for researchers utilizing the linked CCHS in novel applications, such as the development of health risk algorithms. Combining multiple cycles of the CCHS has been a common approach to address sample size issues. However, as risk algorithms have turned to more rare linked longitudinal outcomes where weights can have a dramatic impact, researchers have begun developing risk algorithms from combined cycles of the CCHS without use of the weights. One of the purposes of these survey weights is to adjust for non-response. Any non-response bias may be minimal or non-existent among the CCHS respondents after applying the weights. By not incorporating the weights, however, novel areas of research may be magnifying any non-response bias that does exist among the unweighted CCHS respondents. Generating lifetime measures of burden for unhealthy behaviours from the linked CCHS is likely to involve a multitude of unweighted risk algorithms. For researchers leveraging the linked CCHS to create such lifetime measures, it is essential to understand the differences between unweighted linked CCHS respondents and their counterparts in the community-dwelling and general population.

1.5 PERIOD LIFE TABLE APPROACHES

The obvious challenges of utilizing the linked CCHS to generate lifetime measures of burden for unhealthy behaviours has not deterred researchers. One interesting innovation is the integration of CCHS-based mortality risk algorithms with the period life table approach. This methodology is designed to estimate the impact of unhealthy behaviours on period life expectancy. A multivariable risk algorithm for mortality can be developed from a CCHS cohort consisting of multiple cycles of the linked CCHS. Including multiple behaviours and various socioeconomic factors in the algorithm leads to a well-adjusted coefficient for each exposure included in the algorithm. These well-adjusted coefficients can be interpreted as estimates of the causal relationships between the exposures and the outcome. Indicators of disability or diseases are not generally included in the algorithm since they lie along the causal pathway between the exposures of interest and death. Applying this mortality risk algorithm to different CCHS cycles generates baseline period life expectancies for the population at different points in time. A cause-deleted risk approach can be employed to create counterfactual period life expectancies where a particular unhealthy behaviour has been theoretically removed. This is accomplished by using cause-deleted CCHS cycles where the behavioural exposure of interest has been recoded to its 'healthy' level for all respondents. Applying the algorithm to these cause-deleted CCHS cycles provides corresponding cause-deleted period life expectancies. The difference between the baseline period life expectancy of a CCHS cycle and its cause-deleted period life expectancy provides an estimate of the mortality burden associated with the unhealthy behaviour in question. Comparison of burdens across multiple CCHS cycles offers an opportunity to evaluate the extent to which the mortality burden of an unhealthy behaviour is changing over time. This scheme can generate estimates for each unhealthy behaviour that is included in the algorithm. The burden of all the unhealthy behaviours combined can be generated by using cause-deleted CCHS cycles where all unhealthy behaviours are simultaneously coded to their 'healthy' level. Several recent publications have used variations of this methodology.[45,47]

The cause-deleted period life table approach can be extended to estimate the impact of unhealthy behaviours on disability. Disability is a construct that is difficult to assess using administrative healthcare databases. Standard measures of disability are generally not captured within administrative healthcare databases for the entire population. The CCHS has various

questions that can be utilized to identify the extent to which a respondent is disabled. Multivariable risk algorithms can be developed for these indicators of disability using the same linked CCHS cycles and exposures utilized in the development of the mortality risk algorithm. The disability risk algorithms can then be applied to the respondents of different cycles of the CCHS to estimate baseline age-specific prevalence measures of disability for the population at different points in time. Using the Sullivan method, the baseline period life table (i.e., generated by mortality risk algorithm) of a CCHS cycle can be combined with its corresponding baseline age-specific prevalence measures of disability. This results in a baseline estimate of life years lived at each level of disability for each CCHS cycle. Applying the disability risk algorithms to the cause-deleted CCHS cycles of a particular unhealthy behaviour provides corresponding cause-deleted age-specific prevalence measures of disability where the unhealthy behaviour theoretically does not exist. The Sullivan method can then also be used to combine the unhealthy behaviour's cause-deleted period life table (i.e., generated by mortality risk algorithm) of a CCHS cycle with its corresponding cause-deleted age-specific prevalence measures of disability. This results in a cause-deleted estimate of life years lived at each level of disability for each CCHS cycle. The difference between a CCHS cycle's baseline estimate of disability life years and its cause-deleted estimate of disability life years provides an estimate of the disability burden associated with the unhealthy behaviour. Comparison of disability burdens across multiple CCHS cycles offers an opportunity to evaluate the extent to which the disability burden of an unhealthy behaviour is changing over time. This methodology decomposes the impact of behaviours and socioeconomic factors on both mortality and disability.

The cause-deleted period life table approach can also be utilized to estimate the impact of unhealthy behaviours on healthcare costs. Unlike disability, healthcare costs can be accurately estimated from linked healthcare administrative databases rather than self-reported measures in CCHS. In Ontario, linked CCHS respondents at ICES can be individually linked to all records of publicly funded healthcare use, and methodologies have been developed to estimate the person-level costs associated with each type of healthcare use.[127] Multivariable risk algorithms can be developed for healthcare costs using the same linked CCHS cycles and exposures utilized in the development of the mortality risk algorithm. The cost risk algorithms can then be applied to the respondents of different cycles of the CCHS to estimate baseline age-specific healthcare costs at different points in time. Using the Sullivan method, the baseline period life table (i.e., generated

by mortality risk algorithm) of a CCHS cycle can be combined with its corresponding baseline age-specific healthcare costs. This results in a baseline estimate of period lifetime healthcare costs for each CCHS cycle. Applying the cost risk algorithms to the cause-deleted CCHS cycles of a particular unhealthy behaviour provides corresponding cause-deleted age-specific healthcare costs where the unhealthy behaviour theoretically does not exist. The Sullivan method can then also be used to combine the unhealthy behaviour's cause-deleted period life table (i.e., generated by mortality risk algorithm) of a CCHS cycle with its corresponding cause-deleted age-specific healthcare costs. This results in a cause-deleted estimate of period lifetime healthcare costs for each CCHS cycle. The difference between a CCHS cycle's baseline estimate of period lifetime healthcare costs and its cause-deleted estimate of period lifetime healthcare costs provides an estimate of the healthcare cost burden associated with the unhealthy behaviour. Comparison of healthcare cost burdens across multiple CCHS cycles offers an opportunity to evaluate the extent to which the healthcare cost burden of an unhealthy behaviour is changing over time. This methodology incorporates the differential impact of behaviours and socioeconomic factors on both mortality and healthcare costs.

1.6 EVOLUTION TOWARDS COHORT LIFE TABLES

These cause-deleted risk approaches represent an important advancement in the science but they still have several limitations. Founded on a period life table framework, these approaches implicitly assume that younger generations will have the same behavioural and socioeconomic experience of older generations once they reach those ages. Individuals from different birth cohorts, however, are unlikely to experience the same behaviour trajectories across their lifetimes. For example, the prevalence of smoking among younger generations is much lower than it has been historically. As a result, the proportion of elderly in the future who have any history of smoking will likely be much lower than what we observe today. In addition, younger generations of smokers have higher levels of education than older generations of smokers have had historically, suggesting that they will have a higher likelihood of quitting. As a result, the ratio of former smokers to current smokers among the elderly will likely be much higher than what we observe today. It is unrealistic to assume that the trajectory of behaviors will be the same across different generations. Unfortunately, cross-sectional health surveys such as the CCHS are not designed to inform causality around behaviour change. For example, a

respondent may report to be physically inactive because they have already entered functional decline and become disabled. At the same time, respondents who report to be disabled are more likely to have been physically inactive in the past. It is unclear whether physical inactivity is a cause or product of disability in this scenario. Cross-sectional surveys only provide snapshots in time without accounting for temporal relationships between risk factors and health states. This raises questions around the risk estimates utilized in the multivariable algorithms, and whether they can actually be interpreted as estimates of causal relationships. This also calls into question the plausibility of some cause-deleted scenarios where an unhealthy behaviour theoretically does not exist. Many individuals who are observed to be physically inactive and sedentary late in life, would have been physically active and non-sedentary earlier in life. Some portion of these changes in behaviour are a product of aging, and should not be included as part of the burden estimate for unhealthy behaviours. Incorporating behaviour change would address many of the drawbacks associated with the period life table framework.

Researchers have attempted to solve this issue by developing transition models for various exposures and health states of interest. This is accomplished by leveraging differences between the repeated cross-sectional national health surveys. For example, if surveys are administered every two years, individuals who were X years of age in one survey could be assumed to be represented by individuals X+2 years of age in the subsequent survey. Age-specific transition probabilities for a particular exposure, behaviour, or health state can then be derived by comparing prevalence estimates between successive surveys. Fundamental to these methods is the assumption that the population is a closed cohort, which implies that the net migration between surveys is zero. Otherwise, any differences in prevalence between surveys could be the product of individuals entering or leaving the population. This can be a problematic assumption since national health surveys only target the household population. As a result, the most debilitated segments of the population are excluded; individuals in retirement homes, long-term care, and other institutional care. As the health of individuals decline in old age, the most debilitated are likely to transition out of the household population and not be eligible to participate in subsequent surveys. Furthermore, potential changes in the health of the target household population over time because of insufficient increases in the capacity of long-term care facilities, coupled with concerns over decreasing response rates to national health surveys brings into the question the comparability of respondents from one survey to the next. On the

surface, it would appear that deriving transitions through simple comparisons of successive surveys may be problematic and crude.

The specificity of transitions between successive surveys can sometimes be improved by leveraging longitudinal information that is captured within the cross-sectional surveys. Although respondents in cross-sectional surveys are not followed over time, they are occasionally asked time-dependent content that can inform us of past behaviours and health states. For example, the CCHS inquires whether current non-smokers smoked in the past, and whether current non-drinkers drank in the past. In some cases, respondents are even asked about the duration of time that they have displayed a behaviour or experienced a health state. For example, former smokers in the CCHS are asked about the length of time since they have quit smoking. Using this information, it can be inferred whether the individuals would have been a current smoker or former smoker at the time of a previously administered iteration of the cross-sectional survey. Unfortunately, time-dependent content is not habitually captured for every behavior or health state of interest, and in the few occasions it is captured, the degree of specificity provided is often insufficient to determine an individual's exact behavior or health state at the time of a previously survey administration. Furthermore, surveys obviously cannot capture information indicative of future transitions, since they have yet to occur. This still leaves us with numerous potential transitions to evaluate.

The number of potential transitions can be simplified by focusing only on net transitions (i.e., allowing transitions only in one direction). Although in many cases, individuals may transition towards or away from unhealthy behaviors over time, it is the net transitions that are of particular interest. The net transitions allow for projections from one survey to the next, and potentially over a lifetime. In examining only net transitions, the number of potential transitions drastically decreases. For example, the proportion of the population that are physically inactive increases with age. In this case, only transitions that move individuals towards ever-increasing levels of physical inactivity are necessary to project population changes in physical activity between successive cross-sectional surveys. In other cases, the potential transitions are limited in direction by irreversibility. For example, respondents who report to be physically limited due to a chronic condition can only maintain their present level of debilitation or worsen. For the most part, it is not realistic for the respondent to remedy their chronic condition and become fully

healthy again. This also suggests that this same respondent in the past was healthier or at worst, at the same level of debilitation that they are presently reporting. Taking these details into account, respondents can only project into certain levels of health behaviours or health states going forward or backward in time.

The linkage of national health surveys to heath administrative data provides further opportunities to enrich our knowledge of respondents between successive surveys. For example, linkages can identify respondents who are not eligible for successive surveys for reasons such as death or emigration. This allows for transitions that are derived from differences in prevalence between successive surveys to no longer depend on a closed cohort assumption that is unlikely to be true. Linkages can also identify the duration of time a respondent has experienced a disease or the duration of time until a respondent develops a disease. This enables us to evaluate whether a respondent has a particular disease at the time of their current survey, a survey administered 'X' years earlier, or 'X' years later. This creates a longitudinal framework that is common to respondents across different survey administrations. Respondents who present with a particular disease (e.g., diabetes) at the time of their survey should therefore project onto respondents of a successive survey (i.e., 2 years later) who also had presented with the same disease at least two years earlier (i.e., at the time of the earlier survey). Similarly, linkages to administrative healthcare databases can provides us with patterns of healthcare utilization and costs in the intervals between successive surveys that can also improve specification of transitions. This provides overlapping information between respondents to different survey administrations, and therefore a more direct form of connecting similar respondents between successive surveys. The culmination of the basic assumptions aforementioned, and the linkage to administrative healthcare databases turns the challenge of estimating transitions into a missing data problem.

1.7 DEVELOPMENT OF TRANSITION MODELS

Multiple imputation can be leveraged to transform two separate survey administrations into a quasi-longitudinal cohort where respondents have two sets of values rather than the one pertinent only to their immediate survey administration. In this scenario, the two survey administrations encapsulate an observation period with the earlier survey administration (i.e., base cohort) providing pre-survey responses at the beginning of the observation period, and the later survey administration (i.e., target cohort) providing post-survey responses at the end of the

observation period. This implies that we need to impute post-survey values for the base cohort and pre-survey values for the target cohort. If we assume that transitions only occur in a net direction, and take into account that many health states are not reversible, several of the pre-survey and post-survey values can be deduced. The remaining missing values can be ascertained through imputation models. For each post-survey response of interest, we can estimate values through imputation models using the pre-survey responses, and additional linked information such as diseases, healthcare utilization, and costs. Similarly, for each pre-survey response of interest, we can estimate values through imputation models using the post-survey responses of interest, and the same linked information. These imputation models result in a longitudinal cohort where both the base and target cohorts have pre-survey and post-survey response values. Given the inclusion of linked information (diseases, healthcare utilization, and costs etc.), the imputation models can also potentially produce imputed values that are reflective of fundamental differences between the base and target cohorts. For example, the target cohort at the time of their survey administration (post-survey relative to observation window) are residents of the household population, whereas the base cohort is likely to also include individuals who have transitioned into retirement homes and long-term care by the time they reach the end of the observation window. Once the quasi-longitudinal cohort is created, multivariable transition algorithms can be derived for each of the exposures that were transitioned across the observation window. Similar to how multiple CCHS cycles are pooled together into a larger cohort for the development of the mortality, disability, and cost risk algorithms, several quasi-longitudinal CCHS cohorts can be derived and pooled together for the development of the multivariable transition algorithms.

The period life table framework is no longer necessary once multivariable transition algorithms are available for each exposure of interest. Each multivariable transition algorithm provides age-specific estimates of an exposure's distribution based on the distribution of behaviour and socioeconomic indicators at a prior age. For example, a multivariable transition algorithm that explains transitions from current smoking to former smoking would provide the probability of becoming a former smoker over an age interval given the distribution of behaviours and socioeconomic indicators at the beginning of the age interval. Applying all these multivariable transition algorithms simultaneously to an age group's initial set of behaviour and socioeconomic indicators creates a new distribution of behaviour and socioeconomic factors that

represents their projected distribution at a later age. Rather than combining different birth cohorts into a period life table for each CCHS cycle, a separate life table can be constructed for each age cohort in a CCHS cycle. In this scenario, the behaviour and socioeconomic factors are projected forward through the life table using the age group's initial set of behaviour and socioeconomic factors. The multivariable transition algorithms are iteratively applied on each age interval's distribution of behaviours and socioeconomic indicators to produce their distribution for the next age interval in the life table. The mortality, disability, and cost risk algorithms can be applied to these life tables in the same manner that they were applied to the period life tables. However, unlike the estimates based on the period life table framework, this methodology would produce specific estimates for each age cohort of a CCHS cycle. As a result, the resulting burdens of unhealthy behaviours on life expectancy, lifetime disability, and lifetime healthcare costs would be more realistic of actual lifetimes. In addition, the requirement that the algorithms represent causal relationships between exposures and outcomes can be relaxed with this methodology. If multivariable transitions models are developed for transitions across different levels of disability, it is no longer necessary to preclude these levels of disability from inclusion in the mortality and cost algorithms since the transition algorithms will project transitions into these health states based on behaviours and socioeconomic indicators. Treating the levels of disability as health states through which a CCHS respondent transitions is also a more desirable preposition. Unlike mortality and healthcare costs which are based on linked longitudinal outcomes, levels of disability are self-reported measures from the CCHS which means the disability algorithm is not based on predicted risk. Inclusion of the disability levels in the mortality and cost algorithms would also improve their ability to discriminate risk between respondents since mortality and costs are more closely tied to health states.

This methodology can be taken one step further. Instead of creating life tables for different age cohorts, life tables could be generated for individual CCHS respondents. Given that we would have multivariable algorithms that would decompose the impact of behaviours and sociodemographic indicators on behaviour change, mortality, disability, and healthcare costs, there is no reason why this methodology cannot be applied at the individual level rather than on different age cohorts. At this point, the overall methodology can be evaluated in a similar manner to multivariable predictive algorithms. Typically, predictive algorithms are derived using a cohort with a limited observation window (e.g., 5 years), and then validated on another cohort

with the same limited observation window. In order to validate this methodology, outcomes should be evaluated in an observation window beyond the limited observation window that was used for the derivation of the algorithms. For example, if predicted deaths in future years (e.g., beyond the 5 years observation window used for derivation of algorithms) are well calibrated to observed deaths for various subgroups, it would suggest that the combined algorithms can accurately project the risk of mortality into the future. Several of the CCHS cycles now have over a decade of follow-up. Demonstrating accurate projections of mortality, disability, and healthcare costs for CCHS respondents of different ages, and from different periods (i.e., different CCHS cycles), would suggest that derived lifetime estimates from this methodology could be interpreted as realistic approximations of reality. The lifetime estimates and their counterfactual estimates could then be used to determine whether individuals with healthy behaviours actually do or do not have more disability and healthcare costs than individuals with unhealthy behaviours. As a corollary, it would also be possible to examine the more general question of whether increases in life expectancy over time has been accompanied by more or less disability and healthcare costs.

1.8 OBJECTIVES

The purpose of this thesis was twofold: 1) to develop a framework that can utilize linked CCHS to evaluate the burden of unhealthy behaviours over a lifetime, and 2) assess the representativeness household survey respondents in the CCHS for linkage studies. This was accomplished through fives studies that utilized linked Ontario subsamples of the CCHS that are available at ICES. This organization holds an extensive repository of data holdings that can be linked to the CCHS.

The first objective was accomplished through three studies. The first study assessed the burden of unhealthy behaviours on period life expectancy utilizing a cause-deleted approach, expanding on methods developed in prior research.[45] This required the development of multivariable risk algorithms for mortality that included five unhealthy behaviours (smoking, leisure physical inactivity, inactive transport, leisure sedentary activity, and poor diet). The second study assessed the burden of unhealthy behaviours on lifetime healthcare costs. This involved development of phase-based multivariable risk algorithms for healthcare costs based on proximity to death. These were combined with the previously derived multivariable risk

algorithms for mortality utilizing a period life table approach. The third study focused on the generation of quasi-longitudinal CCHS cohorts through imputation that could be leveraged for the development of multivariable transition algorithms. In particular, algorithms for transitions to different levels of immobility were created, but these methods can also be applied to transitions of unhealthy behaviours, or any exposures that are allowed to transition in the quasi-longitudinal cohorts. Such transition algorithms would allow for the creation of cohort life tables rather than period life tables. Combining these algorithms with the aforementioned algorithms for mortality and healthcare costs could potentially generate a framework by which to cause-delete the impact of an unhealthy behaviour on mortality, healthcare costs, and transitions simultaneously.

The second objective was accomplished through two studies. The first of these studies evaluated the representativeness of CCHS respondents for health outcome linkage studies. Novel research (i.e., as in this thesis) that utilize the linked CCHS routinely aggregates multiple linked cycles, and often do not incorporate survey weights in the manner that is generally prescribed. This study assessed the comparability of linked CCHS respondents over time, which like all national household surveys, has been experiencing declines in response and consent to linkage. At the same time, it provided insight into the challenges of generating life expectancies from survey respondents who are only sampled from the household population, which excludes residents of retirement homes and long-term care facilities. The second of these studies evaluated the representativeness of CCHS respondents for healthcare utilization and cost outcome linkage studies. High-healthcare use and costs are often concentrated within a small segment of the general population. Naturally, many of these individuals are not in the household population, but rather retirement homes and long-term care facilities. This would suggest that for many CCHS respondents, the costliest phases of their life are not generally captured in the period immediately following their survey administration. These studies highlighted the difficulties of generating lifetime healthcare costs from the household population, and the necessity to ensure estimates realistically incorporate the time individuals may live outside of the household population over the duration of their lifespans.

1. Data and Associated Derived Measures

2.1 SOURCE OF DATA HOLDINGS

The data from this thesis is held securely in coded form at ICES, an independent, non-profit institute whose legal status under Ontario's privacy law allows it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement. The use of the data in this thesis is authorized under section 45 of Ontario's Personal Health Information Protective Act (PHIPA) and does not require review by a Review Ethics Board. The datasets were linked using encoded identifiers and analyzed at ICES. Parts of this material are based on data compiled and provided by Statistics Canada, CIHI and the Ontario Ministry of Health. However, the analyses, conclusions, opinions, and statements expressed in the material are those of the author, and not necessarily those of Statistics Canada, CIHI or the Ontario Ministry of Health.

2.2 POPULATION FRAMES

Registered Person Database of Ontario

The Registered Person Database (RPDB) is a population-based registry maintained by the Ministry of Health and Long-Term Care in Ontario. It contains all unique health numbers that have been issued for Ontario Health Insurance Plan (OHIP), and their corresponding periods of coverage. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. Each individual record in RPDB is augmented by ICES with additional information which includes date of birth, sex, date of death, yearly postal codes, geographic indicators, and date of last contact with the healthcare system. The two most commonly utilized geographic indicators in ICES studies are neighbourhood income quintile and rurality index of Ontario score (RIO).[128] The neighbourhood income quintile is obtained through the postal code conversion file which utilizes data from the nearest census.[129] Within each census, the average income per single person equivalent is calculated for each dissemination area. These areas are ranked within each metropolitan area, census agglomeration, or residual area. The population within each of these areas are then divided into fifths. RIO is a 0 to 100 score which is widely used as an aid to define rural areas. It was originally used by the Ministry of Health and Long-term Care and Ontario

Medical Association to determine incentive payment levels for physicians. It is habitually categorized into urban (0-9), suburban (10-39), and rural (40+). RPDB is routinely used at ICES to derive population-based cohorts that are representative of the Ontario population in a given year or over a specific timeframe.

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian population. It is primarily used for health surveillance and population health research. The surveys are conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions (e.g., long-term care and correctional facilities) and collective dwellings (e.g., retirement and group homes). The survey employs a complex multistage sampling strategy that results in a weight being assigned to each one of its participants. This weight indicates the number of individuals in the household population that is represented by the participant, after adjusting for unequal selection probabilities and non-response. The CCHS is designed to produce robust estimates at the health region level. As a result, it samples at much higher rates in sparsely populated regions, and much lower rates in heavily populated regions. Unweighted Ontario subsamples appear much older and rural than the Ontario population. The details of the survey methodology have been previously published and are available online.[130] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content have been linked to RPDB at ICES. These include 1.1 (2000-2001), 2.1(2003-2004), 3.1(2005-2006), and 2007 to 2014.

2.3 HEALTH SERVICES AND UTILIZATION DATA

Multiple types of administrative data were linked using unique encoded identifiers to the population frames for the purpose of this thesis. This included records from the Ontario Health Insurance Plan (OHIP) database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), Ontario Mental Health Reporting System (OMHRS) database, National Rehabilitation Reporting System (NRS)

database, Same-Day Surgery (SDS) database, National Ambulatory Care Reporting System (NACRS) database, and Home Care Database (HCD). The OHIP database contains claims to Ontario's publicly funded health insurance system by health care practitioners for inpatient and outpatient services. The ODB database contains prescription claims covered by the provincial drug program, which is restricted to those aged 65 and older, LTC residents, or service recipients of home care, social assistance and special drug programs. The CCRS database contains information for episodes of complex continuing care within hospitals and long-term care facilities, where individuals receive 24-hour nursing services. Reporting to CCRS database was only mandated for all long-term care facilities in 2009. The DAD contains information for all admissions and transfers to inpatient care within acute, rehabilitation, and chronic care institutions. The OMHRS database contains information for all admissions to adult inpatient hospital care beds designated for mental health. Before April 2006, these admissions were captured within the DAD. The NRS database contains client data collected from participating adult inpatient rehabilitation facilities and programs. The SDS database contains information for all visits to day surgery institutions. The NACRS database contains information for all patient visits to community and hospital ambulatory care centres. The HCD is a clinical client centric database that captures all services that are provided by or coordinated by Community Care Access Centres.

2.4 COMORBIDITY MEASURES

Identification of Individual Diseases

Comorbidities were identified throughout the thesis in a variety of ways. ICES-derived validated disease algorithms were utilized to identify several individual diseases. These algorithms have been created at ICES and validated through medical chart review. Each of these algorithms use a combination of records from different administrative healthcare databases (e.g., hospital, outpatient, emergency department, and drug prescription records) with specific diagnostic codes. ICES applies these algorithms to their administrative healthcare databases on an annual basis to generate cumulative diseases-specific cohorts with the earliest estimated diagnosis dates. These ICES-derived cohorts were utilized to identify dementia, asthma, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and hypertension.[131–136] Other diseases were identified through an in-hospital diagnosis or two separate relevant OHIP

billings within two years of each other. These diseases included renal disease, history of stroke, osteoporosis, heart arrhythmias, ischemic heart disease, history of cancer, and osteoarthritis. For mental illness, an in-hospital diagnosis or three separate relevant OHIP billings within three years of each other was employed since many OHIP codes in this area are fairly generic (e.g., general anxiety). The look back windows used when evaluating the presence of pre-existing diseases without a validated algorithm varied depending on the thesis chapter. The sixteen diseases were routinely summed in several chapters to provide an indicator of multimorbidity.

Classification of Health Profiles

The Canadian Institute for Health Information's (CIHI) population grouping methodology 1.1 was utilized to classify individuals based on their current level of illness.[137] This software profiles each person in the population by applying a case-mix classification system to person-level clinical information within a two year observation window. The preliminary version available at ICES is able to utilize DAD, NACRS, and SDS codes from April 2002 onwards, as well as OHIP for any time period. This version was modified for the purposes of this thesis to allow for observation windows specific to each individual in a cohort, rather than having a fixed observation window for all individuals in a cohort. The CIHI population grouping methodology identifies 226 health conditions, and aggregates them into 16 health profiles. These profiles were grouped into severe illness (palliative, major cancer, major mental illness, major chronic, and major acute), moderate illness (other cancer, other mental illness, moderate chronic, and moderate acute), and healthy (minor chronic, minor acute, major or healthy newborn, obstetrics, healthy, and non-user) using their observed mortality risk in the general population. The extent to which an individual had prior history of severe illness was determined by applying this methodology to multiple intervals within a lookback window preceding the observation window utilized to ascertain the current level of illness. Individuals with more intervals of severe illness were considered to have a greater history of severe illness.

2.5 HEALTHCARE COSTS

ICES-derived costing methodologies were applied to various types of utilization to generate corresponding healthcare costs in several chapters.[127] For healthcare utilization funded by the global budgets of institutions, healthcare costs per utilization unit is estimated using a top-

down approach that employs a case-mix methodology. In this scenario, the total annual cost of this system of a particular type of utilization is divided by its total annual weighted days to generate a corresponding cost per weighted day. An individual's annual cost with respect to a particular utilization is the product of its weighted days (i.e., adjusted for intensity of resource utilization) and cost per weighted days. The costs for utilization captured through the CCRS, DAD, NRS, OMHRS, NACRS, and SDS were estimated using this approach. Long-term care utilization indirectly estimated through OHIP and ODB were also estimated using this approach. The costs for utilization from the HCD, OHIP, and ODB were directly estimated from explicit fee payments, unless the healthcare professional involved in the encounter was compensated through a capitation system. In this latter case, costs for utilization were estimated using the compensation typically paid to the healthcare professionals with explicit fee payments. ICES-derived costing methodologies are only available from 2002 onwards. These methodologies were adapted for specific utilization prior to 2002 for several chapters in this thesis. All the pertinent costing information was available in these instances, except for the cost per weighted days estimates required for the top-down costing approaches. These missing values were extrapolated from the estimates available in the later years (i.e., from 2002 to 2020).

3. Burden of Unhealthy Behaviours on Ontario Period Life Expectancy in 2000-2004, 2005-2009, and 2010-2014

3.1 ABSTRACT

Background

Unhealthy behaviours -- such as smoking, leisure physical inactivity, non-active transport, leisure sedentary activity, and poor diet -- are leading risk factors for mortality. This study estimated the burden of these unhealthy behaviours on period life expectancy for the Ontario population.

Data and Methods

40- to 89-year-old Ontario respondents from multiple cycles of the Canadian Community Health Survey were pooled and linked to vital statistics data to ascertain deaths. 5-year predictive algorithms for mortality were developed and calibrated to general population mortality rates in different time periods (2000-2004, 2005-2009, and 2010-2014). Using a cause-deleted approach, the calibrated algorithms were applied to weighted linked Ontario subsamples to generate era-specific period life tables and corresponding counterfactual tables. The life years lost to each unhealthy behaviour, and all unhealthy behaviours combined were estimated by time period.

Results

Life years lost to unhealthy behaviours was 7.5 years in 2000-2004, 7.0 years in 2005-2009, and 6.7 years in 2010-2014. This reduction in unhealthy behaviour burden represents over 41% of the period life expectancy gains experienced during this time. In 2009-2014, 25.2% of the burden was attributed to non-active transport, 24.2% to smoking, 20.7% to leisure physical inactivity, 16.2% to poor diet, and 13.7% to leisure sedentary activity.

Interpretation

The large detrimental effect of unhealthy behaviours on period life expectancy in Ontario has been decreasing over time. Smoking behaviour is now responsible for only about a fifth of the unhealthy behaviour burden. Behaviours associated with physical inactivity and sedentary activity form most of the unhealthy behaviour burden. The burden of poor diet was likely

underestimated because diet was only ascertained through frequency intake of fruits and vegetables.

3.2 INTRODUCTION

It is widely acknowledged that unhealthy behaviours are leading risk factors for mortality. Smoking behaviour has historically received the largest attention among the various unhealthy behaviours. This attention has been largely warranted given that it is the leading cause of premature of death around the world.[138] Smokers have mortality rates more than two time higher than non-smokers, and live at least 10 years less than non-smokers.[44,47,48,139] The burden of smoking in many developed countries, however, is steadily decreasing because of large declines in smoking behaviour.[140] This has shifted the focus in many of these countries to physical inactivity and sedentary activity, unhealthy behaviours that are rapidly increasing in developed countries.[50] Studies have demonstrated that physically inactive individuals live at least three years less than active individuals.[53,141] Sedentary activity has demonstrated a similar impact on life expectancy.[54,55] In recent years, various developed countries have seen their life expectancy falter, reversing upward trends that have been consistent for almost a century.[142] Life expectancies could continue to deteriorate if societies become even more physically inactive and sedentary. Unfortunately, understanding the current and future potential burden of individual unhealthy behaviours on a population's life expectancy is a challenge. Unhealthy behaviours typically cluster within the population, and are more routinely observed among lower socioeconomic groups.[119,120,143] The burden of unhealthy behaviours on mortality has been typically estimated through indirect methods that combine population attributable fractions with aggregate data. These approaches are unable to adequately address the distribution of other baselines risks at the individual level.

The linkage of national population-based health surveys to mortality records has provided opportunities for more direct approaches. These surveys provide cross-sectional measures of lifestyle and sociodemographic information not generally captured within administrative healthcare databases. The linkage of these surveys to mortality records creates opportunities to develop multivariable predictive algorithms that can assess the relationship between unhealthy behaviours and mortality while simultaneously adjusting for numerous other factors. This

provides a more accurate picture of the causal relationship between each unhealthy behaviour and mortality. Using a cause-deleted approach, the mortality risk associated with a survey respondent can be estimated with and without the contribution attributed to each unhealthy behaviour. The difference between the baseline estimate and a counterfactual estimate where an unhealthy behaviour has been 'deleted' can be interpreted as a measure of the unhealthy behaviour's burden on mortality risk. Period life tables can be generated by applying these multivariable risk algorithms to different iterations of national population-based health surveys. Extending the cause-deleted approach to the period life tables provides burden estimates in terms of life years lost. This approach was previously utilized with the linked Canadian Community Health Survey (CCHS) to estimate the burden of unhealthy behaviours in Canada to be 6.0 years in 2009-2010.[47] Prior CCHS-based studies have examined multiple unhealthy behaviours (smoking, leisure physical inactivity, alcohol, and poor diet), but have not included non-active transport and leisure sedentary activity in their analysis.[45,99,111]

This study used the linked CCHS to estimate the mortality burden attributable to five behavioural risk factors in Ontario (smoking, leisure physical inactivity, leisure sedentary activity, inactive transport, and poor diet) using a multivariable predictive approach. This involved (1) development of 5-year predictive all-cause mortality risk algorithms using an unweighted CCHS cohort of survey respondents from multiple linked Ontario subsamples (1.1 (2000-2001) to 2013-2014), (2) calibration of these risk algorithms to general population mortality rates in different time periods to account for the sampling frame of the CCHS (i.e. household population), (3) application of these calibrated risk algorithms to linked Ontario subsamples to generate period-specific baseline and counterfactual period life tables, and (4) estimation of the life years lost for each unhealthy behaviour and all unhealthy behaviours combined within each period.

3.3 DATA & METHODS

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian

population. The survey employs a complex multistage sampling strategy and is conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions (e.g., long-term care facilities and penitentiaries) and collective dwellings (e.g., retirement and group homes). The details of the survey methodology have been previously published and are available online.[9] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content were linked to the Registered Person Database (RPDB) at ICES. This population-based registry contains all unique health numbers issued for the Ontario Insurance Health Plan (OHIP), and their corresponding period of coverage. Each individual record is augmented by ICES with additional information which includes date of birth, sex, date of death, yearly postal codes, geographic indicators, and date of last contact with the healthcare system. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. These groups are also excluded from the CCHS sampling frame. The linked Ontario subsamples from 1.1 (2000-2001), 2.1 (2003-04), 3.1(2005-06), and 2007 to 2014 were combined to create an initial linked CCHS cohort of 233,835 respondents for this study.

Administrative Healthcare Databases

Multiple types of administrative data were linked to the CCHS surveys to facilitate imputation of missing survey information and ascertainment of linked exposures. This included records from the OHIP database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), and Ontario Mental Health Reporting System (OMHRS) database, and Home Care Database (HCD). The OHIP database contains claims to Ontario's publicly funded health insurance system by health care practitioners for inpatient and outpatient services. The ODB database contains prescription claims covered by the provincial drug program, which is restricted to those aged 65 and older, LTC residents, or service recipients of home care, social assistance and special drug programs. The CCRS database contains information for episodes of complex continuing care within hospitals and long-term care facilities, where individuals receive 24-hour nursing services. The DAD contains information for all admissions and transfers to inpatient care within acute, rehabilitation, and chronic care institutions. The OMHRS database contains information for all

admissions to adult inpatient hospital care beds designated for mental health. The HCD is a clinical client centric database that captures all services that are provided by or coordinated by Community Care Access Centres.

Analytical Cohort

An analytical cohort was constructed from the linked CCHS cohort for derivation of the multivariable mortality risk algorithms. The index date (i.e., beginning of the observation window) of each respondent was designated as 90 days after their survey date. This washout period between the survey date and the index date was designed to minimize reverse causality between exposures and outcomes. Respondents under the age of 40 at their index date were excluded because health behaviours and sociodemographic factors are fluid at younger ages and are unlikely to be representative of lifetime exposures. Respondents 90 years of age or older were also excluded since they were extremely rare (<0.6%), and unlikely to demonstrate variation in health behaviours. These age restrictions reduced the cohort to 143,889 respondents. Respondents who identified as pregnant at time of their survey were removed since they were unlikely to report behaviours representative of their lives historically. Respondents who were not alive and residents of Ontario at their index date, or not eligible for OHIP within 6 months of their index date were also excluded. The final linked CCHS cohort consisted of 143,277 respondents. Individuals in this cohort were followed up to a maximum of five years or death, and censored at either the beginning of any time period of 6 months or longer where they were OHIP ineligible.

Unhealthy Behaviours

Unhealthy behaviours were the main exposures of interest, and were included in all algorithms. Descriptions of the unhealthy behaviour exposures are provided in Table 3.1. The CCHS has an extensive number of questions which evaluate present and past smoking behavior. Smoking behavior was defined by combining survey responses regarding smoking status, daily cigarette consumption, and years since having quit daily smoking consumption. The CCHS assesses participation in numerous different types of leisure physical activities in the three months prior to the survey. For each activity, an average daily expenditure is calculated using the frequency, average duration, and energy cost associated with the activity. Summing the average daily expenditures of all the ascertained leisure activities provides an overall average daily

expenditure in metabolic equivalent of task (MET) units. Leisure physical inactivity was defined by categorizing this aggregated summary measure. The CCHS uses several questions to evaluate the amount of time walking or cycling while commuting or conducting errands. Non-active transport was defined by categorizing the total weekly hours associated with these activities. The CCHS assesses participation in different types of leisure sedentary activity (i.e., time on the computer, playing video games, or reading) in the three months prior to the survey. Leisure sedentary activity was defined by categorizing the total weekly hours spent in these activities. Dietary quality is difficult to assess using health surveys. The CCHS captures fruit and vegetable consumption through a series of questions that evaluate frequency of intake rather than the amount consumed or number of servings. Frequencies are initially captured either per day, week, month, or year, but then subsequently converted to daily estimates of consumption frequency. The different dietary items examined in the CCHS include fruit, salad, carrot, potato, juice, and other vegetables.

Other Risk Factors

Additional risk factors were selected for the multivariable risk algorithms that varied in the degree to which they were distal to proximal risk factors of death. These included age, calendar date, sociodemographic factors, neighbourhood indicators, measures of immobility, and level of illness. The sociodemographic factors included were educational level, immigrant status, body mass index, sense of belonging, marital status, home ownership and food insecurity. Neighbourhood indicators for each respondent were supplemented to the cohort through linkages to the Registered Person's Database, which provided the postal code of each respondent in the year of their survey administration. The postal code was utilized in conjunction with postal code conversion files to identify whether the neighbourhood was low income (i.e. lowest quintile), and the rurality index of Ontario to identify whether the neighbourhood was urban, suburban, or rural.[128,129] Immobility was assessed by examining a respondent's self-reported inability to perform basic tasks, as well as their self-reported inability to participate in everyday activities due to a long-term health condition. Level of illness was assessed by applying the Canadian Institute for Health Information's (CIHI) population grouping methodology to OHIP billings within one year of the index date of each respondent.[137] This methodology identified the health profile of each respondent which was subsequently categorized into levels of illness severity.

Multiple Imputation

Missing information due to survey non-response or differences in the survey modules between cycles were important issues for consideration. Non-response to specific survey questions is habitually very low (e.g., <5%) within the CCHS. Most missing information is a product of changes in the cycles over time, which can be considered to be missing at random. Markov Chain Monte Carlo methods was utilized to generate 10 imputed datasets. The risk factors selected for the multiple risk algorithms and 5-year mortality information were included in the multiple imputation. In addition, the multiple imputation was supplemented by auxiliary information that could improve the accuracy and precisions of imputed values. For example, various additional sociodemographic factors and indicators of general health from the CCHS were incorporated such as labour force participation, source of main income, self-perceived stress, and self-perceived health. Pre-existing diseases that were identified using healthcare utilization records were also included. ICES-derived validated disease algorithms were utilized to identify dementia, asthma, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and hypertension.[131–136] Other diseases were identified based upon an in-hospital diagnosis or multiple OHIP billings occurring within a specified number of years of each other. These diseases included renal disease, history of stroke, osteoporosis, heart arrhythmias, ischemic heart disease, history of cancer, osteoarthritis, and mental illness. 10-year look back windows were used when evaluating the presence of pre-existing diseases without a validated algorithm.

Other auxiliary information utilized for multiple imputation included measures of past morbidity and healthcare costs. The CIHI population grouping methodology was applied to OHIP records in a 10-year window ending 12 months before the index date to assess prior history of illness. The number of intervals with severe and moderate illness were tabulated separately and included in the multiple imputation. Healthcare costs were generated by applying ICES-derived costing methodologies to various types of utilization that occurred in the four years prior to the index date and the four year observation window following the index date.[127] The different types of costs were aggregated into three sectors: chronic care (long-term care, complex continuing care), inpatient care (acute hospitalizations, and mental health hospitalizations), and community care (physician and other healthcare professional encounters, home care services, laboratory testing, and drug prescriptions). The rate at which healthcare costs

accumulate has been demonstrated to be closely associated with mortality risk.[144–146] The sector-specific costs were converted to rates and included in the multiple imputation.

Statistical Analysis

Cox proportional hazards models were utilized to evaluate the association between unhealthy behaviours and 5-year mortality risk of death. Multiple risk algorithms were constructed which varied in their level of adjustment, thereby providing a potential range for the causal relationship between each unhealthy behaviour and mortality. In total, four risk algorithms for mortality were constructed: a base, distal, intermediate, and proximal algorithm. Each algorithm incorporated additional risk factors considered more proximal than those utilized in the prior algorithm. On one end of this spectrum was the base algorithm which represented the least adjusted algorithm (essentially just unhealthy behaviours), and at the other end of the spectrum was the proximal algorithm which represented the most adjusted algorithm (all potential risk factors).

More precisely, the base algorithm included all unhealthy behaviours, along with age and calendar date. Age was employed as a time-dependent covariate (i.e., updated after each year of follow-up) with spline functions to account for potential proportional hazards assumption violations. Calendar date was included to account for temporal trends in mortality risk. A simple unhealthy diet score was derived from the dietary indicators using the parameter estimates of the base model. This score was categorized and utilized in an updated base model and all subsequent algorithms. This under-adjusted algorithm was expected to overestimate the causal relationship between unhealthy behaviours and mortality. The distal algorithm added sociodemographic indicators, and typically would be considered the algorithm that best portrays the causal relationships between unhealthy behaviours and mortality. However, many individuals likely only became physically inactive and sedentary after losing mobility or becoming ill. Thus, the distal algorithm was perceived as a slightly under-adjusted algorithm for many unhealthy behaviours. The intermediate algorithm added measures of mobility, and was considered a slightly over-adjusted algorithm since immobility is along the causal pathway to mortality. The proximal algorithm added level of illness, and was considered an over-adjusted algorithm, which was expected to underestimate causal relationships between unhealthy behaviours and mortality. Algorithms that included immobility and illness included interactions with age. Risk factors were

assessed for multicollinearity. Discrimination of the algorithms was evaluated utilizing the c-statistic, and calibration was evaluated across deciles and age groups. All statistical analysis was performed utilizing SAS Enterprise Guide 7.1.

Calculation of Burden Estimates

The risk algorithms were utilized to estimate the life years lost from each unhealthy behaviour on the Ontario population in three different time periods: 2000-2004, 2005-2009, and 2010-2014. To facilitate these calculations for the Ontario population, the risk algorithms were converted to 1-year algorithms and calibrated to general population mortality rates in each of the time periods. A 1-year baseline hazard was derived for each risk algorithm by averaging its five annual baseline hazards. Each of the four resulting 1-year risk algorithms were applied to the linked Ontario subsamples of the CCHS in different time periods, and the weighted age-specific mortality rates were calibrated to corresponding general population mortality rates obtained from RPDB. This produced a set of four calibrated risk algorithms for each of the three time periods. Each time period's set of calibrated risk algorithms was applied to linked Ontario subsamples of the CCHS in its corresponding time period to generate baseline period life tables from 40 to 89 years of age. Each period life table was extended to 99 years of age utilizing general population mortality rates of its corresponding time period obtained from RPDB. As a result, all subsequent life expectancy and life-years lost estimates were conditional upon having survived until age 40. Counterfactual period life tables for each time period's set of calibrated risk algorithms were also generated where unhealthy behaviours did not theoretically exist. This was accomplished by repeating the aforementioned process that generated the baseline period life tables, but rather than utilizing the original CCHS respondents, utilizing counterfactual respondents where the unhealthy behaviour of interest had been recoded to a 'healthy' reference. The difference in life years between a baseline period life table and a counterfactual period life table where an unhealthy behaviour had been 'deleted' was considered an estimate of its burden. For each unhealthy behaviour burden calculation within each time period, estimates from the base and distal algorithms were averaged to provide an upper bound for the life years lost, and estimates from the intermediate and proximal algorithms were averaged to provide a lower bound for the life years lost. The actual burden estimates of life years lost themselves were obtained by averaging the estimates from the distal and intermediate algorithms.

3.4 RESULTS

The analytical cohort had 11,361 deaths within its 692,933 person-years of follow-up. This cohort is described in Table 3.2. Each unhealthy behaviour was highly prevalent within the analytical cohort. The majority of CCHS respondents demonstrated either current or past history of daily smoking behaviour. Most respondents were observed to have non-healthy levels of leisure physical inactivity, non-active transport, and leisure sedentary activity. Only a quarter had an adequate diet. Age and sex standardized mortality rates of the unhealthy behaviours are presented in Table 3.3. Dose-response relationships were observed between each health behaviour and mortality, with the unhealthiest level of each behaviour exhibiting the largest mortality risk. Among the dietary items, low consumption of fruits, vegetables, salad, and carrots increased mortality risk, while low potato and juice consumption decreased mortality risk. The unhealthy diet score that was derived from the mortality risks associated with these dietary items is displayed in Table 3.4. The categorical exposures of the four risk algorithms that incorporated this unhealthy diet score are presented in Table 3.5. In general, the mortality risk associated with each unhealthy behaviour was increasingly attenuated as additional risk factors were added to the base algorithm. All the risk algorithms demonstrated high discrimination (C-statistics > 0.8), with discrimination improving from 0.820 (0.816-0.824) in the base algorithm to 0.863 (0.859-0.866) in the proximal algorithm. All risk algorithms demonstrated adequate calibration, except within the oldest age group (85-89) where risk algorithms underestimated mortality risk.

The application cohorts are presented in Table 3.6. In general, health behaviours demonstrated slight improvements over time. For example, current daily smoking, leisure physical inactivity, non-active transport, and less than adequate diets appear to be decreasing. In contrast, leisure sedentary activities appear to be increasing. This was reflective in the burdens of the unhealthy behaviours which are presented in Table 3.7. The overall burden of unhealthy behaviours decreased from 7.5 years in 2000-2004 to 6.7 years in 2010-2014. This decrease represented over 40% of the increase in period life expectancy. Over this time, the burden of smoking was reduced from 2.4 (2.3-2.5) years to 2.1 (2.0-2.2) years. It was less clear the extent to which the burden of poor diet, leisure physical inactivity, non-active transport, and leisure sedentary activity was changing over time. These burden estimates varied widely depending on the risk algorithm utilized. Nevertheless, the unhealthy behaviours associated with physical

inactivity and sedentary when combined consistently represented over half of the unhealthy behaviour burden in all time periods regardless of which risk algorithms were utilized.

3.5 DISCUSSION

This study combined multivariable risk algorithms with a cause-deleted approach to estimate the burden of unhealthy behaviours on period life expectancy in Ontario. Four risk algorithms were developed for mortality using population-based household surveys linked to mortality records. The algorithms differed in their level of adjustment to provide insight into the uncertainty surrounding the causal relationships between unhealthy behaviours and mortality. These algorithms were calibrated to the general population in different time periods. The calibrated algorithms were applied to CCHS in different time periods to generate baseline period life tables and corresponding counterfactual tables where unhealthy behaviours theoretically did not exist. The life years lost to each unhealthy behaviour was estimated for different time periods

According to this study's results, the burden of unhealthy behaviours appears to be decreasing in Ontario, but it continues to have a large detrimental effect on period life expectancy. Smoking still represents a large burden in Ontario despite the declining prevalence in daily smoking behaviours. It will continue to have a significant impact on period life expectancy in the future because of the burden associated with former daily smokers. The complete benefits of public health programs in reducing smoking behaviour will likely not be entirely realized until well into the future, when the proportion of current and former daily smokers are both dramatically reduced. The burden associated with poor diet appears to be one of the smallest, but given the limited scope of dietary questions in the CCHS, it is unlikely that the burden of poor diet in Ontario is fully captured in this study. Evidence from other research that have utilized specialized cycles of the CCHS focused on diet suggests that the burden of unhealthy diet in Ontario may be decreasing.[147] The extent to which unhealthy behaviours associated with physical inactivity (i.e. leisure physical inactivity and non-active transport) and sedentary activity are a burden on period life expectancy is difficult to quantify using cross-sectional healthy surveys. Nevertheless, even conservative estimates from over-adjusted algorithms suggest that they form the majority of unhealthy burden in Ontario. Public health programs that can successfully reduce these unhealthy behaviours in the population have the potential to significantly increase period life expectancy in Ontario.

Future research should focus on the development and incorporation of multivariable risk algorithms for behaviour change. Cross-sectional surveys only provide snapshots in time without accounting for temporal relationships between risk factors and health states. Risk factors observed in a cross-sectional survey may also not necessarily be indicative of past or future exposure. For example, a survey respondent who is typically physically inactive, may simply have happened to be physically active at the time of their survey administration. This raises questions around the risk estimates utilized in the multivariable algorithms, and whether they can actually be interpreted as estimates of causal relationships. Moreover, period life tables do not provide realistic estimates of cohort life expectancies since they assume that younger birth cohorts will exhibit the exposures exhibited by older birth cohorts later in life. The burden of younger highly sedentary birth cohorts is likely to be larger in the future than the burden presently observed in older birth cohorts. Incorporating multivariable risk algorithms for behaviour change into the period life tables would allow for construction of more realistic behavior trajectories for each birth cohort. Many individuals who are observed to be physically inactive and sedentary late in life, would have been physically active and non-sedentary earlier in life. Some portion of these changes in behaviour are a product of aging. Unhealthy behaviours, however, also tend to cluster late in life which suggests that individuals who exhibit unhealthy behaviours early in life are more likely to accumulate more unhealthy behaviours late in life. Cause-deleting a single unhealthy behaviour without accounting for its impact on other unhealthy behaviours is unrealistic. Multivariable risk algorithms for behaviour change would provide opportunities to 'delete' the behaviour change attribution associated with different unhealthy behaviours. Cause-deleting the mortality risk and behaviour change risk attributed to an unhealthy behaviour simultaneously would provide a more realistic counterfactual scenario and estimate of its burden on life expectancy.

3.6 TABLES

Table 3.1: Definitions of Healthy Behaviour Exposures

Unhealthy Behaviours	Definition
Smoking	
Heavy	≥ Pack/day daily smoking within last 4 years
Light	< Pack/day daily smoking within last 4 years
Recent Heavy	≥ Pack/day daily smoking 4 to less than 20 years ago
Recent Light	< Pack/day daily smoking 4 to less than 20 years ago
Non-Recent Heavy	≥ Pack/day daily smoking 20 or more years ago
Non-Recent Light	< Pack/day daily smoking 20 or more years ago
Non-Smoker	Never Daily Smoker
Leisure Physical Inactivity	
High	0 to < 0.5 Daily averaged MET
Moderate	0.5 to < 2.5 Daily averaged MET
Low	≥ 2.5 Daily averaged MET
Non-active Transport	
High	0 to < 2 weekly hours of active transport
Moderate	2 to < 6 weekly hours of active transport
Low	≥ 6 daily weekly hours of active transport
Leisure Sedentary Activity	
High	≥ 5.5 weekly hours
Moderate	2.5 to < 5.5 weekly hours
Low	0 to < 2.5 weekly hours
Fruit Consumption	
Inadequate	0 to < 0.5 times daily
Less than optimal	0.5 to < 3 or ≥ 4 times daily
Optimal	3 to < 4 times daily
Other Vegetable Consumption	
Inadequate	0 to < 0.5 times daily
Less than optimal	0.5 to < 3 or ≥ 4 times daily
Optimal	3 to < 4 times daily
Salad Consumption	
Low	< 1 time daily
High	≥ 1 time daily
Carrot Consumption	
No	None in past year
Yes	Any in past year
Potato Consumption	
High	≥ 1 time daily
Low	< 1 time daily
Juice Consumption	
High	≥ 2 times daily
Low	< 2 times daily

Table 3.2: Baseline characteristics of analytical cohort

Exposures	Respondents	Prevalence
CCHS Cycle		
1.1 (2000-2001)	18,755	13.1
2.1 (2003-2004)	19,833	13.8
3.1 (2005-2006)	19,586	13.7
2007-2008	21,740	15.2
2009-2010	20,915	14.6
2011-2012	20,764	14.5
2013-2014	21,684	15.1
Sex		
Female	79,132	55.2
Male	64,145	44.8
Age Groups		
40-44	17,343	12.1
45-49	15,326	10.7
50-54	17,283	12.1
55-59	18,902	13.2
60-64	18,471	12.9
65-69	16,844	11.8
70-74	14,225	9.9
75-79	11,992	8.4
80-84	8,511	5.9
85-89	4,380	3.1
Smoking		
Heavy	9,807	6.8
Light	22,316	15.6
Recent Heavy	6,498	4.5
Recent Light	11,334	7.9
Non-Recent Heavy	8,081	5.6
Non-Recent Light	18,289	12.8
Non-Smoker	66,953	46.7
Leisure Physical Inactivity		
High	34,746	24.3
Moderate	66,590	46.5
Low	41,941	29.3
Non-active Transport		
High	47,350	41.7
Moderate	55,047	31.0
Low	40,880	27.3
Leisure Sedentary Activity		
High	47,010	33.0
Moderate	55,521	38.4
Low	40,746	28.5

Fruit Consumption		
Inadequate	36,805	25.7
Less than optimal	91,232	63.7
Optimal	15,240	10.6
Other Vegetable Consumption		
Inadequate	32,376	22.6
Less than optimal	101,612	70.9
Optimal	9,289	6.5
Salad Consumption		
Low	78,227	54.6
High	65,050	45.4
Carrot Consumption		
No	16,427	11.5
Yes	126,850	88.5
Potato Consumption		
High	21,516	15.0
Low	121,761	85.0
Juice Consumption		
High	16,215	11.3
Low	127,063	88.7
Unhealthy Diet Score≠		
Extremely Inadequate (≥ 7 Pts)	34,940	24.4
Very Inadequate (5 to <7 Pts)	33,342	23.3
Inadequate (3 to <5 Pts)	37,430	26.1
Adequate (0 to <3 Pts)	37,565	26.2
Education		
Less than High school	32,349	22.6
High school	34,945	24.4
Post-secondary	75,983	53.0
Immigrant Status		
Yes	32,393	22.6
No	110,885	77.4
Body Mass Index		
Morbidly Obese (>=40)	4,515	3.2
Underweight (<20)	4,643	3.2
Other (20 to <40)	134,119	93.6
Sense of Belonging		
Very Weak	13,206	9.2
Not Very Weak	130,071	90.8
Marital Status		
Single/Divorced	54,954	38.4
Married/Common-Law	88,323	61.6
Home Ownership		
No	28,899	20.2
Yes	114,378	79.8

Food Insecurity		
Insecure	12,765	8.9
Secure	130,512	91.1
Neighborhood Income		
Lowest Quintile	27,728	19.4
Non-lowest Quintile	115,549	80.6
Rurality		
Rural	26,519	18.5
Suburban	46,110	32.2
Urban	70,648	49.3
Degree of Immobility		
Need Help with Tasks	20,459	14.3
Some Limitations	31,569	22.0
No Physical Limitations	91,248	63.7
Level of Illness		
Severely Ill	19,659	13.7
Moderately Ill	53,315	37.2
Relatively Healthy	70,303	49.1

[≠]Assignment of Dietary Points Described in Table 3.4

Table 3.3: Death Rates of Health Behaviour Exposures

Unhealthy Behaviours	Crude Death Rate per 1000 Person Years	Standardized≠ Rate per 1000 Person Years
Smoking		
Heavy	1,910	3,110
Light	1,520	2,150
Recent Heavy	2,120	2,180
Recent Light	1,210	1,380
Non-Recent Heavy	2,550	1,350
Non-Recent Light	1,700	960
Non-Smoker	1,030	860
Leisure Physical Inactivity		
High	2,670	2,040
Moderate	1,170	1,040
Low	730	750
Non-active Transport		
High	2,120	1,634
Moderate	1,100	970
Low	660	750
Leisure Sedentary Activity		
High	2,140	1,500
Moderate	1,280	1,170
Low	710	960
Fruit Consumption		
Inadequate	1,840	1,780
Less than optimal	1,300	1,090
Optimal	980	900
Other Vegetable Consumption		
Inadequate	2,070	1,740
Less than optimal	1,250	1,100
Optimal	760	820
Salad Consumption		
Low	1,700	1,470
High	1,040	930
Carrot Consumption		
No	2,380	2,160
Yes	1,270	1,110
Potato Consumption		
High	2,800	1,670
Low	1,160	1,130
Juice Consumption		
High	1,850	1,500
Low	1,340	1,190

Unhealthy Diet Score		
Extremely Inadequate	2,450	1,920
Very Inadequate	1,400	1,160
Inadequate	1,050	970
Adequate	790	800

≠Age-Sex Standardized to CCHS 1.1 (2000-2001) respondents

Table 3.4: Assignment of Unhealthy Diet Score Points

Dietary Item	Points
Fruit Consumption	
Inadequate	2.1
Less than optimal	1.0
Optimal	Reference
Other Vegetable Consumption	
Inadequate	3.2
Less than optimal	1.5
Optimal	Reference
Salad Consumption	
Low	1.9
High	Reference
Carrot Consumption	
No	2.7
Yes	Reference
Potato Consumption	
High	2.6
Low	Reference
Juice Consumption	
High	2.6
Low	Reference

Table 3.5: Hazard Ratios of Categorical Exposures in Risk Algorithms

Exposures	Algorithms			
	Base	Distal	Intermediate	Proximal
Sex				
Female	0.68 (0.65, 0.70)	0.60 (0.58, 0.63)	0.56 (0.53, 0.58)	0.61 (0.58, 0.63)
Male	Ref.	Ref.	Ref.	Ref.
Smoking				
Heavy	2.97 (2.76, 3.21)	2.54 (2.35, 2.74)	2.36 (2.19, 2.55)	2.18 (2.02, 2.35)
Light	2.30 (2.18, 2.44)	2.06 (1.94, 2.18)	2.01 (1.89, 2.13)	1.95 (1.84, 2.06)
Recent Heavy	2.12 (1.95, 2.31)	1.97 (1.81, 2.15)	1.88 (1.73, 2.05)	1.71 (1.57, 1.87)
Recent Light	1.57 (1.45, 1.70)	1.53 (1.40, 1.66)	1.50 (1.38, 1.63)	1.41 (1.30, 1.53)
Non-Recent Heavy	1.42 (1.31, 1.53)	1.38 (1.27, 1.49)	1.34 (1.25, 1.45)	1.29 (1.20, 1.39)
Non-Recent Light	1.17 (1.10, 1.24)	1.18 (1.11, 1.25)	1.18 (1.11, 1.25)	1.17 (1.11, 1.24)
Non-Smoker	Ref.	Ref.	Ref.	Ref.
Leisure Physical Inactivity				
High	1.76 (1.66, 1.87)	1.63 (1.54, 1.73)	1.43 (1.35, 1.52)	1.36 (1.28, 1.44)
Moderate	1.19 (1.13, 1.26)	1.16 (1.10, 1.23)	1.10 (1.04, 1.17)	1.07 (1.01, 1.13)
Low	Ref.	Ref.	Ref.	Ref.
Non-active Transport				
High	1.65 (1.51, 1.79)	1.60 (1.47, 1.74)	1.44 (1.33, 1.56)	1.33 (1.23, 1.43)
Moderate	1.18 (1.08, 1.29)	1.18 (1.08, 1.29)	1.13 (1.04, 1.24)	1.09 (1.01, 1.19)
Low	Ref.	Ref.	Ref.	Ref.
Sedentary Activity				
High	1.36 (1.27, 1.46)	1.31 (1.22, 1.41)	1.19 (1.11, 1.28)	1.14 (1.06, 1.22)
Moderate	1.14 (1.07, 1.21)	1.12 (1.04, 1.19)	1.07 (1.00, 1.14)	1.05 (0.98, 1.12)
Low	Ref.	Ref.	Ref.	Ref.
Unhealthy Diet Score				
Extremely Inadequate	1.59 (1.50, 1.69)	1.38 (1.30, 1.47)	1.34 (1.27, 1.43)	1.30 (1.22, 1.38)
Very Inadequate	1.27 (1.19, 1.35)	1.20 (1.13, 1.28)	1.19 (1.12, 1.27)	1.17 (1.10, 1.24)
Inadequate	1.13 (1.06, 1.20)	1.09 (1.02, 1.16)	1.08 (1.01, 1.15)	1.07 (1.00, 1.13)
Adequate	Ref.	Ref.	Ref.	Ref.
Education				
Less than High school		1.12 (1.07, 1.17)	1.11 (1.06, 1.16)	1.10 (1.05, 1.15)
High school		1.07 (1.02, 1.12)	1.09 (1.03, 1.14)	1.07 (1.01, 1.12)
Post-secondary		Ref.	Ref.	Ref.
Immigrant Status				
Yes		0.90 (0.86, 0.95)	0.91 (0.87, 0.95)	0.91 (0.87, 0.95)
No		Ref.	Ref.	Ref.
Body Mass Index				

Morbidly Obese (>=40)	1.60 (1.44, 1.78)	1.36 (1.22, 1.50)	1.28 (1.15, 1.42)
Underweight (<20)	2.02 (1.88, 2.17)	1.90 (1.77, 2.04)	1.71 (1.59, 1.84)
Other (20 to <40)	Ref.	Ref.	Ref.
Sense of Belonging			
Very Weak	1.38 (1.30, 1.46)	1.24 (1.17, 1.32)	1.19 (1.12, 1.26)
Not Very Weak	Ref.	Ref.	Ref.
Marital Status			
Single/Divorced	1.13 (1.09, 1.18)	1.14 (1.10, 1.19)	1.15 (1.10, 1.19)
Married/Common-Law	Ref.	Ref.	Ref.
Home Ownership			
No	1.20 (1.15, 1.26)	1.17 (1.12, 1.23)	1.15 (1.10, 1.20)
Yes	Ref.	Ref.	Ref.
Food Insecurity			
Insecure	1.29 (1.21, 1.39)	1.06 (0.99, 1.14)	1.01 (0.94, 1.08)
Secure	Ref.	Ref.	Ref.
Neighborhood Income			
Lowest Quintile	1.07 (1.02, 1.12)	1.05 (1.01, 1.10)	1.04 (1.00, 1.09)
Non-lowest Quintile	Ref.	Ref.	Ref.
Rurality			
Rural	1.06 (1.01, 1.12)	1.06 (1.00, 1.11)	1.10 (1.05, 1.16)
Suburban	1.07 (1.02, 1.11)	1.06 (1.02, 1.11)	1.08 (1.03, 1.13)
Urban	Ref.	Ref.	Ref.
Immobility at Age 60[¥]			
Need Help with Tasks		3.28 (3.02, 3.56)	1.96 (1.80, 2.13)
Physical Limitations		1.64 (1.51, 1.78)	1.26 (1.16, 1.37)
No Physical Limitations		Ref.	Ref.
Immobility at Age 80[¥]			
Need Help with Tasks		1.98 (1.88, 2.08)	1.61 (1.53, 1.69)
Physical Limitations		1.31 (1.24, 1.39)	1.18 (1.12, 1.25)
No Physical Limitations		Ref.	Ref.
Illness at Age 60[¥]			
Severely Ill			6.82 (6.26, 7.44)
Moderately Ill			1.47 (1.35, 1.61)
Relatively Healthy			Ref.
Illness at Age 80[¥]			
Severely Ill			3.01 (2.83, 3.20)
Moderately Ill			1.24 (1.16, 1.32)
Relatively Healthy			Ref.

[¥] **Exposure estimated at specific age because of age interaction**

Table 3.6: Characteristics of application cohorts: 2000-2004, 2005-2009, and 2013-2014

Exposures	Prevalence by Era		
	2000-2004	2005-2009	2010-2014
Sex			
Female	52.1	51.0	51.8
Male	47.9	49.0	48.2
Age Groups			
40-44	19.7	17.9	14.8
45-49	17.1	16.7	14.7
50-54	14.7	14.6	15.8
55-59	12.3	13.5	14.1
60-64	9.6	11.0	12.1
65-69	8.4	8.6	9.8
70-74	7.5	6.6	7.0
75-79	5.7	5.6	5.7
80-84	3.4	3.8	3.8
85-89	1.6	1.7	2.1
Smoking			
Heavy	7.3	5.4	4.3
Light	16.5	15.4	13.9
Recent Heavy	4.5	4.1	3.2
Recent Light	8.2	8.2	7.6
Non-Recent Heavy	4.4	4.2	4.2
Non-Recent Light	10.2	10.8	11.3
Non-Smoker	48.8	51.9	55.4
Leisure Physical Inactivity			
Inactive	27.5	24.9	24.1
Moderately Inactive	47.1	47.1	45.3
Active	25.4	28.0	30.6
Non-active Transport			
High	44.7	40.2	37.6
Moderate	31.6	31.9	29.3
Low	23.7	27.9	33.2
Leisure Sedentary Activity			
Sedentary	22.5	24.5	30.9
Moderately Sedentary	37.3	37.8	38.3
Non-Sedentary	40.2	37.7	30.9
Fruit Consumption			
Inadequate	25.9	24.3	24.5
Less than optimal	63.8	65.1	64.8
Optimal	10.3	10.6	10.7
Other Vegetable Consumption			
Inadequate	22.4	21.8	22.6

Moderately Inadequate	72.4	71.5	69.9
Adequate	5.1	6.7	7.4
Salad Consumption			
Low	53.5	51.6	52.7
High	46.5	48.4	47.3
Carrot Consumption			
No	10.2	12.3	13.5
Yes	89.8	87.7	86.5
Potato Consumption			
High	15.9	11.4	8.5
Low	84.1	88.6	91.5
Juice Consumption			
High	14.5	12.4	8.2
Low	85.5	87.6	91.8
Unhealthy Diet Score			
Extremely Inadequate	24.5	21.9	20.6
Very Inadequate	25.8	23.0	21.1
Inadequate	26.0	26.8	27.4
Adequate	23.7	28.3	30.9
Education			
Less than High school	23.3	18.2	14.0
High school	26.9	23.5	23.3
Post-secondary	49.8	58.3	62.7
Immigrant Status			
Yes	34.3	36.3	36.6
No	65.7	63.7	63.4
Body Mass Index			
Morbidly Obese (>=40)	1.9	2.6	3.3
Underweight (<20)	3.6	3.3	3.4
Other (20 to <40)	94.5	94.2	93.4
Sense of Belonging			
Very Weak	19.6	17.5	20.6
Not Very Weak	80.4	82.5	79.4
Marital Status			
Single/Divorced	24.7	25.3	26.9
Married/Common-Law	75.3	74.7	73.1
Home Ownership			
No	19.6	17.5	20.6
Yes	80.4	82.5	79.4
Food Insecurity			
Insecure	10.2	9.2	8.9
Secure	89.8	90.8	91.1
Neighborhood Income			
Lowest Quintile	17.8	17.8	17.7
Non-lowest Quintile	82.2	82.2	82.3

Rurality			
Rural	9.4	9.1	8.7
Suburban	21.9	21.3	21.7
Urban	68.8	69.6	69.6
Degree of Immobility			
Need Help with Tasks	11.2	12.0	12.4
Some Limitations	19.7	18.7	18.6
No Physical Limitations	69.1	69.3	69.0
Level of Illness			
Severely Ill	10.4	11.2	11.6
Moderately Ill	33.8	34.1	35.8
Relatively Healthy	55.7	54.7	52.7

Table 3.7: Burden Estimates for CCHS 2000-2004, 2005-2009, and 2013-2014

Scenario	Life Expectancy	Life Years Lost	% Of Unhealthy Burden
2000-2004			
Baseline	82.1	-	-
No Smoking	84.5 (84.4-84.6)	2.4 (2.3-2.5)	25.9%
No Leisure Physical Inactivity	84.1 (83.7-84.5)	2.0 (1.6-2.4)	21.3%
No Non-Active Transport	84.4 (84.0-84.7)	2.3 (1.9-2.6)	25.5%
No Sedentary Activity	83.1 (82.8-83.4)	1.0 (0.7-1.3)	11.1%
No Poor Diet	83.7 (83.6-83.9)	1.6 (1.5-1.8)	16.2%
No Unhealthy Behaviours	89.6 (88.6-90.4)	7.5 (6.5-8.3)	100.0%
2005-2009			
Baseline	83.1	-	-
No Smoking	85.3 (85.2-85.4)	2.2 (2.1-2.3)	25.3%
No Leisure Physical Inactivity	84.9 (84.6-85.3)	1.8 (1.5-2.2)	20.8%
No Non-Active Transport	85.3 (84.9-85.6)	2.2 (1.8-2.5)	25.4%
No Sedentary Activity	84.2 (83.9-84.5)	1.1 (0.8-1.4)	12.2%
No Poor Adequate Diet	84.5 (84.4-84.8)	1.4 (1.3-1.7)	16.3%
No Unhealthy Behaviours	90.1 (89.2-90.9)	7.0 (6.1-7.8)	100.0%
2010-2014			
Baseline	84.2	-	-
No Smoking	86.3 (86.2-86.4)	2.1 (2.0-2.2)	24.2%
No Leisure Physical Inactivity	86.0 (85.6-86.4)	1.8 (1.4-2.2)	20.7%
No Non-Active Transport	86.4 (86.0-86.6)	2.2 (1.8-2.4)	25.2%
No Sedentary Activity	85.4 (85.1-85.7)	1.2 (0.9-1.5)	13.7%
No Poor Diet	85.6 (85.5-85.9)	1.4 (1.3-1.7)	16.2%
No Unhealthy Behaviours	90.9 (90.0-91.6)	6.7 (5.8-7.4)	100.0%

4. Burden of Unhealthy Behaviours on Period Lifetime Healthcare Costs of Ontarians in 2000-2004, 2005-2009, and 2010-2014

4.1 ABSTRACT

Background

Unhealthy behaviours -- such as smoking, leisure physical inactivity, non-active transport, leisure sedentary activity, and poor diet -- simultaneously impact mortality and healthcare costs. It is unclear whether unhealthy behaviours result in greater or lower lifetime healthcare costs. This study estimated the burden of these unhealthy behaviours on period lifetime healthcare costs for the Ontario population.

Data and Methods

40- to 89-year-old Ontario respondents from multiple cycles of the Canadian Community Health Survey (1.1, 2.1, 3.1, and 2007-2014) were pooled and linked to health administrative data within a 2-year observation window. The follow-up was divided into three phases based on proximity to death. Predicative algorithms for healthcare costs within each phase were developed and calibrated to healthcare costs in the general population for different time periods (2000-2004, 2005-2009, and 2010-2014). These calibrated algorithms were applied in conjunction with previously derived period-specific predictive mortality algorithms to linked subsamples of the CCHS. This process generated baseline and counterfactual period life tables with corresponding period lifetime healthcare costs. The difference between baseline and counterfactual estimates provided measures of burden for unhealthy behaviours on period lifetime healthcare costs.

Results

Unhealthy behaviours increased period lifetime healthcare costs across all time periods: 10.2% in 2000-2004, 11.6% in 2005-2009, and 12.9% in 2010-2014. Smoking and poor diet reduced period lifetime healthcare costs, while non-active transport and leisure sedentary activity increased period lifetime healthcare costs.

Interpretation

The impact of unhealthy behaviours on lifetime healthcare costs is complex and can vary drastically. Further research is needed to improve our understanding of how unhealthy behaviours influence lifetime healthcare costs.

4.2 INTRODUCTION

Unhealthy behaviours simultaneously impact mortality and healthcare costs. In general, individuals with unhealthy behaviours have much shorter life expectancies than individuals with healthy behaviours.[47,49,53,55,56] Unsurprisingly, these individuals with unhealthy behaviours also typically use healthcare services at a much higher rate and more intensely than those with healthy behaviours.[92,99,105] This translates into more rapid accumulation of healthcare costs among the individuals with unhealthy behaviours.[111] It is plausible however that individuals with healthy behaviours actually accumulate more healthcare costs over their lifetimes because of their longer life expectancies. Unfortunately, long-term longitudinal studies that follow individuals until death and simultaneously capture measures of health behaviours and healthcare costs are extremely rare. Longitudinal studies that do exist are often limited in sample size, have insufficient follow-up, or are not generalizable. Given the lack of appropriate data, it is not uncommon for researchers to take behavioural exposures, socioeconomic factors, measures of disability, healthcare costs, and mortality statistics from different sources.[115,116] This makes it difficult to decompose the impact of different unhealthy behaviours that often occur in tandem, and more frequently within low socioeconomic groups.[117–120] Studies that have attempted to incorporate differences in life expectancies when evaluating the lifetime burden of smoking on healthcare costs have produced mixed results.[113,114,121,122]

National health surveys have been our main source of health behaviour surveillance. The Canadian Community Health Survey (CCHS) is a population-based survey routinely administered by Statistics Canada.[123] Similar to other population health surveys worldwide, the CCHS was originally designed to provide cross-sectional measures of exposures (e.g. lifestyle and sociodemographic information) and health status not generally captured within the healthcare system, nor readily available through other sources. The linking of these surveys to administrative healthcare databases has provided opportunities to directly evaluate the impact of

unhealthy behaviours on different forms of healthcare utilization and mortality using the same population. The extent of linkage has also reached a point in certain jurisdictions where the cost of almost all the different types of publicly funded healthcare associated with a survey respondent can be directly estimated.[127] This is a major development for researchers that traditionally have had to piece together healthcare costs from sources that were not even consistent with the sources providing the measures of disease, disability, and mortality. This provides opportunities to more accurately assess the economic impact of unhealthy behaviours, and potentially determine whether individuals with healthy behaviours actually do have higher lifetime healthcare costs than individuals with unhealthy behaviours.

The availability of actual costing data has introduced its own set of challenges. Costing data is characterized by distributions that are challenging to model using standard statistical approaches.[148] Moreover, the rate at which healthcare costs accumulate typically increases as one approaches death. The period of time immediately before death is particularly expensive and typically constitutes a very large portion of one's lifetime healthcare costs.[149] Longitudinal incidental methods of costing are being increasingly utilized since it is not practical for most healthcare studies to follow all participants until death. Phase-based healthcare costing is one such approach that arises from the realization that healthcare costs are often closely aligned with certain phases within one's life.[150–152] In this approach, follow-up is divided into important phases that are associated with different levels of healthcare costs. The cost within each phase is modelled separately and superimposed on survival curves based on the probability of each phase occurring at each interval along the survival curve. These approaches can be utilized to compare the lifetime healthcare costs of different populations. Unfortunately, the majority of these methods do not address the simultaneous impact of covariates on survival and healthcare cost accumulation. As a result, differences in lifetime healthcare costs between two populations may not be only a product of differences in the exposure of interest (e.g., smoking). This problem can be addressed by utilizing survival models and phase-based cost models which have the same covariates.[153] This provides opportunities to generate counterfactual estimates where the impact of a covariate has been 'deleted' from the survival trajectory and the trajectory of healthcare cost accumulation simultaneously.

This study used the linked CCHS to estimate the portion of lifetime healthcare costs attributable to five behavioural risk factors in Ontario (smoking, leisure physical inactivity, leisure sedentary activity, inactive transport, and poor diet) using a multivariable predictive approach. This involved (1) development of predictive healthcare cost risk algorithms for three phases (>18 months from death, 18 to > 6 months, and last 6 months of life) using an unweighted CCHS cohort of survey respondents from multiple linked Ontario subsamples (1.1 (2000-2001) to 2013-2014), (2) calibration of these phase-based cost risk algorithms to equivalent phase-base costs in the general population in different time periods to account for the sampling frame of the CCHS (i.e. household population), (3) application of these calibrated phase-based cost risk algorithms to link Ontario subsamples to generate period-specific baseline and counterfactual phase-based costs per age, (4) creation of period-specific baseline and counterfactual lifetime costs by combining the calibrated period-specific baseline and counterfactual phase-based costs with corresponding period-specific baseline and counterfactual period life tables derived in prior research (i.e., Chapter 3), (5) estimation of the change in period lifetime healthcare costs associated with the removal of each unhealthy behaviour and all unhealthy behaviours combined within each period, and (6) a sensitivity analysis where unhealthy behaviours were not cause-deleted from the two phases which are closer to death.

4.3 DATA & METHODS

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian population. The survey employs a complex multistage sampling strategy and is conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions (e.g., long-term care facilities and penitentiaries) and collective dwellings (e.g., retirement and group homes). The details of the survey methodology have been previously published and are available online.[9] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content were linked to the Registered Person Database (RPDB) at ICES. This population-based registry contains all unique health numbers issued for

the Ontario Insurance Health Plan (OHIP), and their corresponding period of coverage. Each individual record is augmented by ICES with additional information which includes date of birth, sex, date of death, yearly postal codes, geographic indicators, and date of last contact with the healthcare system. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. These groups are also excluded from the CCHS sampling frame. The linked Ontario subsamples from 1.1 (2000-2001), 2.1 (2003-04), 3.1(2005-06), and 2007 to 2014 were combined to create an initial linked CCHS cohort of 233,835 respondents for this study.

Health Services & Utilization Data

Multiple types of administrative data were linked to the linked CCHS surveys to facilitate imputation of missing survey information and ascertainment of healthcare costs. This included records from the OHIP database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), and Ontario Mental Health Reporting System (OMHRS) database, and Home Care Database (HCD). The OHIP database contains claims to Ontario's publicly funded health insurance system by health care practitioners for inpatient and outpatient services. The ODB database contains prescription claims covered by the provincial drug program, which is restricted to those aged 65 and older, LTC residents, or service recipients of home care, social assistance and special drug programs. The CCRS database contains information for episodes of complex continuing care within hospitals and long-term care facilities, where individuals receive 24-hour nursing services. The DAD contains information for all admissions and transfers to inpatient care within acute, rehabilitation, and chronic care institutions. The OMHRS database contains information for all admissions to adult inpatient hospital care beds designated for mental health. The HCD is a clinical client centric database that captures all services that are provided by or coordinated by Community Care Access Centres.

Analytical Cohort

An analytical cohort was constructed from the linked CCHS cohort for derivation of the multivariable healthcare cost algorithms. This cohort was constructed in prior work (see Chapter 3). Briefly, this cohort consisted of 143,277 respondents who were followed up to a maximum of five years or death, and censored at either the beginning of any time period of 6 months or longer

where they were OHIP ineligible. The unhealthy behaviour exposures are provided in Table 4.1, and ascertainment of these unhealthy behaviours by the CCHS can be found in prior work (Chapter 3). The only difference was poor diet being categorized by three levels in this chapter rather than the four levels in prior work (i.e., the two intermediate levels of diet are collapsed into one). The other risk factors that were included in the risk algorithms and auxiliary information utilized for multiple imputation are also described in the prior work. Markov Chain Monte Carlo methods were utilized to address missing information due to survey non-response or differences in the survey modules between cycles, resulting in 10 imputed datasets.

Statistical Analysis

Healthcare costs were estimated for three different phases of life: regular life (>18 months from death), near end-of-life (18 to > 6 months from death), and end-of-life (last 6 months of life). The follow-up of the analytical cohort was divided into these three phases using follow-up within the first two years, and corresponding healthcare costs were tabulated within each of the phases. Costs were inflation-adjusted to 2018 and expressed in annual cost per capita terms. Gamma generalized linear models were utilized to evaluate the association between unhealthy behaviours and healthcare costs within each phase. Multiple risk algorithms were constructed within each phase which varied in their level of adjustment. This was designed to produce a potential range for the causal relationship between each unhealthy behaviour and healthcare costs (paralleling methods in Chapter 3). In total, four risk algorithms for were constructed for each phase: a base, distal, intermediate, and proximal algorithm. Each algorithm incorporated additional risk factors considered more proximal than those utilized in the prior algorithm. On one end of this spectrum were the base algorithms which represented the least adjusted algorithms (essentially just unhealthy behaviours), and at the other end of the spectrum was the proximal algorithms which represented the most adjusted algorithm (all potential risk factors).

More precisely, the base algorithm included all unhealthy behaviours, along with age and calendar date. Age was employed with spline functions to account for non-linear relationships. Calendar date was included to account for temporal trends in healthcare costs. This severely under-adjusted algorithm was expected to overestimate the causal relationship between unhealthy behaviours and mortality. The distal algorithm added sociodemographic indicators,

and typically would be considered the algorithm that best portrays the causal relationships between unhealthy behaviours and healthcare costs. However, many individuals likely only became physically inactive and sedentary after losing mobility or becoming ill. Thus, the distal algorithm was perceived as a slightly under-adjusted algorithm for many unhealthy behaviours. The intermediate algorithm added measures of immobility, and was considered a slightly over-adjusted algorithm since immobility is along the causal pathway to healthcare costs. The proximal algorithm added level of illness, and was considered an over-adjusted algorithm, which was expected to underestimate causal relationships between unhealthy behaviours and healthcare costs. Algorithms that included immobility and illness included interactions with age. Risk factors were assessed for multicollinearity. Calibration was evaluated across deciles and age groups. All statistical analysis was performed utilizing SAS Enterprise Guide 7.1.

Calculation of Burden Estimates

The risk algorithms were utilized to estimate the attribution of period lifetime healthcare costs from each unhealthy behaviour in three different time periods: 2000-2004, 2005-2009, and 2010-2014. Each of the four risk algorithms for each phase were applied to the linked Ontario subsamples of the CCHS in different time periods, and the weighted age-specific cost estimates were calibrated to corresponding healthcare costs in the general population. Phase-based healthcare costs for the general population were obtained by applying ICES-derived costing methodologies to utilization of Ontarians in RPDB. As a result, four calibrated risk algorithms were produced for each phase and period combination. These calibrated algorithms were applied to linked Ontario subsamples of the CCHS in their corresponding period to generate baseline phase-specific healthcare cost estimates from 40 to 89 years of age within each time period. Phase-based healthcare costs for ages 90 to 99 years of age were obtained from the general population. All subsequent lifetime healthcare costs estimates were therefore conditional upon having survived until age 40.

Counterfactual estimates were also generated where unhealthy behaviours did not theoretically exist. This was accomplished by repeating the aforementioned process that generated the baseline estimates, but rather than utilizing the original CCHS respondents, utilizing counterfactual respondents where the unhealthy behaviour of interest had been recoded to a ‘healthy’ reference. These baseline and counterfactual phase-based healthcare cost estimates

by age were superimposed on corresponding baseline and counterfactual period life tables derived in prior research (Chapter 3). The probability of each phase occurring within each age in a lifetable was obtained in a straightforward manner since the phases are defined by proximity to death. The difference in lifetime healthcare costs between a baseline and a counterfactual estimate where an unhealthy behaviour had been 'deleted' was considered an estimate of its burden. For each unhealthy behaviour burden calculation within each time period, estimates from the base and distal algorithms were averaged to provide an upper bound for the life years lost, and estimates from the intermediate and proximal algorithms were averaged to provide a lower bound for the life years lost. The actual burden estimates themselves were obtained by averaging the estimates from the distal and intermediate algorithms. A sensitivity analysis was performed where unhealthy behaviours were not 'deleted' from the two phases closer to death, since it may not be feasible to reverse several of the unhealthy behaviours this late in life.

4.4 RESULTS

The analytical cohort resulted in three phased-based cohorts that are described in Tables 4.2 to 4.4. As expected, phases closer to death had higher rates of healthcare costs. Healthcare costs accumulated at a rate of approximately $250 per person-month in the regular life phase, $1,750 in the near end-of-life phase, and $5,700 in the end-of-life phase. Unhealthy behaviours were more prevalent the closer a phase-based cohort was to death, which was a reflection of the higher risk of mortality associated with the unhealthy behaviours. The categorical exposures of the healthcare cost risk algorithms for each phase are displayed in Tables 4.5 to 4.7. Dose-response relationships were observed between each health behaviour and healthcare costs within each phase, with the unhealthiest level of each behaviour typically exhibiting the largest risk. The one exception was smoking within the near end-of-life and end-of-life phases, where the unhealthiest levels of smoking exposure actually had a lower risk for healthcare costs. Within each phase, the risk of healthcare costs associated with healthy behaviours were attenuated as more proximal risk factors were added to the base algorithm. All risk algorithms demonstrated adequate calibration across deciles and age groups with respect to the rates of healthcare cost accumulation with predicted estimates never differing more than 10% from actual values in categories representing at least 10% of a phase-based cohort.

The application cohorts are presented in Table 4.8. In general, health behaviours demonstrated slight improvements over time. For example, current daily smoking, leisure physical inactivity, non-active transport, and less than adequate diets appear to be decreasing. In contrast, leisure sedentary activities appear to be increasing. The impact of unhealthy behaviours over time are presented in Table 4.9. Overall, the five unhealthy behaviours combined to increase lifetime healthcare costs across all periods: 10.2% (2.5% to 17.7%) in 2000-2004, 11.6% (4.0% to 18.9%) in 2005-2009, and 12.9% (5.6% to 19.8%) in 2010-2014. However, individual unhealthy behaviours did not have necessarily the same effect on lifetime healthcare costs. Smoking and poor diet lowered lifetime healthcare costs, while inactive transport and sedentary activity raised lifetime healthcare costs. Whether an unhealthy behavior lowered or raised lifetime healthcare costs was not related to the extent to which it placed a burden on life expectancy, since smoking had comparable burdens to non-active transport on period life expectancy (See Chapter 3) but very different impacts on lifetime healthcare costs. In the sensitivity analysis presented in Table 4.10, where unhealthy behaviours were not 'deleted' from phases closer to death, the burden of the five unhealthy behaviours was attenuated: 5.1% (-2.4% to 12.7%) in 2000-2004, 7.1% (0.4% to 14.5%) in 2005-2009, and 8.7% (1.5% to 15.7%) in 2010-2014.

4.5 DISCUSSION

This study combined multivariable risk algorithms for healthcare costs with a cause-deleted approach to estimate the burden of unhealthy behaviours on period lifetime healthcare costs in Ontario. A population-based household survey linked to healthcare costs was utilized to develop risk algorithms with respect to three different phases of life that were based on proximity to death. In each phase, multiple risk algorithms were developed that differed in their level of adjustment to provide insight into the uncertainty surrounding the causal relationships between unhealthy behaviours and healthcare costs. These algorithms were calibrated to phase-based healthcare costs of the general population in different time periods. The calibrated algorithms were applied to the CCHS in conjunction with previously derived mortality risk algorithms to generate baseline and counterfactual lifetime healthcare costs in different time periods. The difference between these baseline and counterfactual estimates provided a measure of burden on period lifetime healthcare costs for each unhealthy behaviour.

According to this study's results, the burden of unhealthy behaviours on period lifetime healthcare costs appears to be increasing in Ontario, but the impact varies by unhealthy behaviour. Smoking, leisure physical inactivity, and poor diet reduced lifetime healthcare costs, but non-active transport and sedentary activity increased them. Noting that smoking and poor diet are decreasing over time, and sedentary activity is simultaneously increasing, these results may have implications for cost escalation. This would suggest that lifetime healthcare costs could increase in Ontario in the future. These trends were still apparent even when healthcare costs in the last 18 months of life were held constant, albeit to a lesser degree. Nevertheless, attribution of lifetime healthcare costs may still be confounded since it is unclear the degree to which some unhealthy behaviours are a product of others. For example, smokers may have had a higher propensity than non-smokers to use non-active transport or become sedentary later in life because of deteriorating health. Unhealthy behaviours tend to cluster late in life which suggests that individuals who exhibit unhealthy behaviours early in life are more likely to accumulate more unhealthy behaviours late in life. Cross-sectional surveys do not provide insight into temporal relationships between risk factors and health states. This raises questions around the risk estimates utilized in the multivariable algorithms, and whether they can actually be interpreted as estimates of causal relationships. Moreover, period life tables do not provide realistic estimates of life expectancies since they assume that younger birth cohorts will exhibit the exposures exhibited by older birth cohorts later in life. Disentangling the contributions of unhealthy behaviours on lifetime healthcare costs requires more sophisticated methods.

Future research should focus on the development and incorporation of multivariable risk algorithms for behaviour change and different health states. Incorporating multivariable risk algorithms for behaviour change into the period life tables would allow for construction of more realistic behavior trajectories for each birth cohort, as well as more realistic counterfactual estimates. Incorporating multivariable risk algorithms for different health states could also improve the accuracy and reliability of estimates since behaviour change and healthcare costs are often associated with experiencing certain health states. For example, treatment of various cancers can cost anywhere from $12,000 to $44,000 in Ontario, and those who survive engage in limited physical activity and spend a large proportion of their day sedentary.[154,155] Once more advanced methods are developed, they could be potentially utilized to generate projections that can be validated with future measures of actual deaths and healthcare costs.

4.6 TABLES

Table 4.1: Definitions of Healthy Behaviour Exposures

Unhealthy Behaviours	Definition
Smoking	
Heavy	≥ Pack/day daily smoking within last 4 years
Light	< Pack/day daily smoking within last 4 years
Recent Heavy	≥ Pack/day daily smoking 4 to less than 20 years ago
Recent Light	< Pack/day daily smoking 4 to less than 20 years ago
Non-Recent Heavy	≥ Pack/day daily smoking 20 or more years ago
Non-Recent Light	< Pack/day daily smoking 20 or more years ago
Non-Smoker	Never Daily Smoker
Leisure Physical Inactivity	
High	0 to < 0.5 Daily averaged MET
Moderate	0.5 to < 2.5 Daily averaged MET
Low	≥ 2.5 Daily averaged MET
Non-active Transport	
High	0 to < 2 weekly hours of active transport
Moderate	2 to < 6 weekly hours of active transport
Low	≥ 6 daily weekly hours of active transport
Leisure Sedentary Activity	
High	≥ 5.5 weekly hours
Moderate	2.5 to < 5.5 weekly hours
Low	0 to < 2.5 weekly hours
Unhealthy Diet Score	
Extremely Inadequate	≥ 7 Points
Less than Inadequate	3 to <7 Points
Adequate	0 to <3 Points

Table 4.2: Description of Regular Phase Cohort

Exposures	Respondents	Prevalence	Healthcare Costs per Person-Month ($)¥
CCHS Cycle			
1.1 (2000-2001)	18,357	13.1	257
2.1 (2003-2004)	19,433	13.8	240
3.1 (2005-2006)	19,222	13.7	260
2007-2008	21,315	15.2	261
2009-2010	20,494	14.6	261
2011-2012	20,356	14.5	260
2013-2014	21,244	15.1	250
Sex			
Female	77,803	55.4	289
Male	62,618	44.6	267
Age Groups			
40-44	17,313	12.3	107
45-49	15,253	10.9	120
50-54	17,182	12.2	149
55-59	18,752	13.4	179
60-64	18,245	13.0	235
65-69	16,510	11.8	336
70-74	13,808	9.8	430
75-79	11,438	8.1	533
80-84	7,959	5.7	639
85-89	3,961	2.8	783
Smoking			
Heavy	9,539	6.8	384
Light	21,820	15.5	300
Recent Heavy	6,301	4.5	330
Recent Light	11,146	7.9	264
Non-Recent Heavy	7,781	5.5	285
Non-Recent Light	17,855	12.7	233
Non-Smoker	65,981	47.0	227
Leisure Physical Inactivity			
High	33,400	23.8	348
Moderate	65,521	46.7	239
Low	41,500	29.6	198
Non-active Transport			
High	57,961	41.3	321
Moderate	43,679	31.1	221
Low	38,782	27.6	185
Leisure Sedentary Activity			
High	45,912	32.7	345
Moderate	54,038	38.5	229

Low	40,470	28.8	199
Unhealthy Diet Score			
Extremely Inadequate	33,721	24.0	332
Less than Adequate	69,537	49.5	243
Adequate	37,164	26.5	210
Education			
Less than High school	31,186	22.2	320
High school	34,310	24.4	247
Post-secondary	74,925	53.4	234
Immigrant Status			
Yes	31,734	22.6	238
No	108,687	77.4	260
Body Mass Index			
Morbidly Obese (>=40)	4,416	3.1	505
Underweight (<20)	4,301	3.1	475
Other (20 to <40)	131,704	93.8	240
Sense of Belonging			
Very Weak	12,643	9.0	420
Not Very Weak	127,778	91.0	239
Marital Status			
Single/Divorced	53,448	38.1	302
Married/Common-Law	86,973	61.9	228
Home Ownership			
No	27,998	19.9	367
Yes	112,423	80.1	228
Food Insecurity			
Insecure	12,464	8.9	446
Secure	127,958	91.1	237
Neighborhood Income			
Lowest Quintile	27,001	19.2	313
Non-lowest Quintile	113,420	80.8	241
Rurality			
Rural	25,940	18.5	249
Suburban	45,145	32.1	256
Urban	69,336	49.4	257
Degree of Immobility			
Need Help with Tasks	19,134	13.6	611
Some Limitations	30,944	22.0	304
No Physical Limitations	90,342	64.3	165
Level of Illness			
Severely Ill	17,677	12.6	774
Moderately Ill	52,719	37.5	278
Relatively Healthy	70,025	49.9	110

¥ Age-sex standardized to CCHS 1.1 Respondents (Except for Sex and Age Group Exposures)

Table 4.3: Description of Near End-of-Life Phase Cohort

Exposures	Respondents	Prevalence	Healthcare Costs per Person-Month ($)[¥]
CCHS Cycle			
1.1 (2000-2001)	929	14.0	1,510
2.1 (2003-2004)	948	14.3	1,670
3.1 (2005-2006)	841	12.7	1,760
2007-2008	988	14.9	1,670
2009-2010	927	14.0	1,950
2011-2012	935	14.1	1,870
2013-2014	1,053	15.9	1,770
Sex			
Female	3,147	47.5	1,730
Male	3,474	52.5	1,780
Age Groups			
40-44	70	1.1	1,540
45-49	135	2.0	1,250
50-54	233	3.5	2,000
55-59	362	5.5	1,850
60-64	531	8.0	1,650
65-69	748	11.3	1,910
70-74	970	14.7	1,700
75-79	1,282	19.4	1,670
80-84	1,307	19.7	1,740
85-89	983	14.8	1,890
Smoking			
Heavy	615	9.3	1,540
Light	1,096	16.6	1,650
Recent Heavy	419	6.3	1,960
Recent Light	486	7.3	1,670
Non-Recent Heavy	637	9.6	2,000
Non-Recent Light	1,060	16.0	1,780
Non-Smoker	2,308	34.9	1,820
Leisure Physical Inactivity			
High	2,865	43.3	2,050
Moderate	2,638	39.8	1,570
Low	1,118	16.9	1,420
Non-active Transport			
High	3,991	60.3	2,130
Moderate	1,705	25.8	1,260
Low	925	14.0	1,070
Leisure Sedentary Activity			
High	3,334	50.4	2,130
Moderate	2,284	34.5	1,450

Low	1,004	15.2	1,300
Unhealthy Diet Score			
Extremely Inadequate	2,601	39.3	2,070
Less than Adequate	2,997	45.3	1,540
Adequate	1,023	15.4	1,560
Education			
Less than High school	2,698	40.7	1,730
High school	1,502	22.7	1,870
Post-secondary	2,421	36.6	1,700
Immigrant Status			
Yes	1,486	22.4	1,810
No	5,135	77.6	1,730
Body Mass Index			
Morbidly Obese (>=40)	256	3.9	2,340
Underweight (<20)	641	9.7	2,840
Other (20 to <40)	5,724	86.5	1,620
Sense of Belonging			
Very Weak	1,077	16.3	2,430
Not Very Weak	5,544	83.7	1,610
Marital Status			
Single/Divorced	3,474	52.5	1,690
Married/Common-Law	3,148	47.5	1,770
Home Ownership			
No	2,024	30.6	1,960
Yes	4,597	69.4	1,660
Food Insecurity			
Insecure	660	10.0	2,230
Secure	5,961	90.0	1,680
Neighborhood Income			
Lowest Quintile	1,661	25.1	1,870
Non-lowest Quintile	4,960	74.9	1,720
Rurality			
Rural	1,336	20.2	1,760
Suburban	2,265	34.2	1,680
Urban	3,020	45.6	1,790
Degree of Immobility			
Need Help with Tasks	2,705	40.9	2,500
Some Limitations	1,601	24.2	1,580
No Physical Limitations	2,315	35.0	1,070
Level of Illness			
Severely Ill	3,469	52.4	2,560
Moderately Ill	2,155	32.5	1,080
Relatively Healthy	997	15.1	690

[¥] Age-sex standardized to CCHS 1.1 Respondents (Except for Sex and Age Group Exposures)

Table 4.4: Description of End-of-Life Phase Cohort

Exposures	Respondents	Prevalence	Healthcare Costs per Person-Month ($)[¥]
CCHS Cycle			
1.1 (2000-2001)	714	14.1	5,290
2.1 (2003-2004)	708	14.0	5,510
3.1 (2005-2006)	661	13.1	5,230
2007-2008	771	15.2	5,880
2009-2010	721	14.2	6,320
2011-2012	722	14.3	6,030
2013-2014	764	15.1	5,440
Sex			
Female	2,374	46.9	5,940
Male	2,687	53.1	5,480
Age Groups			
40-44	57	1.1	4,800
45-49	114	2.3	4,520
50-54	180	3.6	5,110
55-59	280	5.5	5,770
60-64	390	7.7	6,120
65-69	559	11.0	6,450
70-74	738	14.6	5,550
75-79	1,000	19.8	5,780
80-84	975	19.3	5,510
85-89	768	15.2	5,530
Smoking			
Heavy	485	9.6	5,070
Light	834	16.5	5,580
Recent Heavy	337	6.7	6,100
Recent Light	361	7.1	5,700
Non-Recent Heavy	508	10.0	5,720
Non-Recent Light	790	15.6	5,820
Non-Smoker	1,748	34.5	5,640
Leisure Physical Inactivity			
High	2,297	45.4	6,130
Moderate	1,963	38.8	5,560
Low	801	15.8	4,860
Non-active Transport			
High	3,135	62.0	6,390
Moderate	1,256	24.8	4,800
Low	670	13.2	4,290
Leisure Sedentary Activity			
High	2,542	50.2	6,320
Moderate	1,765	34.9	5,150

Low	755	14.9	5,090
Unhealthy Diet Score			
Extremely Inadequate	2,120	41.9	5,800
Less than Adequate	2,201	43.5	5,640
Adequate	740	14.6	5,450
Education			
Less than High school	2,055	40.6	5,810
High school	1,165	23.0	5,630
Post-secondary	1,841	36.4	5,520
Immigrant Status			
Yes	1,155	22.8	5,780
No	3,906	77.2	5,690
Body Mass Index			
Morbidly Obese (>=40)	202	4.0	6,790
Underweight (<20)	565	11.2	6,300
Other (20 to <40)	4,295	84.9	5,580
Sense of Belonging			
Very Weak	936	18.5	6,190
Not Very Weak	4,125	81.5	5,590
Marital Status			
Single/Divorced	2,651	52.4	5,700
Married/Common-Law	2,410	47.6	5,670
Home Ownership			
No	1,564	30.9	6,090
Yes	3,497	69.1	5,520
Food Insecurity			
Insecure	521	10.3	6,940
Secure	4,540	89.7	5,630
Neighborhood Income			
Lowest Quintile	1,274	25.2	5,770
Non-lowest Quintile	3,787	74.8	5,690
Rurality			
Rural	1,016	20.1	5,470
Suburban	1,724	34.1	5,310
Urban	2,321	45.9	6,100
Degree of Immobility			
Need Help with Tasks	2,261	44.7	6,150
Some Limitations	1,167	23.1	5,740
No Physical Limitations	1,633	32.3	5,020
Level of Illness			
Severely Ill	3,073	60.7	6,880
Moderately Ill	1,351	26.7	4,200
Relatively Healthy	637	12.6	3,330

¥ Age-sex standardized to CCHS 1.1 Respondents (Except for Sex and Age Group Exposures)

Table 4.5: Cost Ratios of Categorical Exposures in Regular Life Risk Algorithms

Exposures	Algorithms			
	Base	Distal	Intermediate	Proximal
Sex				
Female	1.06 (1.04, 1.08)	1.00 (0.98, 1.02)	0.95 (0.93, 0.97)	1.03 (1.02, 1.05)
Male	Ref.	Ref.	Ref.	Ref.
Smoking				
Heavy	1.46 (1.41, 1.51)	1.27 (1.23, 1.32)	1.19 (1.15, 1.23)	1.08 (1.04, 1.11)
Light	1.26 (1.23, 1.29)	1.15 (1.12, 1.18)	1.11 (1.08, 1.14)	1.03 (1.01, 1.06)
Recent Heavy	1.30 (1.25, 1.35)	1.26 (1.21, 1.31)	1.18 (1.14, 1.23)	1.07 (1.03, 1.12)
Recent Light	1.11 (1.07, 1.16)	1.11 (1.08, 1.15)	1.08 (1.05, 1.12)	1.05 (1.02, 1.08)
Non-Recent Heavy	1.20 (1.15, 1.24)	1.19 (1.14, 1.23)	1.15 (1.11, 1.19)	1.10 (1.06, 1.14)
Non-Recent Light	1.01 (0.98, 1.04)	1.02 (0.99, 1.05)	1.02 (0.99, 1.05)	1.02 (0.99, 1.04)
Non-Smoker	Ref.	Ref.	Ref.	Ref.
Leisure Physical Inactivity				
High	1.41 (1.38, 1.45)	1.29 (1.26, 1.32)	1.12 (1.10, 1.15)	1.08 (1.06, 1.10)
Moderate	1.09 (1.07, 1.11)	1.06 (1.04, 1.08)	1.00 (0.99, 1.02)	0.99 (0.98, 1.01)
Low	Ref.	Ref.	Ref.	Ref.
Non-active Transport				
High	1.51 (1.45, 1.57)	1.45 (1.39, 1.51)	1.30 (1.25, 1.35)	1.19 (1.15, 1.23)
Moderate	1.14 (1.11, 1.18)	1.11 (1.08, 1.15)	1.06 (1.03, 1.10)	1.02 (0.98, 1.07)
Low	Ref.	Ref.	Ref.	Ref.
Sedentary Activity				
High	1.71 (1.65, 1.77)	1.53 (1.48, 1.59)	1.33 (1.28, 1.37)	1.20 (1.16, 1.24)
Moderate	1.19 (1.16, 1.22)	1.15 (1.11, 1.18)	1.09 (1.06, 1.12)	1.06 (1.03, 1.09)
Low	Ref.	Ref.	Ref.	Ref.
Unhealthy Diet Score				
Extremely Inadequate	1.32 (1.28, 1.35)	1.14 (1.11, 1.17)	1.10 (1.07, 1.12)	1.07 (1.04, 1.09)
Less than adequate	1.06 (1.04, 1.08)	1.01 (1.00, 1.03)	1.01 (1.00, 1.03)	1.01 (0.99, 1.03)
Adequate	Ref.	Ref.	Ref.	Ref.
Education				
Less than High school		1.10 (1.08, 1.13)	1.09 (1.07, 1.12)	1.07 (1.05, 1.09)
High school		0.99 (0.97, 1.01)	1.01 (0.99, 1.02)	0.99 (0.97, 1.01)
Post-secondary		Ref.	Ref.	Ref.
Immigrant Status				
Yes		0.96 (0.95, 0.98)	1.00 (0.98, 1.02)	1.01 (0.99, 1.02)
No		Ref.	Ref.	Ref.
Body Mass Index				
Morbidly Obese (>=40)		1.70 (1.62, 1.78)	1.44 (1.38, 1.51)	1.35 (1.30, 1.41)

Underweight (<20)	1.48 (1.41, 1.56)	1.36 (1.30, 1.43)	1.24 (1.18, 1.29)
Other (20 to <40)	Ref.	Ref.	Ref.
Sense of Belonging			
Very Weak	1.30 (1.26, 1.34)	1.17 (1.14, 1.20)	1.12 (1.09, 1.15)
Not Very Weak	Ref.	Ref.	Ref.
Marital Status			
Single/Divorced	1.09 (1.07, 1.11)	1.08 (1.06, 1.10)	1.04 (1.02, 1.05)
Married/Common-Law	Ref.	Ref.	Ref.
Home Ownership			
No	1.25 (1.22, 1.28)	1.19 (1.17, 1.22)	1.14 (1.12, 1.17)
Yes	Ref.	Ref.	Ref.
Food Insecurity			
Insecure	1.60 (1.55, 1.65)	1.29 (1.25, 1.33)	1.17 (1.14, 1.21)
Secure	Ref.	Ref.	Ref.
Neighborhood Income			
Lowest Quintile	1.06 (1.04, 1.08)	1.04 (1.02, 1.06)	1.04 (1.02, 1.06)
Non-lowest Quintile	Ref.	Ref.	Ref.
Rurality			
Rural	0.93 (0.91, 0.95)	0.91 (0.89, 0.93)	0.95 (0.93, 0.97)
Suburban	0.99 (0.98, 1.01)	0.98 (0.96, 0.99)	0.98 (0.96, 0.99)
Urban	Ref.	Ref.	Ref.
Immobility at Age 60[¥]			
Need Help with Tasks		2.83 (2.75, 2.90)	2.05 (2.00, 2.10)
Physical Limitations		1.73 (1.70, 1.76)	1.44 (1.42, 1.47)
No Physical Limitations		Ref.	Ref.
Immobility at Age 80[¥]			
Need Help with Tasks		2.13 (2.05, 2.21)	1.79 (1.73, 1.85)
Physical Limitations		1.52 (1.47, 1.57)	1.37 (1.32, 1.41)
No Physical Limitations		Ref.	Ref.
Illness at Age 60[¥]			
Severely Ill			5.79 (5.64, 5.94)
Moderately Ill			2.33 (2.30, 2.37)
Relatively Healthy			Ref.
Illness at Age 80[¥]			
Severely Ill			3.68 (3.55, 3.82)
Moderately Ill			1.86 (1.80, 1.92)
Relatively Healthy			Ref.

[¥] **Exposure estimated at specific age because of age interaction**

Table 4.6: Cost Ratios of Categorical Exposures in Near End-of-Life Risk Algorithms

Exposures	Algorithms			
	Base	Distal	Intermediate	Proximal
Sex				
Female	0.96 (0.89, 1.04)	0.93 (0.86, 1.02)	0.89 (0.82, 0.97)	0.94 (0.87, 1.02)
Male	Ref.	Ref.	Ref.	Ref.
Smoking				
Heavy	0.77 (0.66, 0.89)	0.75 (0.64, 0.87)	0.75 (0.65, 0.87)	0.76 (0.66, 0.88)
Light	0.86 (0.76, 0.97)	0.83 (0.73, 0.93)	0.84 (0.74, 0.95)	0.81 (0.72, 0.92)
Recent Heavy	0.93 (0.78, 1.10)	0.93 (0.78, 1.11)	0.88 (0.74, 1.04)	0.79 (0.67, 0.93)
Recent Light	0.86 (0.73, 1.01)	0.84 (0.71, 0.99)	0.85 (0.72, 1.00)	0.84 (0.72, 0.99)
Non-Recent Heavy	1.08 (0.93, 1.26)	1.09 (0.94, 1.27)	1.03 (0.89, 1.19)	1.01 (0.87, 1.16)
Non-Recent Light	0.95 (0.84, 1.07)	0.97 (0.86, 1.09)	0.96 (0.85, 1.08)	0.97 (0.87, 1.09)
Non-Smoker	Ref.	Ref.	Ref.	Ref.
Leisure Physical Inactivity				
High	1.29 (1.14, 1.47)	1.28 (1.13, 1.45)	1.17 (1.04, 1.33)	1.10 (0.98, 1.24)
Moderate	1.09 (0.97, 1.23)	1.09 (0.97, 1.23)	1.04 (0.93, 1.17)	0.98 (0.87, 1.10)
Low	Ref.	Ref.	Ref.	Ref.
Non-active Transport				
High	1.91 (1.51, 2.40)	1.85 (1.48, 2.32)	1.62 (1.30, 2.01)	1.47 (1.19, 1.81)
Moderate	1.24 (1.00, 1.53)	1.24 (1.01, 1.53)	1.18 (0.97, 1.42)	1.09 (0.91, 1.31)
Low	Ref.	Ref.	Ref.	Ref.
Sedentary Activity				
High	1.47 (1.30, 1.67)	1.45 (1.28, 1.65)	1.32 (1.16, 1.49)	1.27 (1.12, 1.45)
Moderate	1.12 (0.96, 1.31)	1.12 (0.96, 1.31)	1.05 (0.88, 1.24)	1.07 (0.89, 1.28)
Low	Ref.	Ref.	Ref.	Ref.
Unhealthy Diet Score				
Extremely Inadequate	1.31 (1.17, 1.48)	1.20 (1.06, 1.35)	1.14 (1.01, 1.28)	1.13 (1.00, 1.27)
Less than adequate	1.11 (0.99, 1.24)	1.09 (0.97, 1.22)	1.06 (0.94, 1.18)	1.02 (0.91, 1.14)
Adequate	Ref.	Ref.	Ref.	Ref.
Education				
Less than High school		1.04 (0.95, 1.15)	1.04 (0.95, 1.14)	1.01 (0.93, 1.11)
High school		1.06 (0.95, 1.17)	1.04 (0.94, 1.16)	1.03 (0.93, 1.13)
Post-secondary		Ref.	Ref.	Ref.
Immigrant Status				
Yes		1.04 (0.94, 1.15)	1.06 (0.96, 1.17)	1.04 (0.94, 1.14)
No		Ref.	Ref.	Ref.
Body Mass Index				
Morbidly Obese (>=40)		1.16 (0.94, 1.42)	1.11 (0.90, 1.37)	1.05 (0.86, 1.28)

Underweight (<20)	1.46 (1.26, 1.67)	1.40 (1.22, 1.62)	1.25 (1.08, 1.45)
Other (20 to <40)	Ref.	Ref.	Ref.
Sense of Belonging			
Very Weak	1.10 (0.98, 1.24)	1.01 (0.90, 1.14)	1.02 (0.91, 1.14)
Not Very Weak	Ref.	Ref.	Ref.
Marital Status			
Single/Divorced	0.94 (0.86, 1.02)	0.96 (0.88, 1.05)	0.97 (0.89, 1.05)
Married/Common-Law	Ref.	Ref.	Ref.
Home Ownership			
No	1.13 (1.03, 1.25)	1.12 (1.02, 1.23)	1.16 (1.06, 1.27)
Yes	Ref.	Ref.	Ref.
Food Insecurity			
Insecure	1.17 (1.00, 1.38)	1.04 (0.88, 1.22)	0.96 (0.82, 1.12)
Secure	Ref.	Ref.	Ref.
Neighborhood Income			
Lowest Quintile	0.99 (0.90, 1.08)	0.96 (0.88, 1.05)	0.99 (0.91, 1.09)
Non-lowest Quintile	Ref.	Ref.	Ref.
Rurality			
Rural	1.00 (0.90, 1.11)	1.00 (0.90, 1.11)	1.02 (0.92, 1.13)
Suburban	1.02 (0.93, 1.12)	1.01 (0.93, 1.10)	1.04 (0.95, 1.13)
Urban	Ref.	Ref.	Ref.
Immobility at Age 60¥			
Need Help with Tasks		2.32 (1.99, 2.70)	1.75 (1.50, 2.04)
Physical Limitations		1.61 (1.45, 1.80)	1.44 (1.30, 1.61)
No Physical Limitations		Ref.	Ref.
Immobility at Age 80¥			
Need Help with Tasks		1.88 (1.61, 2.19)	1.42 (1.22, 1.65)
Physical Limitations		1.15 (1.02, 1.31)	1.06 (0.94, 1.20)
No Physical Limitations		Ref.	Ref.
Illness at Age 60¥			
Severely Ill			4.34 (3.70, 5.09)
Moderately Ill			3.01 (2.64, 3.44))
Relatively Healthy			Ref.
Illness at Age 80¥			
Severely Ill			1.68 (1.42, 2.00)
Moderately Ill			1.54 (1.34, 1.77)
Relatively Healthy			Ref.

¥ Exposure estimated at specific age because of age interaction

Table 4.7: Cost Ratios of Categorical Exposures in End-of-Life Risk Algorithms

	Algorithms			
Exposures	**Base**	**Distal**	**Intermediate**	**Proximal**
Sex				
Female	1.11 (1.02, 1.20)	1.09 (1.00, 1.18)	1.08 (0.99, 1.18)	1.10 (1.01, 1.19)
Male	Ref.	Ref.	Ref.	Ref.
Smoking				
Heavy	0.76 (0.65, 0.88)	0.75 (0.65, 0.87)	0.75 (0.65, 0.87)	0.76 (0.66, 0.88)
Light	0.87 (0.77, 0.98)	0.87 (0.77, 0.98)	0.87 (0.78, 0.99)	0.88 (0.78, 0.99)
Recent Heavy	0.96 (0.82, 1.14)	0.95 (0.80, 1.12)	0.94 (0.80, 1.12)	0.88 (0.75, 1.04)
Recent Light	0.92 (0.78, 1.08)	0.92 (0.78, 1.08)	0.92 (0.78, 1.08)	0.93 (0.79, 1.09)
Non-Recent Heavy	1.13 (0.96, 1.31)	1.14 (0.98, 1.34)	1.14 (0.97, 1.33)	1.09 (0.94, 1.27)
Non-Recent Light	0.99 (0.88, 1.12)	1.00 (0.88, 1.13)	0.99 (0.88, 1.12)	1.00 (0.89, 1.13)
Non-Smoker	Ref.	Ref.	Ref.	Ref.
Leisure Physical Inactivity				
High	1.19 (1.06, 1.33)	1.16 (1.03, 1.30)	1.14 (1.01, 1.28)	1.14 (1.02, 1.29)
Moderate	1.07 (0.95, 1.20)	1.06 (0.94, 1.19)	1.05 (0.93, 1.18)	1.04 (0.93, 1.16)
Low	Ref.	Ref.	Ref.	Ref.
Non-active Transport				
High	1.50 (1.31, 1.72)	1.47 (1.29, 1.68)	1.45 (1.27, 1.66)	1.36 (1.18, 1.56)
Moderate	1.14 (0.96, 1.35)	1.13 (0.96, 1.34)	1.13 (0.96, 1.33)	1.08 (0.91, 1.28)
Low	Ref.	Ref.	Ref.	Ref.
Sedentary Activity				
High	1.31 (1.15, 1.50)	1.30 (1.13, 1.49)	1.28 (1.12, 1.47)	1.26 (1.09, 1.45)
Moderate	1.08 (0.95, 1.24)	1.08 (0.95, 1.24)	1.07 (0.94, 1.23)	1.06 (0.93, 1.21)
Low	Ref.	Ref.	Ref.	Ref.
Unhealthy Diet Score				
Extremely Inadequate	1.08 (0.96, 1.22)	1.04 (0.92, 1.18)	1.04 (0.92, 1.17)	1.05 (0.93, 1.19)
Less than adequate	1.09 (0.97, 1.23)	1.10 (0.98, 1.23)	1.09 (0.97, 1.22)	1.08 (0.97, 1.22)
Adequate	Ref.	Ref.	Ref.	Ref.
Education				
Less than High school		1.04 (0.95, 1.14)	1.05 (0.95, 1.15)	1.05 (0.96, 1.14)
High school		1.01 (0.91, 1.12)	1.01 (0.91, 1.12)	1.00 (0.90, 1.11)
Post-secondary		Ref.	Ref.	Ref.
Immigrant Status				
Yes		0.98 (0.89, 1.07)	0.98 (0.89, 1.08)	0.97 (0.88, 1.06)
No		Ref.	Ref.	Ref.
Body Mass Index				
Morbidly Obese (>=40)		1.16 (0.95, 1.42)	1.14 (0.93, 1.39)	1.20 (0.99, 1.47)

Underweight (<20)	1.16 (1.01, 1.32)	1.14 (1.00, 1.31)	1.10 (0.96, 1.25)
Other (20 to <40)	Ref.	Ref.	Ref.
Sense of Belonging			
Very Weak	1.04 (0.94, 1.16)	1.03 (0.93, 1.15)	1.00 (0.90, 1.12)
Not Very Weak	Ref.	Ref.	Ref.
Marital Status			
Single/Divorced	0.96 (0.88, 1.04)	0.96 (0.88, 1.05)	0.97 (0.90, 1.06)
Married/Common-Law	Ref.	Ref.	Ref.
Home Ownership			
No	1.08 (0.98, 1.18)	1.07 (0.98, 1.17)	1.09 (0.99, 1.19)
Yes	Ref.	Ref.	Ref.
Food Insecurity			
Insecure	1.08 (0.93, 1.24)	1.05 (0.91, 1.21)	1.04 (0.91, 1.20)
Secure	Ref.	Ref.	Ref.
Neighborhood Income			
Lowest Quintile	1.02 (0.93, 1.11)	1.01 (0.93, 1.11)	1.05 (0.96, 1.15)
Non-lowest Quintile	Ref.	Ref.	Ref.
Rurality			
Rural	0.91 (0.83, 1.01)	0.92 (0.83, 1.02)	0.92 (0.83, 1.02)
Suburban	0.89 (0.81, 0.97)	0.89 (0.82, 0.97)	0.91 (0.84, 0.99)
Urban	Ref.	Ref.	Ref.
Immobility at Age 60[¥]			
Need Help with Tasks		1.16 (0.99, 1.34)	0.95 (0.81, 1.10)
Physical Limitations		1.10 (0.94, 1.28)	0.92 (0.79, 1.08)
No Physical Limitations		Ref.	Ref.
Immobility at Age 80[¥]			
Need Help with Tasks		1.08 (0.97, 1.21)	1.03 (0.93, 1.15)
Physical Limitations		1.06 (0.93, 1.20)	1.04 (0.92, 1.18)
No Physical Limitations		Ref.	Ref.
Illness at Age 60[¥]			
Severely Ill			3.04 (2.58, 3.58)
Moderately Ill			1.63 (1.35, 1.96)
Relatively Healthy			Ref.
Illness at Age 80[¥]			
Severely Ill			2.14 (1.86, 2.46)
Moderately Ill			1.40 (1.21, 1.63)
Relatively Healthy			Ref.

[¥] **Exposure estimated at specific age because of age interaction**

Table 4.8: Characteristics of application cohorts: 2000-2004, 2005-2009, and 2013-2014

Exposures	Prevalence by Era		
	2000-2004	2005-2009	2010-2014
Sex			
Female	52.1	51.0	51.8
Male	47.9	49.0	48.2
Age Groups			
40-44	19.7	17.9	14.8
45-49	17.1	16.7	14.7
50-54	14.7	14.6	15.8
55-59	12.3	13.5	14.1
60-64	9.6	11.0	12.1
65-69	8.4	8.6	9.8
70-74	7.5	6.6	7.0
75-79	5.7	5.6	5.7
80-84	3.4	3.8	3.8
85-89	1.6	1.7	2.1
Smoking			
Heavy	7.3	5.4	4.3
Light	16.5	15.4	13.9
Recent Heavy	4.5	4.1	3.2
Recent Light	8.2	8.2	7.6
Non-Recent Heavy	4.4	4.2	4.2
Non-Recent Light	10.2	10.8	11.3
Non-Smoker	48.8	51.9	55.4
Leisure Physical Inactivity			
Inactive	27.5	24.9	24.1
Moderately Inactive	47.1	47.1	45.3
Active	25.4	28.0	30.6
Non-active Transport			
High	44.7	40.2	37.6
Moderate	31.6	31.9	29.3
Low	23.7	27.9	33.2
Leisure Sedentary Activity			
Sedentary	22.5	24.5	30.9
Moderately Sedentary	37.3	37.8	38.3
Non-Sedentary	40.2	37.7	30.9
Unhealthy Diet Score			
Extremely Inadequate	24.5	21.9	20.6
Less than adequate	51.8	49.8	48.5
Adequate	23.7	28.3	30.9
Education			
Less than High school	23.3	18.2	14.0

High school	26.9	23.5	23.3
Post-secondary	49.8	58.3	62.7
Immigrant Status			
Yes	34.3	36.3	36.6
No	65.7	63.7	63.4
Body Mass Index			
Morbidly Obese (>=40)	1.9	2.6	3.3
Underweight (<20)	3.6	3.3	3.4
Other (20 to <40)	94.5	94.2	93.4
Sense of Belonging			
Very Weak	19.6	17.5	20.6
Not Very Weak	80.4	82.5	79.4
Marital Status			
Single/Divorced	24.7	25.3	26.9
Married/Common-Law	75.3	74.7	73.1
Home Ownership			
No	19.6	17.5	20.6
Yes	80.4	82.5	79.4
Food Insecurity			
Insecure	10.2	9.2	8.9
Secure	89.8	90.8	91.1
Neighborhood Income			
Lowest Quintile	17.8	17.8	17.7
Non-lowest Quintile	82.2	82.2	82.3
Rurality			
Rural	9.4	9.1	8.7
Suburban	21.9	21.3	21.7
Urban	68.8	69.6	69.6
Degree of Immobility			
Need Help with Tasks	11.2	12.0	12.4
Some Limitations	19.7	18.7	18.6
No Physical Limitations	69.1	69.3	69.0
Level of Illness			
Severely Ill	10.4	11.2	11.6
Moderately Ill	33.8	34.1	35.8
Relatively Healthy	55.7	54.7	52.7

Table 4.9: Cost Burden Estimates for CCHS 2000-2004, 2005-2009, and 2013-2014

Scenario	Lifetime Costs ($)	Costs Per Life Year	% Change in Lifetime Costs
2000-2004			
Baseline	205,600	4,886	-
No Smoking	217,049 (215,550 - 219,207)	4,874 (4,824 - 4,923)	5.6 (5.0, 6.4)
No Leisure Physical Inactivity	207,331 (202,257 - 211,189)	4,676 (4,531 - 4,850)	0.8 (-1.5, 2.5)
No Non-Active Transport	189,689 (185,405 - 195,397)	4,244 (4,139 - 4,464)	-7.7 (-9.7, -5.1)
No Leisure Sedentary Activity	186,372 (180,405 - 195,397)	4,301 (4,152 - 4,491)	-9.4 (-12.1, -6.8)
No Poor Diet	209,842 (206,631 - 211,057)	4,810 (4,677 - 4,860)	2.1 (0.6, 2.5)
No Unhealthy Behaviours	184,593 (168,969 - 200,831)	3,685 (3,327 - 4,164)	-10.2 (-17.7, -2.5)
2005-2009			
Baseline	223,276	5,180	-
No Smoking	234,014 (232,641 - 236,238)	5,160 (5,113 - 5,228)	4.8 (4.3, 5.7)
No Leisure Physical Inactivity	224,526 (219,209 - 228,572)	4,968 (4,823 - 5,143)	0.6 (-1.7, 2.2)
No Non-Active Transport	206,397 (201,209 - 212,424)	4,530 (4,421 - 4,751)	-7.6 (-9.5, -5.0)
No Leisure Sedentary Activity	200,779 (193,985 - 207,373)	4,523 (4,354 - 4,736)	-10.1 (-13.0, -7.3)
No Poor Adequate Diet	227,320 (223,904 - 228,591)	5,099 (4,967 - 5,150)	1.8 (0.4, 2.2)
No Unhealthy Behaviours	197,313 (180,752 - 214,709)	3,900 (3,528 - 4,395)	-11.6 (-18.9, -4.0)
2010-2014			
Baseline	237,391	5,366	-
No Smoking	247,191 (245,835 - 249,260)	5,339 (5,295 - 5,400)	4.1 (3.7, 4.9)
No Leisure Physical Inactivity	238,168 (232,711 - 242,312)	5,152 (5,008 - 5,323)	0.3 (-1.9. 2.0)
No Non-Active Transport	219,695 (215,143 - 225,774)	4,717 (4,611 - 4,932)	-7.5 (-9.3, -5.0)
No Leisure Sedentary Activity	211,088 (203,519 - 218,584)	4,626 (4,446 - 4,861)	-11.1 (-14.2, -9.0)
No Poor Diet	241,020 (237,431 - 242,337)	5,281 (5,151 - 5,330)	1.5 (0.1, 2.0)
No Unhealthy Behaviours	206,785 (190, 257 - 224,392)	4,031 (3,668 - 4,518)	-12.9 (-19.8, -5.6)

Table 4.10: Sensitivity Analysis of Cost Burden Estimates for CCHS 2000-2004, 2005-2009, and 2013-2014

Scenario	Lifetime Costs ($)	Costs Per Life Year	% Change in Lifetime Costs
2000-2004			
Baseline	205,600	4,886	-
No Smoking	216,169 (214,745 - 218,174)	4,854 (4,806 - 4,919)	5.1 (4.6, 5.9)
No Leisure Physical Inactivity	210,716 (206,211 - 213,763)	4,752 (4,619 - 4,910)	2.5 (0.4, 3.8)
No Non-Active Transport	199,147 (195,422 - 203,681)	4,455 (4,362 - 4,653)	-3.1 (-4.8, -1.1)
No Leisure Sedentary Activity	191,605 (186,286 - 196,655)	4,422 (4,285 - 4,599)	-6.8 (-9.3, -4.5)
No Poor Diet	212,375 (209,784 - 213,183)	4,868 (4,748 - 4,909)	3.3 (2.2, 3.5)
No Unhealthy Behaviours	195,073 (179,319 - 210,835)	3,894 (3,531 - 4,371)	-5.1 (-12.7, 2.4)
2005-2009			
Baseline	223,276	5,180	-
No Smoking	233,163 (231,869 - 235,213)	5,141 (5,096 - 5,205)	4.4 (3.9, 5.2)
No Leisure Physical Inactivity	227,701 (222,920 - 230,942)	5,039 (4,905 - 5,197)	2.0 (-0.1, 3.3)
No Non-Active Transport	215,676 (211,702 - 220,558)	4,734 (4,636 - 4,933)	-3.4 (-5.1, -1.4)
No Leisure Sedentary Activity	206,453 (200,225 - 212,392)	4,651 (4,494 - 4,850)	-7.5 (-10.2, -5.0)
No Poor Adequate Diet	229,786 (226,985 - 230,658)	5,154 (5,035 - 5,196)	2.9 (1.8, 3.2)
No Unhealthy Behaviours	207,484 (190,741 - 224,459)	4,101 (3,723 - 4,595)	-7.1 (-14.5, 0.4)
2010-2014			
Baseline	237,391	5,366	-
No Smoking	246,331 (245,036 - 248,245)	5,320 (5,278 - 5,378)	3.8 (3.3, 4.5)
No Leisure Physical Inactivity	241,231 (236,275 - 244,584)	5,218 (5,084 - 5,373)	1.6 (-0.4. 2.9)
No Non-Active Transport	228,972 (224,944 - 233,951)	4,916 (4,821 – 5,110)	-3.5 (-5.1, -1.6)
No Leisure Sedentary Activity	217,513 (210,502 - 224,347)	4,767 (4,598 – 4,989)	-8.4 (-11.2, -5.6)
No Poor Diet	243,413 (240,320 - 244,335)	5,333 (5,216 - 5,374)	2.5 (1.4, 2.8)
No Unhealthy Behaviours	216,655 (199,807 - 234,056)	4,223 (3,852 - 4,712)	-8.7 (-15.7, -1.5)

5. Assessment of Impact of Unhealthy Behaviours on Transitions towards Immobility using Linked Cross-sectional National Health Surveys

5.1 ABSTRACT

Background

Utilizing cross-sectional national health surveys to assess the impact of unhealthy behaviours on survey measures of disability can be challenging. This study applied imputation models to successive iterations of a linked national health survey to generate quasi-longitudinal exposures that could be utilized to assess the impact of unhealthy behaviours (smoking, leisure physical inactivity, non-active transport, leisure sedentary activity, and poor diet) on transitions towards immobility.

Data and Methods

Linked Ontario subsamples of the Canadian Community Health Survey (CCHS) roughly 4 years apart (i.e., base and target cohorts) were utilized to approximate transition cohorts for 5 different time periods (CCHS 1.1(2000-2001) to CCHS 3.1(2005-2006), CCHS 2.1 (2003-2004) to CCHS 2007-2008, CCHS 3.1 (2005-2006) to CCHS 2009-2010, and CCHS 2011-2012) Respondents who did not survive until the end of their 4-year window were excluded. The five cohorts were aggregated together, and fully conditional specification methods were utilized to generate missing survey information (e.g., health behaviours, sociodemographic indicators, immobility etc.) either at the beginning or end of each respondent's four-year window. A final transition cohort was constructed utilizing the period-specific transition cohorts demonstrating comparable exposure distributions (Cohen's $d \leq 0.10$) between their base and target cohorts. Three immobility-specific cohorts were derived from this cohort using baseline values, from which multivariable logistic models were constructed for transitions from none to moderate immobility, and from moderate to severe immobility.

Results

The base and target cohorts of the four more recent period-specific transition cohorts demonstrated comparable pre- and post-period exposure distributions (Cohen's $d \leq 0.10$). In the earliest period, the majority of the pre- and post-period exposure distributions were not

comparable between the base and target cohort, potentially the product of fundamental differences between the cohorts (e.g., different education levels). This period-specific cohort was excluded from the final transition cohort. Transitions towards moderate immobility were predicted by all unhealthy behaviours except poor diet, and transitions towards severe immobility were predicted by all unhealthy behaviours except sedentary activity.

Interpretation

The impact of unhealthy behaviours on transitions towards immobility can be evaluated utilizing quasi-longitudinal cohorts generated by combining multiple imputation with repeated iterations of national health surveys. These preliminary methods however do not consider that some individuals will become ill and transition out of the household population over time. Future research should examine the incorporation of linked healthcare administrative data that could help adjust for expected differences in the sampling frames over time. The resulting multivariable transition models could be incorporated into cohort life tables or microsimulations.

5.2 INTRODUCTION

Assessing the population health effects of unhealthy behaviours can be a challenge. National health surveys that are predominantly cross-sectional have been our main source of health behaviour surveillance. The lack of population-based longitudinal information prevents researchers from answering many important questions in a straightforward and robust manner. As a result, estimates of life expectancy, lifetime morbidity or disability, and lifetime healthcare costs are often grounded on period life tables. This can be problematic because not all birth cohorts have the same experience across their lifetime.[156] For example, younger cohorts are less likely to be smokers than older cohorts, while at the same time more likely to be sedentary. Some researchers have attempted to address these challenges by deriving age-specific transition probabilities through comparisons of birth cohorts across different iterations of a national health survey.[157–159] This is often a crude and tedious endeavour because adjusting for multiple factors simultaneously involves extensive stratification of the birth cohorts. At a certain point, the data becomes too sparse to stratify further. The linking of these surveys to administrative healthcare databases has provided access to longitudinal information beyond what is collected in the surveys. Unfortunately, because of the challenge with survey length and responder burden,

comprehensive longitudinal data is routinely lacking in population health surveys for health behaviours, sociodemographic factors, and measures of disability. Estimation of transition probabilities for these exposures will still have to largely depend on repeated iterations of household health surveys for the foreseeable future. More sophisticated methods have arisen to address these challenges, but they still generally only produce matrices of transitional probabilities rather than comprehensive multivariable transition models.

Multiple imputation has the potential to transform two separate survey administrations into a quasi-longitudinal cohort where respondents have two sets of values rather than one pertinent to their immediate survey administration. In this scenario, the two survey administrations encapsulate an observation window where the earlier survey (i.e., base cohort) provides responses at the beginning of the observation window, and the later survey (i.e., target cohort) provides responses at the end of the observation window. This implies that we need to impute pre-period values for the base cohort and post-period values for the target cohort. In some cases, exposures can be assumed to not change over time (e.g., immigrant status, education level at older ages), which reduces the unknowns. In other cases, time-dependent content from the actual survey may provide the missing values. For example, former smokers in many health survey are routinely asked how long since they have quit smoking, enabling us to infer their smoking status in prior years.

The number of unknowns can be further reduced by focusing only on net transitions (i.e., only allowing transitions to occur in one direction). After all, it is the net transitions that are of primary interest and could be utilized in cohort life tables or microsimulations to generate lifetime estimates. For many exposures, there is a life course trajectory.[160–162] For example, people’s weight tends to increase throughout middle life and then decreases toward the end of life.[163,164] Whereas, physical activities can fluctuate but decline overall. For other exposures, there are also limited by irreversibility. For example, respondents who report to be physically limited due to a chronic condition may maintain their present level of debilitation or worsen. This also suggests that this same respondent in the past was healthier or at worst, at the same level of debilitation than they are presently reporting. Taking these details into account, respondents can only project into certain levels of health behaviours or health states going forward or backward in time.

This study used utilized the linked CCHS to estimate the impact of unhealthy behaviours on transitions towards immobility. This involved (1) the construction of five hypothetical transition cohorts using linked CCHS approximately 4 years apart, (2) aggregation of the hypothetical transitions cohorts into a preliminary transition cohort and completion of missing values by inference where possible, (3) imputation of remaining missing values utilizing imputation models (i.e., each post-period exposure imputed utilizing all pre-period exposures, and each pre-period exposure imputed utilizing all post-period exposures), (4) construction of a final transition cohort using period-specific transition cohorts demonstrating comparable values between their base and target cohorts, (5) division of the final transition cohort into three cohorts using their pre-period levels of immobility (none, moderate, or severe), and (6) construction of multivariable transition models for transitions from none to moderate immobility, and from moderate to severe immobility.

5.3 DATA & METHODS

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian population. The survey employs a complex multistage sampling strategy and is conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions and collective dwellings. The details of the survey methodology have been previously published and are available online.[9] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content were linked to the Registered Person Database (RPDB) at ICES. This population-based registry contains all unique health numbers issued for the Ontario Insurance Health Plan (OHIP), and their corresponding period of coverage. Each individual record is augmented by ICES with additional information which includes date of birth, sex, date of death, yearly postal codes, geographic indicators, and date of last contact with the healthcare system. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. These groups are also excluded from the CCHS

sampling frame. The linked Ontario subsamples from 1.1 (2000-2001), 2.1 (2003-04), 3.1(2005-06), and 2007 to 2014 were combined to create an initial linked CCHS cohort of 233,835 respondents for this study.

Health Services & Utilization Data

Multiple types of administrative data were linked to the linked CCHS surveys to facilitate imputation of missing survey information and ascertainment of healthcare costs. This included records from the OHIP database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), and Ontario Mental Health Reporting System (OMHRS) database, and Home Care Database (HCD). The OHIP database contains claims to Ontario's publicly funded health insurance system by health care practitioners for inpatient and outpatient services. The ODB database contains prescription claims covered by the provincial drug program, which is restricted to those aged 65 and older, LTC residents, or service recipients of home care, social assistance and special drug programs. The CCRS database contains information for episodes of complex continuing care within hospitals and long-term care facilities, where individuals receive 24-hour nursing services. The DAD contains information for all admissions and transfers to inpatient care within acute, rehabilitation, and chronic care institutions. The OMHRS database contains information for all admissions to adult inpatient hospital care beds designated for mental health. The HCD is a clinical client centric database that captures all services that are provided by or coordinated by Community Care Access Centres.

Analytical Cohort

An analytical cohort was constructed from the linked CCHS cohort for the construction of the transition cohorts. This cohort was constructed in a similar manner to prior work (see Chapter 3), the only difference was that this cohort excluded respondents 95 years of age and older rather than respondents 90 years of age and older. Briefly, this cohort consisted of 161,905 respondents who were followed up to a maximum of five years or death, and censored at either the beginning of any time period of 6 months or longer where they were OHIP ineligible. The unhealthy behaviour exposures are provided in Table 5.1, and ascertainment of these unhealthy behaviours by the CCHS can be found in prior work (Chapter 3). The only difference was poor diet being categorized by three levels in this chapter rather than the four levels in prior work (i.e.,

the two intermediate levels of diet are collapsed into one). The other risk factors that were included in the risk algorithms and auxiliary information utilized for multiple imputation are also described in the prior work. Markov Chain Monte Carlo methods were utilized to address missing information due to survey non-response or differences in the survey modules between cycles, resulting in 10 imputed datasets.

Construction of Hypothetical Transition Cohorts

Hypothetical transition cohorts were constructed utilizing the linked CCHS cycles approximately 4 years apart within the analytical cohort: CCHS 1.1(2000-2001) to CCHS 3.1(2005-2006), CCHS 2.1 (2003-2004) to CCHS 2007-2008, CCHS 3.1 (2005-2006) to CCHS 2009-2010, and CCHS 2011-2012. In each period, the earlier cycle represented the 'base' cohort for the period of transition supplying pre-period exposures, while the later cycle represented the 'target' cohort of transition supplying post-period exposures (See Figure 1). For the base cohorts, their age at the end of the period was simply four years more than their base cohort age, and for the target cohorts, their age at the beginning of the period was simply four years less than their target cohort age. The cohorts were restricted to individuals who were 40 to 89 years of age at the beginning of the transition period. Individuals in base cohorts that were not alive and eligible for OHIP four years later were excluded. Individuals in target cohorts that were not OHIP eligible four years prior were excluded. Transitions for many exposures within each of the five hypothetical transition cohorts were likely to be rare. In a similar manner to studies that aggregate multiple cycles of the CCHS for analysis of linked outcomes, these five hypothetical transition cohorts were aggregated into an overall transition cohort representing transitions across different four year periods.[47,99,111]

Risk Factors of Interest and Outcomes

Unhealthy behaviours were the main exposures of interest to be included in the transition algorithms for immobility. Descriptions of the unhealthy behaviour exposures are provided in Table 5.1. The CCHS has an extensive number of questions which evaluate present and past smoking behavior. Smoking behavior was defined by combining survey responses regarding smoking status, daily cigarette consumption, and years since having quit daily smoking consumption. The CCHS assesses participation in numerous different types of leisure physical activities in the three months prior to the survey. For each activity, an average daily expenditure

is calculated using the frequency, average duration, and energy cost associated with the activity. Summing the average daily expenditures of all the ascertained leisure activities provides an overall average daily expenditure in metabolic equivalent of task (MET) units. Leisure physical inactivity was defined by categorizing this aggregated summary measure. The CCHS uses several questions to evaluate the amount of time walking or cycling while commuting or conducting errands. Non-active transport was defined by categorizing the total weekly hours associated with these activities. The CCHS assesses participation in different types of leisure sedentary activity (i.e., time on the computer, playing video games, or reading) in the three months prior to the survey. Leisure sedentary activity was defined by categorizing the total weekly hours spent in these activities. Dietary quality is difficult to assess using health surveys. The CCHS captures fruit and vegetable consumption through a series of questions that evaluate frequency of intake rather than the amount consumed or number of servings. Frequencies are initially captured either per day, week, month, or year, but then subsequently converted to daily estimates of consumption frequency. The different dietary items examined in the CCHS include fruit, salad, carrot, potato, juice, and other vegetables. A simple unhealthy diet score derived from these dietary indicators in prior research (Chapter 3) was used to define different levels of unhealthy diet. Immobility was assessed by examining a respondent's self-reported inability to perform basic tasks (severe immobility), as well as their self-reported inability to participate in everyday activities due to a long-term health condition (moderate immobility).

Additional risk factors that were selected for to be included in the transition models for immobility included age, calendar date, sociodemographic factors, and neighbourhood indicators. These risk factors had been included in prior algorithms for mortality and healthcare costs (See Chapter 3 & 4). The sociodemographic factors included were educational level, immigrant status, body mass index, sense of belonging, marital status, home ownership and food insecurity. Neighbourhood indicators for each respondent were supplemented to the cohort through linkages to the Registered Person's Database, which provided the postal code of each respondent in the year of their survey administration. The postal code was utilized in conjunction with postal code conversion files to identify whether the neighbourhood was low income (i.e. lowest quintile), and the rurality index of Ontario to identify whether the neighbourhood was urban, suburban, or rural.[128,129]

Ascertainment of Missing Transition Information

Missing values in the hypothetical transition cohort were inferred where possible. Sociodemographic factors such as educational level, immigrant status, body mass index, sense of belonging, and food insecurity were considered static over time (i.e., static exposures). In these cases, they were either expected to be static because of the ages being analyzed or they involved exposure levels that were relatively rare (<10% prevalence). All other exposures were allowed to transition across observation periods (i.e., transitioning exposures) Low neighbourhood income and rurality was available at the beginning and end of each respondent's designated observation period through linkage. Marital status and home ownership were only allowed to decrease across transition periods, reflecting natural trends at these ages. To operationalize this with regard to marital status meant designating individuals who were not married/common-law at their pre-period exposure to be the same at their post-period exposure, and designating individuals who were married/common-law at their post-period exposure to be the same at their pre-period exposure. As a result, it was only possible for married/common-law individuals to become single/divorced/widowed over time. Home ownership was operationalized in the same manner.

Several three level exposures were also operationalized to transition in a single direction, but only across adjacent levels (e.g., no immobility to moderate immobility, or moderate immobility to severe immobility, but not no immobility to severe immobility). To facilitate this, two binary exposures were created (i.e., severe immobility vs. none or moderate immobility, and severe or moderate immobility vs. no immobility). In a similar manner to marital status and home ownership, potential transitions were inferred where possible using these binary exposures. For example, someone who had no immobility issues at their pre-period exposure could either have no immobility or moderate immobility at their post-period exposure, but not severe immobility. Transitions for several unhealthy behaviours (leisure physical inactivity, inactive transport, leisure sedentary activity, and poor diet) characterized by three levels utilized the same implementation. In these instances, health behaviours were operationalized to only deteriorate with age, which reflects the natural trends observed in the cohorts. For diet and physical activity, exposures were considered static for non-seniors reflecting the lack of an observed trend at younger ages.

The ascertainment of smoking status transitions was a special case because past smoking history was available for individuals in the target cohorts. This implied that not only was their post-period exposure available, but also the pre-period exposure if some basic assumptions were followed. Given that the youngest individuals in the transition cohort were 40 years of age at pre-exposure, it was assumed that smokers could only quit and non-smokers would not initiate smoking. In addition, former smoking categories were characterized by individuals who had at least quit for four years or more, since smokers are regularly attempting to quit but frequently relapsing. This was the reasoning behind defining 'smokers' (see Table 5.1) as smokers and former smokers with less than 4 years since quitting. These assumptions also implied that the only individuals in the base cohorts whose post smoking exposure was unknown were smokers who could remain smokers or become recent former smokers. Based on these assumptions, the pre and post smoking exposures were implemented using indicators for intensity (heavy, light, none) and several binary exposures (non-smoker/current smoker vs. former smokers, non-smoker/current/very recent former smokers (4 to <12 years cessation) vs. non-very recent former smokers (≥12 years cessation), and non-smoker/current/recent smoker vs. non-recent former smoker).

Fully conditional specification methods (also called multivariate imputation by chained equations) were utilized to impute the remaining transition information that could not be inferred. A pre and post imputation model was specified for each of the indicators utilized to operationalize the exposures of interest that were unknown (See Figure 2). In the case of smoking, only a single pre and post binary indicator was unknown (non-smoker/current smoker vs. former smokers). Pre imputation models utilized post indicators, and post imputation models utilized pre indicators. In the case of smoking, interactions between intensity and the binary indicators were also included. For all imputation models, age was included with a beta spline at 75 years of age, survey date was included to adjust for differences in transition over time. This process generated 100 imputed datasets (10 for each imputed baseline dataset). The comparability of the base and target cohorts (overall and by time period) was assessed by examining pre and post exposures to be included in the transition models. An exposure that demonstrated a standardized difference ≤0.1 was considered comparable.

Transition Models

The final transition cohort was split into three cohorts based on the pre-exposure levels of immobility: none, moderate, and severe immobility. A logistic model was constructed assessing transitions from no to moderate immobility using the unweighted healthy cohort, and a logistic model was constructed assessing transitions from moderate immobility to severe immobility using the unweighted moderate immobility cohort. Each transition model included survey date (as years from earliest survey administered), age at pre-period (with a beta spline at 75 years of age), cohort type (base vs. target cohort), sex, unhealthy behaviours (see Table 5.1), and the sociodemographic factors. Backward selection was utilized to select the final models. The relative time of the survey administration and cohort type were retained regardless of statistical significance, with the remaining exposures only retained if a statistical significance of $p<0.20$ was achieved. This lower than traditionally utilized p-value cutoff was implemented to examine whether the association of weaker predictors was also consistent with expectations. The models were re-derived utilizing weights given the exposures selected in the unweighted models. Given the novelty of these methods, an association was considered statistically significant at $p<0.10$. All statistical analysis was performed utilizing SAS Enterprise Guide 7.1.

5.4 RESULTS

Most of the period-specific transition cohorts demonstrated comparable exposures between their base and target cohorts (See Tables 5.2 to 5.6). The one clear exception was the first period-specific cohort where multiple unhealthy behaviours and sociodemographic factors were inconsistent either in the unweighted or weighted cohorts. In general, the base cohort had lower education levels than the target cohort, an exposure which was considered static across the period. Given this discrepancy and others (e.g., lower home ownership in the base cohort), many imputed unhealthy behaviours were imputed as more prevalent in the base than the target cohort. The other period-specific transition cohorts demonstrated predominantly comparable exposures, with the exception of non-active transport in the period-specific transition cohort from CCHS 3.1 to CCHS 2009-2010). The earliest period-specific transition cohort was excluded from the final overall transition cohort (See Table 5.7). The overall base and target cohorts were comparable in the final transition cohort across all exposures.

The final transition cohort was divided into three cohorts depending on their pre-immobility level which are described in tables 5.8 to 5.10. Consistent with expectations, unhealthy behaviours and factors associated with lower socioeconomic status were more common as we moved from the no immobility cohort to the moderate immobility cohort, and finally to severe immobility cohort. The models for transitions towards moderate immobility utilizing the no immobility cohort is presented in Table 5.11. Significant dose-responses were observed between almost all unhealthy behaviours and the likelihood of transitioning to moderate immobility in both the unweighted and weighted models. The one exception was poor diet which was associated with a higher likelihood of transition (p-value < 0.20) but not at significant p-value < 0.10 threshold. The models for transitions towards severe immobility utilizing the moderate immobility cohort is presented in Table 5.12. Significant dose-responses were observed between almost all unhealthy behaviours and the likelihood of transitioning to moderate immobility in both the unweighted and weighted models. The one exception was sedentary activity which was not retained in the model (p-value > 0.20). Sociodemographic factors in both transition models demonstrated relationships to be expected, with the unhealthier levels exhibiting higher likelihoods of transitioning to more immobile levels. The only unexpected result occurred in the model for severe immobility where immigrants demonstrated a much higher risk for transition. There were no statistical differences (>0.5) between the base and target cohorts in any of the models.

5.5 DISCUSSION

This study applied imputation methods to repeated iterations of the CCHS, a cross-sectional national health survey, to generate a quasi-longitudinal dataset that could provide transitions across 4-year intervals. Hypothetical period-specific transition cohorts were constructed utilizing linked Ontario subsamples of the CCHS approximately 4 years apart (i.e., base and target cohorts). These cohorts were aggregated into an overall transition cohort and missing values were inferred where possible. The remaining missing values for the target cohorts (i.e., at the beginning of the 4-year periods) and for the base cohorts (i.e., at the end of the 4-year periods) were generated through imputation models. In general, the period-specific transition cohorts demonstrated comparable distributions between their base and target cohorts. The one exception demonstrated fundamental sociodemographic differences between its base and target

cohort, resulting in its exclusion from the final overall transition cohort. Multivariable transition models for no to moderate immobility, and moderate to severe immobility were constructed utilizing the final transition cohort. Both types of transitions were predicted by multiple unhealthy behaviours and various sociodemographic factors. The final transitional cohort that was created in this study can be further utilized to generate transition models for other exposures (i.e., health behaviours) that were allowed to transition across its 4-year intervals.

The framework used in this study provides a novel means by which to generate multivariable transition probabilities from repeated cross-sectional national health surveys. Longitudinal population-based national health surveys are uncommon, which implies that longitudinal measures of the various exposures (e.g., health behaviours, sociodemographic factors, and health quality measures) generally captured by these surveys are difficult to obtain. In contrast, cross-sectional national health surveys are routinely administered at regular intervals within many countries. The linkage of national health surveys to healthcare administrative data has provided opportunities to assess the relation between the various exposures captures in these surveys and longitudinal outcomes. Nevertheless, these linked sources typically do not contain longitudinal measures of the exposures captured within the cross-sectional national health surveys. As a result, researchers often resort to unsophisticated methods such as the derivation of period life tables or the estimation of basic transition probabilities by comparing birth cohorts across repeated iterations of a national health survey. The former method is not realistic since period life tables assume that different birth cohorts will experience the same age-specific exposures across their lifetimes. The later method typically involves limited adjustment for covariates since adjustment is performed through stratification. Adjustment for multiple covariates is often tedious and impractical. The methodology presented in the current study can provide multivariable transitional models that are more realistic than these historical methods.

Future research should attempt to further this framework through linkages to healthcare administrative data. Although these data do not normally contain longitudinal measures of the exposures of interest captured within the national health surveys, they contain alternative longitudinal information that can help. For example, these databases can be utilized to ascertain linked health exposures (e.g., disease status) at both the beginning and end of an interval or even cumulative measures across an interval (e.g., utilization or healthcare costs). This can be utilized

to improve the specificity of the transitions of survey exposures by either including this information in the imputation models or by stratifying the imputation models on different longitudinal information. Moreover, these improvements can address a fundamental problem with the traditional crude methods and the framework introduced within this study, that these national household surveys do not involve a closed household population. Although the surveys often target the household population, it is unrealistic to assume that all respondents will remain in the household population moving forward. Some of these respondents will transition to retirement homes and long-term care facilities, making them ineligible for future iterations of the survey. The inclusion of linked longitudinal information provides opportunities to address for differences in the sampling frames over time. Without these adjustments, many estimated transitions that coincide with poor health are likely to be underestimated and lead to erroneous conclusions when incorporated into cohort life tables or microsimulations. The linkage of longitudinal national health surveys, which are less common, may also provide opportunities to properly assess these methodologies.

5.6 FIGURES

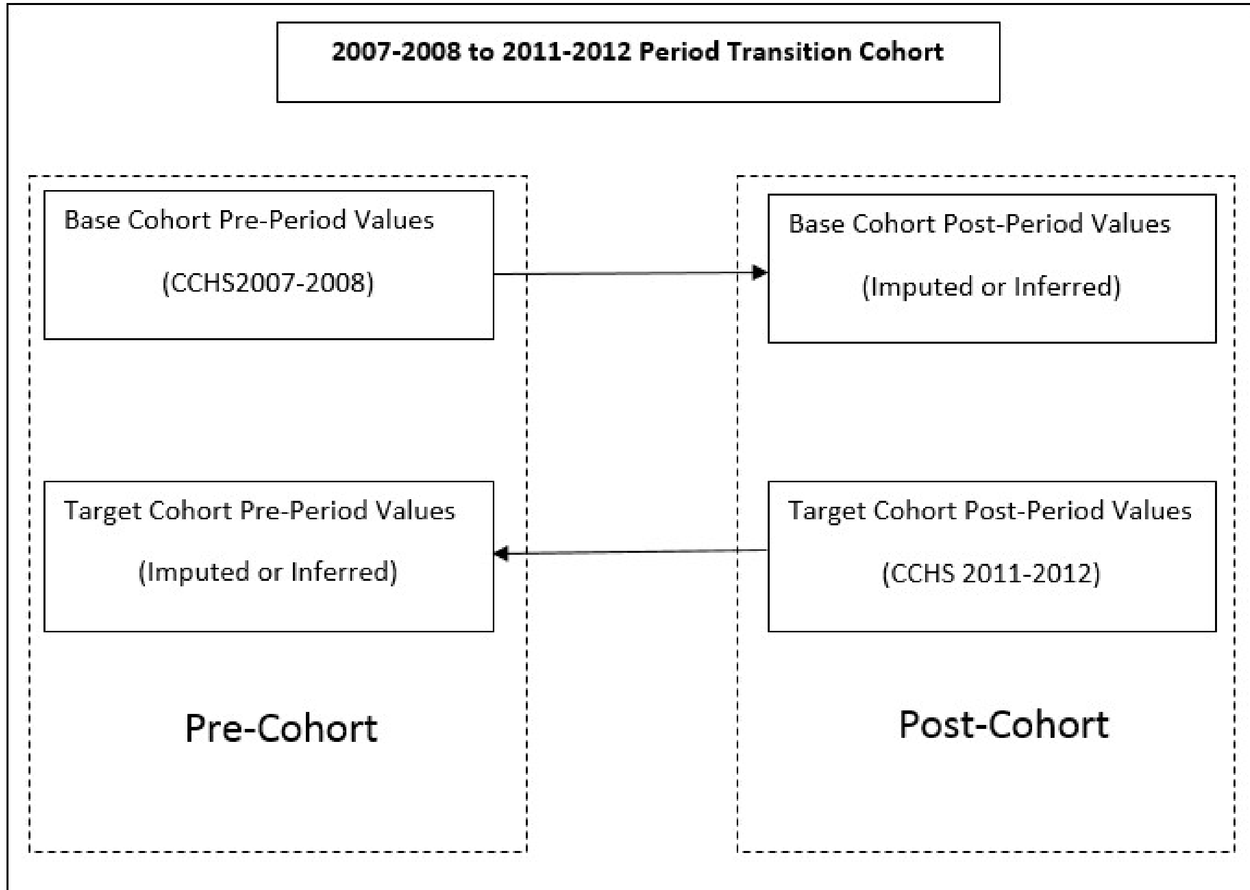

Figure 1: Period Transition Cohort Example

$$\begin{aligned}\boldsymbol{PostImmo1} = {} & \boldsymbol{PreImmo1 + PreImmo2 + PreSmokInt + PreSmokRec1 +} \\ & \boldsymbol{PreSmokRec2 + PreSmokRec3 + PreSmokInt * PreSmokRec1 +} \\ & \boldsymbol{PreSmokInt * PreSmokRec2 + PreSmokInt * PreSmokRec3 +} \\ & \boldsymbol{PreInactivity1 + PreInactivity2 + PreNonactive1 +} \\ & \boldsymbol{PreNonactive2 + PreSedent1 + PreSedent2 + PreDiet1 +} \\ & \boldsymbol{PreDiet2 + Education + Immigrant + BMI + Belonging +} \\ & \boldsymbol{FoodInsecure + PreMarital + PreHome + PreNIncome +} \\ & \boldsymbol{PreRural + Surveydate + PreAge + PreAge75}\end{aligned}$$

$$\begin{aligned}\boldsymbol{PreImmo1} = {} & \boldsymbol{PostImmo1 + PostImmo2 + PostSmokInt + PostSmokRec1} \\ & \boldsymbol{+ PostSmokRec2 + PostSmokRec3 + PostSmokInt * PostSmokRec1} \\ & \boldsymbol{+ PostSmokInt * PostSmokRec2 + PostSmokInt * PostSmokRec3} \\ & \boldsymbol{+ PostInactivity1 + PostInactivity2 + PostNonactive1} \\ & \boldsymbol{+ PostNonactive2 + PostSedent1 + PostSedent2 + PostDiet1} \\ & \boldsymbol{+ PostDiet2 + Education + Immigrant + BMI + Belonging} \\ & \boldsymbol{+ FoodInsecure + PostMarital + PostHome + PostNIncome} \\ & \boldsymbol{+ PostRural + Surveydate + PostAge + PostAge75}\end{aligned}$$

Definitions of Exposures defined by more than one Parameter:

Immo1: Severe vs. Moderate/No Immobility

Immo2: Severe/Moderate vs. No Immobility

SmokInt: Smoking Intensity (Heavy, Light, or None)

SmokRec1: Non-/Current vs. Former Smokers

SmokRec2: Non-/Current /Very Recent Former vs. Recent/Non-Recent Former Smokers

SmokRec3: Non-/Current /Very Recent/Recent Former vs. Non-Recent Former Smokers

Inactive1: High vs. Moderate/Low Leisure Physical Inactivity

Inactive2: High/Moderate vs. Low Leisure Physical Inactivity

Nonactive1: High vs. Moderate/Low Non-active Transport

Nonactive2: High/Moderate vs. Low Non-active Transport

Sedent1: High vs. Moderate/Low Non-active Transport

Sedent2: High/Moderate vs. Low Non-active Transport

Diet1: Extremely Inadequate vs. Less than adequate/Adequate Diet

Diet2: Extremely/Less than Adequate vs. Adequate Diet

Figure 2: Fully Conditional Specification Example for Ascertainment of Severe Immobility

5.7 TABLES

Table 5.1: Definitions of Healthy Behaviour Exposures

Unhealthy Behaviours	Definition
Smoking	
Heavy	≥ Pack/day daily smoking within last 4 years
Light	< Pack/day daily smoking within last 4 years
Recent Heavy	≥ Pack/day daily smoking 4 to less than 20 years ago
Recent Light	< Pack/day daily smoking 4 to less than 20 years ago
Non-Recent Heavy	≥ Pack/day daily smoking 20 or more years ago
Non-Recent Light	< Pack/day daily smoking 20 or more years ago
Non-Smoker	Never Daily Smoker
Leisure Physical Inactivity	
High	0 to < 0.5 Daily averaged MET
Moderate	0.5 to < 2.5 Daily averaged MET
Low	≥ 2.5 Daily averaged MET
Non-active Transport	
High	0 to < 2 weekly hours of active transport
Moderate	2 to < 6 weekly hours of active transport
Low	≥ 6 daily weekly hours of active transport
Leisure Sedentary Activity	
High	≥ 5.5 weekly hours
Moderate	2.5 to < 5.5 weekly hours
Low	0 to < 2.5 weekly hours
Unhealthy Diet Score	
Extremely Inadequate	≥ 7 Points
Less than Inadequate	3 to <7 Points
Adequate	0 to <3 Points

Table 5.2: Exposures of 1st Transition Cohort (CCHS 1.1to CCHS 3.1) ¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base	Target	Base	Target
	(N=17,344)	(N=17,157)	(N=4,817,501)	(N=4,787,721)
Immobility Transitions				
Severe to Severe	10.8	10.2	9.8	8.9
Moderate to Severe	5.2	4.6	4.6	3.7
Moderate to Moderate	16.7	16.1	14.8	14.2
None to Moderate	5.5	4.8	5.5	4.6
None to None	61.8	64.3	65.3	68.5
Pre Immobility				
Severe	10.8	10.2	9.8	8.9
Moderate	21.9	20.7	19.4	17.9
None	67.3	69.1	70.8	73.1
Post Immobility				
Severe	16.0	14.8	14.4	12.7
Moderate	22.2	20.9	20.3	18.8
None	61.8	64.3	65.3	68.5
Pre Smoking				
Heavy	9.1	9.6	7.4	7.9
Light	17.2	17.9	17.0	16.0
Recent Heavy	4.7	5.3	4.2	4.9
Recent Light	8.0	8.9	8.4	8.9
Non-Recent Heavy	5.2	4.3	4.4	3.6
Non-Recent Light	10.9	9.5	10.0	9.3
Non-Smoker	44.9	44.5	48.7	49.4
Post Smoking				
Heavy	7.6	8.0	6.2	6.6
Light	15.1	15.8	14.8	13.9
Recent Heavy	4.9	5.2	4.3	4.7
Recent Light	8.3	7.9	8.7	7.9
Non-Recent Heavy	6.5	5.9	5.5	5.2
Non-Recent Light	12.6	12.6	11.9	12.4
Non-Smoker	44.9	44.5	48.7	49.4
Pre Leisure Inactivity				
High	26.1	23.0	***28.0***	***23.8***
Moderate	47.3	47.2	48.4	47.9
Low	26.7	29.8	***23.5***	***28.3***
Post Leisure Inactivity				
High	27.7	24.6	***29.2***	***24.9***
Moderate	47.1	47.1	48.1	47.7
Low	25.2	28.3	***22.7***	***27.4***
Pre Non-active Transport				
High	***47.3***	***41.5***	***47.2***	***40.9***

Moderate	28.8	32.3	30.1	33.3
Low	23.9	26.2	22.7	25.8
Post Non-active Transport				
High	***48.6***	***42.9***	***48.4***	***42.0***
Moderate	29.3	32.8	30.3	33.7
Low	22.1	24.3	21.4	24.2
Pre Sedentary Activity				
High	25.0	25.2	22.1	20.8
Moderate	38.3	36.6	37.5	34.9
Low	36.7	38.3	40.5	44.4
Post Sedentary Activity				
High	31.2	31.4	28.0	26.5
Moderate	38.9	37.0	38.7	36.2
Low	29.8	31.6	33.3	37.3
Pre Unhealthy Diet				
Extremely Inadequate	***28.3***	***23.8***	25.7	22.1
Less than Adequate	51.5	49.4	51.9	49.3
Adequate	***20.2***	***26.8***	***22.4***	***28.6***
Post Unhealthy Diet				
Extremely Inadequate	***28.9***	***24.4***	26.2	22.6
Less than Adequate	51.2	49.2	51.6	49.1
Adequate	***19.9***	***26.5***	***22.2***	***28.3***
Education				
Less than High school	28.5	25.6	23.5	21.7
High school	27.0	24.1	***28.5***	***24.0***
Post-secondary	***44.6***	***50.3***	***47.9***	***54.4***
Immigrant Status				
Yes	22.0	22.3	33.5	34.8
No	78.0	77.7	66.5	65.2
Body Mass Index				
Morbidly Obese (>=40)	2.0	2.7	1.6	2.2
Underweight (<20)	3.6	2.9	3.6	3.2
Other (20 to <40)	94.4	94.3	94.8	94.6
Sense of Belonging				
Very Weak	10.4	9.4	12.5	10.7
Not Very Weak	89.6	90.6	87.5	89.3
Food Insecurity				
Insecure	***11.3***	***8.5***	10.8	8.0
Secure	***88.7***	***91.5***	89.2	92.0
Pre Marital Status				
Single/Divorced	35.3	35.7	24.7	21.6
Married/Common-Law	64.7	64.3	75.3	78.4
Post Marital Status				
Single/Divorced	39.9	40.1	***30.2***	***25.8***
Married/Common-Law	60.1	59.9	***69.8***	***74.2***

Pre Home Ownership				
No	22.3	19.8	***21.3***	***15.5***
Yes	77.7	80.2	***78.7***	***84.5***
Post Home Ownership				
No	23.0	20.5	***21.9***	***16.0***
Yes	77.0	79.5	***78.1***	***84.0***
Pre Neighborhood Income				
Lowest Quintile	19.5	19.4	18.3	16.4
Non-lowest Quintile	80.5	80.6	81.7	83.6
Post Neighborhood Income				
Lowest Quintile	20.0	20.5	18.2	18.1
Non-lowest Quintile	80.0	79.5	81.8	81.9
Pre Rurality				
Rural	18.4	20.1	9.3	9.8
Suburban	34.7	30.6	21.4	21.5
Urban	47.0	49.3	69.3	68.7
Post Rurality				
Rural	18.7	20.1	9.6	9.5
Suburban	34.8	31.8	22.3	22.0
Urban	46.5	48.1	68.1	68.5

¥ Age-sex standardized to CCHS 1.1 Respondents
Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.3: Exposures of 2nd Transition Cohort (CCHS 2.1 to CCHS 2007-2008) ¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base (N=18,399)	Target (N=19,558)	Base (N=5,227,415)	Target (N=5,166,536)
Immobility Transitions				
Severe to Severe	11.4	11.9	9.7	11.1
Moderate to Severe	4.3	4.4	3.6	4.0
Moderate to Moderate	17.6	18.0	15.8	15.6
None to Moderate	4.9	4.8	4.8	4.5
None to None	61.8	60.9	66.0	64.9
Pre Immobility				
Severe	11.4	11.9	9.7	11.1
Moderate	21.9	22.4	19.5	19.5
None	66.7	65.7	70.8	69.4
Post Immobility				
Severe	15.7	16.3	13.3	15.0
Moderate	22.5	22.8	20.7	20.1
None	61.8	60.9	66.0	64.9
Pre Smoking				
Heavy	8.6	7.7	7.4	6.1
Light	17.2	17.9	16.1	16.7
Recent Heavy	5.2	5.0	4.7	4.2
Recent Light	8.1	8.9	8.2	8.6
Non-Recent Heavy	4.6	4.1	3.9	3.2
Non-Recent Light	10.4	9.8	9.7	9.1
Non-Smoker	45.9	46.5	50.3	52.0
Post Smoking				
Heavy	7.2	6.3	5.8	4.9
Light	15.1	15.7	14.1	14.4
Recent Heavy	5.5	4.8	5.0	3.9
Recent Light	8.7	8.0	8.8	8.0
Non-Recent Heavy	5.7	5.7	4.9	4.8
Non-Recent Light	11.8	12.9	11.1	12.0
Non-Smoker	45.9	46.5	50.3	52.0
Pre Leisure Inactivity				
High	23.5	24.1	25.2	25.0
Moderate	47.6	47.0	46.9	47.4
Low	28.9	28.8	27.8	27.5
Post Leisure Inactivity				
High	24.9	25.5	26.3	26.1
Moderate	47.4	46.8	46.7	47.1
Low	27.6	27.7	27.0	26.8
Pre Non-active Transport				
High	41.0	39.9	41.1	39.1

Moderate	34.1	31.6	33.6	31.6
Low	24.8	28.5	25.4	29.4
Post Non-active Transport				
High	42.9	41.6	42.6	40.5
Moderate	33.7	31.4	33.1	31.3
Low	23.4	27.0	24.2	28.1
Pre Sedentary Activity				
High	24.9	23.9	21.0	19.9
Moderate	38.1	40.1	36.9	38.2
Low	37.0	36.0	42.0	41.9
Post Sedentary Activity				
High	31.4	30.1	27.2	25.7
Moderate	39.0	41.6	38.6	40.6
Low	29.6	28.3	34.2	33.7
Pre Unhealthy Diet				
Extremely Inadequate	24.7	23.9	22.1	21.7
Less than Adequate	50.9	49.4	52.1	49.5
Adequate	24.4	26.7	25.8	28.9
Post Unhealthy Diet				
Extremely Inadequate	25.3	24.6	22.6	22.3
Less than Adequate	50.7	49.1	51.9	49.2
Adequate	24.0	26.3	25.5	28.5
Education				
Less than High school	*24.4*	*22.7*	21.2	20.4
High school	*25.9*	*24.1*	25.9	24.1
Post-secondary	*49.6*	*53.2*	52.9	55.5
Immigrant Status				
Yes	22.9	22.7	35.3	35.9
No	77.1	77.3	64.7	64.1
Body Mass Index				
Morbidly Obese (>=40)	2.5	3.2	2.1	2.7
Underweight (<20)	2.7	3.5	3.0	3.8
Other (20 to <40)	94.8	93.3	94.9	93.5
Sense of Belonging				
Very Weak	9.4	8.8	11.0	11.2
Not Very Weak	90.6	91.2	89.0	88.8
Food Insecurity				
Insecure	*10.1*	*8.0*	9.2	9.1
Secure	*89.9*	*92.0*	90.8	90.9
Pre Marital Status				
Single/Divorced	36.4	36.2	23.3	22.8
Married/Common-Law	63.6	63.8	76.7	77.2
Post Marital Status				
Single/Divorced	40.5	40.0	28.2	26.8
Married/Common-Law	59.5	60.0	71.8	73.2

Pre Home Ownership				
No	19.7	19.2	17.2	16.8
Yes	80.3	80.8	82.8	83.2
Post Home Ownership				
No	20.5	19.9	17.8	17.4
Yes	79.5	80.1	82.2	82.6
Pre Neighborhood Income				
Lowest Quintile	18.8	19.3	16.8	18.3
Non-lowest Quintile	81.2	80.7	83.2	81.7
Post Neighborhood Income				
Lowest Quintile	19.3	20.0	16.3	17.4
Non-lowest Quintile	80.7	80.0	83.7	82.6
Pre Rurality				
Rural	17.9	17.6	9.3	9.7
Suburban	32.1	32.0	22.3	21.2
Urban	50.0	50.4	68.4	69.1
Post Rurality				
Rural	18.4	17.6	9.8	9.7
Suburban	32.0	32.8	22.2	21.6
Urban	49.6	49.6	68.0	68.7

¥ Age-sex standardized to CCHS 1.1 Respondents
Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.4: Exposures of 3rd Transition Cohort (CCHS 3.1 to CCHS 2009-2010) ¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base (N=18,286)	Target (N=18,995)	Base (N=5,475,017)	Target (N=5,450,599)
Immobility Transitions				
Severe to Severe	11.3	12.1	9.4	10.9
Moderate to Severe	3.5	3.8	2.8	3.4
Moderate to Moderate	16.8	17.7	15.1	15.1
None to Moderate	4.8	4.3	4.6	4.0
None to None	63.6	62.0	68.2	66.6
Pre Immobility				
Severe	11.3	12.1	9.4	10.9
Moderate	20.3	21.5	17.9	18.5
None	68.4	66.4	72.7	70.5
Post Immobility				
Severe	14.8	15.9	12.2	14.3
Moderate	21.6	22.0	19.6	19.1
None	63.6	62.0	68.2	66.6
Pre Smoking				
Heavy	8.2	7.0	6.4	5.1
Light	17.3	17.8	15.2	17.2
Recent Heavy	4.9	4.2	4.4	3.5
Recent Light	8.3	8.5	8.4	8.1
Non-Recent Heavy	4.9	3.9	4.1	3.3
Non-Recent Light	11.0	10.9	10.5	9.6
Non-Smoker	45.4	47.7	51.0	53.2
Post Smoking				
Heavy	6.8	5.9	5.3	4.1
Light	15.2	15.6	13.2	15.0
Recent Heavy	5.2	4.1	4.6	3.5
Recent Light	8.9	7.9	8.7	7.2
Non-Recent Heavy	6.0	5.1	5.0	4.3
Non-Recent Light	12.5	13.7	12.2	12.7
Non-Smoker	45.4	47.7	51.0	53.2
Pre Leisure Inactivity				
High	22.5	24.1	23.4	24.1
Moderate	47.2	46.2	47.8	46.6
Low	30.3	29.7	28.8	29.3
Post Leisure Inactivity				
High	23.7	25.3	24.2	24.9
Moderate	47.0	46.0	47.6	46.4
Low	29.3	28.7	28.2	28.7
Pre Non-active Transport				
High	41.4	39.0	40.9	37.9

Moderate	32.9	30.9	33.7	30.9
Low	25.7	30.1	***25.4***	***31.2***
Post Non-active Transport				
High	43.4	41.0	42.7	39.6
Moderate	32.1	30.2	32.9	30.3
Low	24.5	28.8	***24.4***	***30.1***
Pre Sedentary Activity				
High	27.4	28.9	22.5	23.6
Moderate	36.7	37.5	35.4	35.3
Low	36.0	33.7	42.1	41.1
Post Sedentary Activity				
High	34.0	36.0	28.8	30.1
Moderate	37.8	37.4	37.5	36.3
Low	28.3	26.6	33.8	33.6
Pre Unhealthy Diet				
Extremely Inadequate	23.2	22.4	21.2	20.0
Less than Adequate	49.7	49.3	50.0	49.5
Adequate	27.1	28.2	28.8	30.5
Post Unhealthy Diet				
Extremely Inadequate	23.9	23.1	21.7	20.6
Less than Adequate	49.5	49.1	49.9	49.3
Adequate	26.6	27.8	28.4	30.1
Education				
Less than High school	22.1	20.5	18.1	17.1
High school	24.2	24.6	24.1	24.5
Post-secondary	53.7	54.9	57.8	58.4
Immigrant Status				
Yes	22.1	20.5	18.1	17.1
No	24.2	24.6	24.1	24.5
Body Mass Index				
Morbidly Obese (>=40)	2.8	3.5	2.3	2.7
Underweight (<20)	2.6	3.1	2.9	3.3
Other (20 to <40)	94.6	93.5	94.8	94.1
Sense of Belonging				
Very Weak	9.4	8.8	11.0	11.2
Not Very Weak	90.6	91.2	89.0	88.8
Food Insecurity				
Insecure	9.3	8.4	9.0	8.0
Secure	90.7	91.6	91.0	92.0
Pre Marital Status				
Single/Divorced	37.4	36.6	23.8	23.0
Married/Common-Law	62.6	63.4	76.2	77.0
Post Marital Status				
Single/Divorced	41.0	40.1	28.0	27.6
Married/Common-Law	59.0	59.9	72.0	72.4

Pre Home Ownership				
No	19.9	19.4	16.0	17.4
Yes	80.1	80.6	84.0	82.6
Post Home Ownership				
No	20.7	20.2	16.7	18.0
Yes	79.3	79.8	83.3	82.0
Pre Neighborhood Income				
Lowest Quintile	19.8	18.9	17.5	17.7
Non-lowest Quintile	80.2	81.1	82.5	82.3
Post Neighborhood Income				
Lowest Quintile	19.4	19.2	16.5	17.9
Non-lowest Quintile	80.6	80.8	83.5	82.1
Pre Rurality				
Rural	19.2	18.4	9.0	8.9
Suburban	31.4	31.0	21.8	21.4
Urban	49.3	50.5	69.2	69.7
Post Rurality				
Rural	19.5	18.6	9.5	9.0
Suburban	31.5	31.6	22.1	21.7
Urban	49.0	49.8	68.4	69.2

¥ Age-sex standardized to CCHS 1.1 Respondents
Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.5: Exposures of 4th Transition Cohort (CCHS 2007-2008 to CCHS 2011-2012) ¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base (N=20,328)	Target (N=19,144)	Base (N=5,790,758)	Target (N=5,642,552)
Immobility Transitions				
Severe to Severe	12.7	12.2	11.9	10.8
Moderate to Severe	3.6	3.4	3.1	2.8
Moderate to Moderate	18.8	17.4	16.2	14.2
None to Moderate	4.2	4.8	4.0	4.5
None to None	60.6	62.2	64.8	67.7
Pre Immobility				
Severe	12.7	12.2	11.9	10.8
Moderate	22.4	20.8	19.4	17.0
None	64.9	67.0	68.7	72.2
Post Immobility				
Severe	16.3	15.6	15.0	13.6
Moderate	23.0	22.2	20.2	18.7
None	60.6	62.2	64.8	67.7
Pre Smoking				
Heavy	6.4	7.1	4.9	5.8
Light	16.5	17.1	15.1	15.8
Recent Heavy	4.5	4.3	3.7	3.3
Recent Light	8.6	8.6	8.6	8.2
Non-Recent Heavy	4.8	4.4	3.8	3.8
Non-Recent Light	11.2	10.9	9.9	9.5
Non-Smoker	47.9	47.6	53.9	53.5
Post Smoking				
Heavy	5.4	6.1	4.1	4.5
Light	14.6	15.2	13.3	14.1
Recent Heavy	4.5	4.0	3.7	3.5
Recent Light	9.1	7.4	9.2	6.9
Non-Recent Heavy	5.8	5.7	4.6	5.0
Non-Recent Light	12.6	13.9	11.2	12.5
Non-Smoker	47.9	47.6	53.9	53.5
Pre Leisure Inactivity				
High	23.3	22.0	24.9	23.6
Moderate	47.4	45.8	47.3	46.0
Low	29.3	32.2	27.8	30.4
Post Leisure Inactivity				
High	24.3	23.0	25.6	24.4
Moderate	47.2	45.7	47.1	45.7
Low	28.5	31.4	27.3	29.9
Pre Non-active Transport				
High	38.8	37.6	38.2	36.5

Moderate	31.8	30.6	31.6	31.1
Low	29.4	31.8	30.2	32.4
Post Non-active Transport				
High	41.3	39.9	40.3	38.6
Moderate	30.4	29.4	30.3	30.0
Low	28.4	30.6	29.4	31.4
Pre Sedentary Activity				
High	27.7	28.9	23.1	24.9
Moderate	41.6	40.6	40.7	39.4
Low	30.7	30.5	36.2	35.7
Post Sedentary Activity				
High	35.7	36.8	30.8	33.0
Moderate	40.9	40.7	41.1	40.0
Low	23.4	22.6	28.1	27.0
Pre Unhealthy Diet				
Extremely Inadequate	23.3	23.1	21.2	21.3
Less than Adequate	49.5	49.3	50.0	49.5
Adequate	27.2	27.6	28.8	29.1
Post Unhealthy Diet				
Extremely Inadequate	24.0	23.8	21.8	22.0
Less than Adequate	49.5	49.2	50.0	49.4
Adequate	26.6	27.0	28.2	28.6
Education				
Less than High school	19.3	18.7	16.9	15.7
High school	23.8	23.2	23.5	22.7
Post-secondary	56.9	58.1	59.6	61.7
Immigrant Status				
Yes	22.6	21.1	36.4	34.5
No	77.4	78.9	63.6	65.5
Body Mass Index				
Morbidly Obese (>=40)	3.3	3.9	2.7	3.5
Underweight (<20)	2.8	3.4	3.1	3.3
Other (20 to <40)	93.9	92.8	94.2	93.2
Sense of Belonging				
Very Weak	8.5	8.7	10.8	9.7
Not Very Weak	91.5	91.3	89.2	90.3
Food Insecurity				
Insecure	8.8	8.3	9.9	8.4
Secure	91.2	91.7	90.1	91.6
Pre Marital Status				
Single/Divorced	37.3	37.5	25.3	23.9
Married/Common-Law	62.7	62.5	74.7	76.1
Post Marital Status				
Single/Divorced	40.6	40.9	29.2	28.7
Married/Common-Law	59.4	59.1	70.8	71.3

Pre Home Ownership				
No	19.2	18.5	17.7	18.6
Yes	80.8	81.5	82.3	81.4
Post Home Ownership				
No	20.0	19.3	18.4	19.3
Yes	80.0	80.7	81.6	80.7
Pre Neighborhood Income				
Lowest Quintile	19.5	18.0	17.3	16.1
Non-lowest Quintile	80.5	82.0	82.7	83.9
Post Neighborhood Income				
Lowest Quintile	19.2	18.7	17.3	16.8
Non-lowest Quintile	80.8	81.3	82.7	83.2
Pre Rurality				
Rural	17.2	18.7	9.1	9.2
Suburban	32.0	31.1	20.9	21.7
Urban	50.9	50.2	70.0	69.1
Post Rurality				
Rural	17.5	18.5	9.7	9.1
Suburban	31.8	32.0	21.0	22.1
Urban	50.7	49.4	69.3	68.8

[¥] **Age-sex standardized to CCHS 1.1 Respondents**

Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.6: Exposures of 5th Transition Cohort (CCHS 2009-2010 to CCHS 2013-2014) ¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base (N=19,556)	Target (N=20,150)	Base (N=6,085,576)	Target (N=5,911,539)
Immobility Transitions				
Severe to Severe	12.2	12.6	11.1	11.5
Moderate to Severe	3.0	2.9	2.5	2.5
Moderate to Moderate	18.6	18.2	15.9	15.4
None to Moderate	4.2	3.7	3.9	3.9
None to None	62.0	62.5	66.6	66.8
Pre Immobility				
Severe	12.2	12.6	11.1	11.5
Moderate	21.6	21.2	18.4	17.9
None	66.2	66.2	70.5	70.6
Post Immobility				
Severe	15.2	15.5	13.7	13.9
Moderate	22.8	22.0	19.7	19.3
None	62.0	62.5	66.6	66.8
Pre Smoking				
Heavy	5.9	6.9	4.2	4.7
Light	16.3	16.4	15.1	14.6
Recent Heavy	3.9	3.8	3.2	3.1
Recent Light	8.5	8.1	7.9	8.4
Non-Recent Heavy	4.1	4.2	3.3	3.3
Non-Recent Light	11.8	10.8	10.6	9.0
Non-Smoker	49.5	49.9	55.7	56.9
Post Smoking				
Heavy	5.0	5.9	3.5	4.1
Light	14.5	14.4	13.3	12.6
Recent Heavy	4.1	3.5	3.3	2.9
Recent Light	9.2	7.6	8.9	7.7
Non-Recent Heavy	4.8	5.4	3.9	4.1
Non-Recent Light	12.9	13.3	11.5	11.6
Non-Smoker	49.5	49.9	55.7	56.9
Pre Leisure Inactivity				
High	22.8	21.7	24.1	23.9
Moderate	46.7	46.0	46.3	44.4
Low	30.4	32.3	29.7	31.7
Post Leisure Inactivity				
High	23.6	22.5	24.7	24.5
Moderate	46.6	45.9	46.0	44.2
Low	29.8	31.6	29.3	31.3
Pre Non-active Transport				
High	38.1	36.4	37.2	35.3

Moderate	30.6	30.0	30.5	29.6
Low	31.3	33.6	32.2	35.1
Post Non-active Transport				
High	40.9	39.2	39.7	37.8
Moderate	28.8	28.3	28.8	28.0
Low	30.4	32.6	31.5	34.2
Pre Sedentary Activity				
High	31.5	31.1	25.7	25.0
Moderate	37.7	38.1	36.4	35.9
Low	30.8	30.8	37.9	39.0
Post Sedentary Activity				
High	39.2	38.9	33.1	32.1
Moderate	37.9	37.9	37.6	37.2
Low	23.0	23.2	29.3	30.7
Pre Unhealthy Diet				
Extremely Inadequate	22.0	22.2	20.2	19.8
Less than Adequate	49.2	47.3	49.4	46.7
Adequate	28.7	30.5	30.3	33.5
Post Unhealthy Diet				
Extremely Inadequate	22.8	22.9	20.8	20.4
Less than Adequate	49.2	47.3	49.5	46.7
Adequate	28.0	29.8	29.7	32.9
Education				
Less than High school	17.0	16.9	14.1	14.4
High school	24.2	25.2	24.0	24.9
Post-secondary	58.8	57.9	61.9	60.8
Immigrant Status				
Yes	22.4	20.7	37.5	36.1
No	77.6	79.3	62.5	63.9
Body Mass Index				
Morbidly Obese (>=40)	3.5	4.2	2.6	3.3
Underweight (<20)	2.6	3.4	3.0	3.9
Other (20 to <40)	93.9	92.3	94.4	92.8
Sense of Belonging				
Very Weak	8.2	8.8	9.2	8.9
Not Very Weak	91.8	91.2	90.8	91.1
Food Insecurity				
Insecure	8.9	8.7	8.6	9.3
Secure	91.1	91.3	91.4	90.7
Pre Marital Status				
Single/Divorced	37.3	37.3	25.5	23.9
Married/Common-Law	62.7	62.7	74.5	76.1
Post Marital Status				
Single/Divorced	40.4	40.3	29.1	28.0
Married/Common-Law	59.6	59.7	70.9	72.0

Pre Home Ownership				
No	19.1	18.8	17.9	19.7
Yes	80.9	81.2	82.1	80.3
Post Home Ownership				
No	19.9	19.6	18.7	20.4
Yes	80.1	80.4	81.3	79.6
Pre Neighborhood Income				
Lowest Quintile	18.4	18.3	17.5	17.4
Non-lowest Quintile	81.6	81.7	82.5	82.6
Post Neighborhood Income				
Lowest Quintile	19.3	19.6	17.2	18.4
Non-lowest Quintile	80.7	80.4	82.8	81.6
Pre Rurality				
Rural	18.0	18.6	8.4	8.9
Suburban	30.9	30.8	20.8	22.4
Urban	51.2	50.7	70.8	68.7
Post Rurality				
Rural	18.4	18.6	9.1	8.9
Suburban	31.0	31.7	21.1	23.2
Urban	50.7	49.8	69.8	67.9

¥ Age-sex standardized to CCHS 1.1 Respondents

Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.7: Exposures of Overall Transition Cohort¥

Exposures	Unweighted Cohort (%)		Weighted Cohort (%)	
	Base (N=76,569)	Target (N=77,847)	Base (N=22,578,766)	Target (N=22,171,226)
Immobility Transitions				
Severe to Severe	12.0	12.2	10.6	11.1
Moderate to Severe	3.6	3.7	3.0	3.1
Moderate to Moderate	18.0	17.8	15.7	15.1
None to Moderate	4.5	4.4	4.3	4.2
None to None	62.0	61.9	66.4	66.5
Pre Immobility				
Severe	12.0	12.2	10.6	11.1
Moderate	21.5	21.5	18.8	18.2
None	66.5	66.3	70.7	70.7
Post Immobility				
Severe	15.5	15.9	13.6	14.2
Moderate	22.5	22.3	20.0	19.3
None	62.0	61.9	66.4	66.5
Pre Smoking				
Heavy	7.3	7.2	5.6	5.4
Light	16.8	17.4	15.4	16.1
Recent Heavy	4.6	4.4	4.0	3.5
Recent Light	8.4	8.5	8.2	8.3
Non-Recent Heavy	4.6	4.2	3.8	3.4
Non-Recent Light	11.1	10.6	10.2	9.3
Non-Smoker	47.2	47.8	52.8	53.9
Post Smoking				
Heavy	6.1	6.1	4.6	4.4
Light	14.9	15.3	13.5	14.1
Recent Heavy	4.8	4.1	4.1	3.4
Recent Light	9.0	7.7	8.9	7.4
Non-Recent Heavy	5.5	5.5	4.6	4.5
Non-Recent Light	12.5	13.5	11.5	12.2
Non-Smoker	47.2	47.8	52.8	53.9
Pre Leisure Inactivity				
High	23.1	23.0	24.4	24.1
Moderate	47.2	46.3	47.1	46.0
Low	29.7	30.7	28.5	29.8
Post Leisure Inactivity				
High	24.1	24.1	25.2	24.9
Moderate	47.0	46.1	46.9	45.8
Low	28.8	29.8	28.0	29.3
Pre Non-active Transport				
High	39.9	38.2	39.3	37.1

Moderate	32.3	30.8	32.3	30.8
Low	27.8	31.0	28.4	32.1
Post Non-active Transport				
High	42.2	40.5	41.3	39.1
Moderate	31.2	29.8	31.2	29.8
Low	26.7	29.7	27.5	31.1
Pre Sedentary Activity				
High	35.2	35.6	30.1	30.4
Moderate	38.9	39.3	38.7	38.5
Low	26.0	25.1	31.3	31.1
Post Sedentary Activity				
High	38.9	39.3	38.7	38.5
Moderate	26.0	25.1	31.3	31.1
Low	27.9	28.3	23.2	23.5
Pre Unhealthy Diet				
Extremely Inadequate	23.3	22.9	21.2	20.7
Less than Adequate	49.8	48.9	50.3	48.8
Adequate	26.8	28.2	28.5	30.6
Post Unhealthy Diet				
Extremely Inadequate	24.0	23.6	21.7	21.3
Less than Adequate	49.7	48.7	50.3	48.6
Adequate	26.3	27.7	28.0	30.1
Education				
Less than High school	20.6	19.7	17.4	16.8
High school	24.6	24.3	24.4	24.0
Post-secondary	54.8	56.0	58.2	59.2
Immigrant Status				
Yes	22.5	21.6	36.1	35.8
No	77.5	78.4	63.9	64.2
Body Mass Index				
Morbidly Obese (>=40)	3.0	3.7	2.4	3.0
Underweight (<20)	2.7	3.3	3.0	3.6
Other (20 to <40)	94.3	93.0	94.6	93.4
Sense of Belonging				
Very Weak	8.7	8.8	10.2	10.0
Not Very Weak	91.3	91.2	89.8	90.0
Food Insecurity				
Insecure	9.3	8.3	9.2	8.7
Secure	90.7	91.7	90.8	91.3
Pre Marital Status				
Single/Divorced	37.1	36.9	24.5	23.4
Married/Common-Law	62.9	63.1	75.5	76.6
Post Marital Status				
Single/Divorced	40.6	40.3	28.6	27.8
Married/Common-Law	59.4	59.7	71.4	72.2

Pre Home Ownership				
No	19.5	18.9	17.2	18.2
Yes	80.5	81.1	82.8	81.8
Post Home Ownership				
No	20.3	19.7	18.0	18.9
Yes	79.7	80.3	82.0	81.1
Pre Neighborhood Income				
Lowest Quintile	19.1	18.7	17.3	17.4
Non-lowest Quintile	80.9	81.3	82.7	82.6
Post Neighborhood Income				
Lowest Quintile	19.3	19.4	16.8	17.6
Non-lowest Quintile	80.7	80.6	83.2	82.4
Pre Rurality				
Rural	18.0	18.3	9.0	9.2
Suburban	31.6	31.2	21.4	21.7
Urban	50.4	50.4	69.6	69.2
Post Rurality				
Rural	18.4	18.4	9.5	9.2
Suburban	31.6	32.1	21.6	22.2
Urban	50.0	49.6	68.9	68.6

¥ **Age-sex standardized to CCHS 1.1 Respondents**

Bold-*Italic* demonstrate >0.1 standardized difference

Table 5.8: Description of No Immobility Cohort

Exposures	Unweighted Cohort		Weighted Cohort	
	Respondents	Percentage	Population	Prevalence
Immobility Transitions				
None to Moderate	7,113	7.0	1,870,730	5.9
None to None	93,968	93.0	29,724,773	94.1
Cohort Type				
Open (Base Cohorts)	50,147	49.6	15,916,493	50.4
Closed (Target Cohorts)	50,934	50.4	15,679,011	49.6
Sex				
Female	53,507	52.9	15,515,478	49.1
Male	47,574	47.1	16,080,025	50.9
Age Groups				
40 to 44	13,469	13.3	6,411,549	20.3
45 to 49	12,856	12.7	5,946,255	18.8
50 to 54	14,401	14.2	5,136,202	16.3
55 to 59	15,105	14.9	4,397,616	13.9
60 to 64	13,742	13.6	3,331,916	10.5
65 to 69	11,556	11.4	2,539,382	8.0
70 to 74	9,217	9.1	1,889,983	6.0
75 to 79	6,433	6.4	1,223,225	3.9
80 to 84	3,350	3.3	573,325	1.8
85 to 89	951	0.9	146,050	0.5
Pre Smoking				
Heavy	5,744	5.7	1,474,108	4.7
Light	15,737	15.6	4,695,810	14.9
Recent Heavy	4,179	4.2	1,107,277	3.5
Recent Light	8,506	8.5	2,666,987	8.4
Non-Recent Heavy	4,748	4.7	1,100,268	3.5
Non-Recent Light	12,304	12.2	3,138,386	9.9
Non-Smoker	49,862	49.3	17,412,668	55.1
Pre Leisure Inactivity				
High	18,682	18.5	6,521,356	20.6
Moderate	47,517	47.0	14,788,795	46.8
Low	34,881	34.5	10,285,352	32.6
Pre Non-active Transport				
High	36,889	36.5	11,380,516	36.0
Moderate	32,422	32.1	9,922,366	31.4
Low	31,770	31.4	10,292,622	32.6
Pre Sedentary Activity				
High	25,376	25.1	6,022,713	19.1
Moderate	39,961	39.5	11,782,140	37.3
Low	35,744	35.4	13,790,649	43.6
Pre Unhealthy Diet				

Extremely Inadequate	21,028	20.8	5,945,592	18.8
Less than Adequate	50,009	49.5	15,818,615	50.1
Adequate	30,043	29.7	9,831,296	31.1
Education				
Less than High school	18,944	18.7	4,587,014	14.5
High school	25,050	24.8	7,692,307	24.3
Post-secondary	57,087	56.5	19,316,182	61.1
Immigrant Status				
Yes	24,181	23.9	11,912,493	37.7
No	76,899	76.1	19,683,010	62.3
Body Mass Index				
Morbidly Obese (>=40)	1,952	1.9	547,089	1.7
Underweight (<20)	2,528	2.5	861,579	2.7
Other (20 to <40)	96,600	95.6	30,186,835	95.5
Sense of Belonging				
Very Weak	6,804	6.7	2,600,770	8.2
Not Very Weak	94,277	93.3	28,994,733	91.8
Food Insecurity				
Insecure	4,922	4.9	1,870,184	5.9
Secure	96,158	95.1	29,725,319	94.1
Pre Marital Status				
Single/Divorced	34,129	33.8	6,578,299	20.8
Married/Common-Law	66,951	66.2	25,017,204	79.2
Pre Home Ownership				
No	16,019	15.8	4,804,057	15.2
Yes	85,061	84.2	26,791,446	84.8
Pre Neighborhood Income				
Lowest Quintile	17,116	16.9	5,007,892	15.9
Non-lowest Quintile	83,964	83.1	26,587,611	84.1
Pre Rurality				
Rural	18,320	18.1	2,729,312	8.6
Suburban	31,665	31.3	6,613,048	20.9
Urban	51,096	50.5	22,253,143	70.4

Table 5.9: Description of Moderate Immobility Cohort

Exposures	Unweighted Cohort		Weighted Cohort	
	Respondents	Percentage	Population	Prevalence
Immobility Transitions				
Moderate to Severe	5,900	17.4	1,343,605	16.1
Moderate to Moderate	28,085	82.6	7,016,399	83.9
Cohort Type				
Open (Base Cohorts)	16,788	49.4	4,270,450	51.1
Closed (Target Cohorts)	17,198	50.6	4,089,554	48.9
Sex				
Female	18,557	54.6	4,308,701	51.5
Male	15,428	45.4	4,051,303	48.5
Age Groups				
40 to 44	3,161	9.3	1,225,522	14.7
45 to 49	3,663	10.8	1,280,366	15.3
50 to 54	4,568	13.4	1,229,143	14.7
55 to 59	5,242	15.4	1,282,110	15.3
60 to 64	5,010	14.7	1,079,654	12.9
65 to 69	4,069	12.0	804,522	9.6
70 to 74	3,412	10.0	618,264	7.4
75 to 79	2,771	8.2	495,498	5.9
80 to 84	1,556	4.6	262,633	3.1
85 to 89	534	1.6	82,291	1.0
Pre Smoking				
Heavy	2,967	8.7	637,313	7.6
Light	5,705	16.8	1,472,888	17.6
Recent Heavy	2,051	6.0	432,429	5.2
Recent Light	2,848	8.4	702,905	8.4
Non-Recent Heavy	1,941	5.7	379,645	4.5
Non-Recent Light	4,127	12.1	916,174	11.0
Non-Smoker	14,348	42.2	3,818,650	45.7
Pre Leisure Inactivity				
High	8,614	25.3	2,124,489	25.4
Moderate	16,518	48.6	4,080,511	48.8
Low	8,853	26.0	2,155,005	25.8
Pre Non-active Transport				
High	13,252	39.0	3,148,061	37.7
Moderate	11,592	34.1	2,822,434	33.8
Low	9,142	26.9	2,389,509	28.6
Pre Sedentary Activity				
High	11,983	35.3	2,482,082	29.7
Moderate	13,369	39.3	3,272,439	39.1
Low	8,633	25.4	2,605,483	31.2
Pre Unhealthy Diet				

Extremely Inadequate	8,347	24.6	1,906,107	22.8
Less than Adequate	16,982	50.0	4,174,525	49.9
Adequate	8,657	25.5	2,279,372	27.3
Education				
Less than High school	7,812	23.0	1,617,814	19.4
High school	8,058	23.7	1,987,777	23.8
Post-secondary	18,115	53.3	4,754,413	56.9
Immigrant Status				
Yes	6,343	18.7	2,341,820	28.0
No	27,643	81.3	6,018,184	72.0
Body Mass Index				
Morbidly Obese (>=40)	1,709	5.0	384,251	4.6
Underweight (<20)	839	2.5	209,072	2.5
Other (20 to <40)	31,437	92.5	7,766,681	92.9
Sense of Belonging				
Very Weak	3,171	9.3	846,789	10.1
Not Very Weak	30,814	90.7	7,513,215	89.9
Food Insecurity				
Insecure	3,894	11.5	1,033,147	12.4
Secure	30,091	88.5	7,326,857	87.6
Pre Marital Status				
Single/Divorced	13,551	39.9	2,291,190	27.4
Married/Common-Law	20,434	60.1	6,068,814	72.6
Pre Home Ownership				
No	7,170	21.1	1,665,901	19.9
Yes	26,816	78.9	6,694,103	80.1
Pre Neighborhood Income				
Lowest Quintile	6,975	20.5	1,571,465	18.8
Non-lowest Quintile	27,010	79.5	6,788,539	81.2
Pre Rurality				
Rural	6,688	19.7	919,652	11.0
Suburban	11,084	32.6	2,041,683	24.4
Urban	16,213	47.7	5,398,669	64.6

Table 5.10: Description of Severe Immobility Cohort

Exposures	Unweighted Cohort		Weighted Cohort	
	Respondents	Percentage	Population	Prevalence
Immobility Transitions				
Severe to Severe	19,350	100.0	4,794,485	100.0
Cohort Type				
Open (Base Cohorts)	9,635	49.8	2,391,823	49.9
Closed (Target Cohorts)	9,715	50.2	2,402,661	50.1
Sex				
Female	14,051	72.6	3,367,582	70.2
Male	5,299	27.4	1,426,902	29.8
Age Groups				
40 to 44	1,279	6.6	487,799	10.2
45 to 49	1,630	8.4	587,921	12.3
50 to 54	2,039	10.5	609,897	12.7
55 to 59	2,239	11.6	534,183	11.1
60 to 64	2,158	11.2	518,709	10.8
65 to 69	1,793	9.3	414,358	8.6
70 to 74	2,062	10.7	448,352	9.4
75 to 79	2,469	12.8	511,428	10.7
80 to 84	2,385	12.3	451,922	9.4
85 to 89	1,296	6.7	229,916	4.8
Pre Smoking				
Heavy	1,826	9.4	369,340	7.7
Light	3,293	17.0	834,543	17.4
Recent Heavy	910	2.4	176,396	1.8
Recent Light	1,401	3.6	307,870	3.2
Non-Recent Heavy	1,027	5.3	199,414	4.2
Non-Recent Light	2,109	10.9	445,901	9.3
Non-Smoker	8,783	45.4	2,461,020	51.3
Pre Leisure Inactivity				
High	8,436	43.6	2,147,643	44.8
Moderate	8,126	42.0	1,990,154	41.5
Low	2,788	14.4	656,687	13.7
Pre Non-active Transport				
High	10,840	56.0	2,606,038	54.4
Moderate	5,454	28.2	1,380,127	28.8
Low	3,056	15.8	808,319	16.9
Pre Sedentary Activity				
High	9,254	47.8	2,020,344	42.1
Moderate	6,755	34.9	1,710,972	35.7
Low	3,341	17.3	1,063,169	22.2
Pre Unhealthy Diet				
Extremely Inadequate	6,063	31.3	1,494,558	31.2

Less than Adequate	9,038	46.7	2,162,486	45.1
Adequate	4,249	22.0	1,137,440	23.7
Education				
Less than High school	6,577	34.0	1,500,580	31.3
High school	4,389	22.7	1,119,295	23.3
Post-secondary	8,384	43.3	2,174,610	45.4
Immigrant Status				
Yes	4,611	23.8	1,883,763	39.3
No	14,739	76.2	2,910,721	60.7
Body Mass Index				
Morbidly Obese (>=40)	1,400	7.2	302,428	6.3
Underweight (<20)	1,250	6.5	352,831	7.4
Other (20 to <40)	16,700	86.3	4,139,225	86.3
Sense of Belonging				
Very Weak	3,474	18.0	1,075,099	22.4
Not Very Weak	15,876	82.0	3,719,385	77.6
Food Insecurity				
Insecure	3,856	19.9	1,032,059	21.5
Secure	15,495	80.1	3,762,426	78.5
Pre Marital Status				
Single/Divorced	9,603	49.6	1,750,604	36.5
Married/Common-Law	9,747	50.4	3,043,881	63.5
Pre Home Ownership				
No	6,014	31.1	1,376,851	28.7
Yes	13,336	68.9	3,417,633	71.3
Pre Neighborhood Income				
Lowest Quintile	4,840	25.0	1,143,119	23.8
Non-lowest Quintile	14,511	75.0	3,651,366	76.2
Pre Rurality				
Rural	3,601	18.6	444,500	9.3
Suburban	6,145	31.8	1,007,796	21.0
Urban	9,604	49.6	3,342,188	69.7

Table 5.11: Odd Ratios for Transitions from No Immobility to Moderate Immobility

	Unweighted Model		Weighted Model	
Exposures	**OR (90% CI)**	**P-Value**	**OR (90% CI)**	**P-Value**
Open vs. Closed Cohort	0.99 (0.93, 1.06)	0.85	0.97 (0.88, 1.06)	0.54
Year	0.98 (0.95, 1.00)	0.12	0.97 (0.95, 1.01)	0.17
Age	1.04 (1.03, 1.06)	<0.01	1.03 (1.02, 1.05)	<0.01
Age Above 75	1.06 (1.02, 1.09)	<0.01	1.09 (1.04, 1.13)	<0.01
Sex				
Female	1.18 (1.02, 1.38)	0.07	1.27 (1.05, 1.53)	0.04
Male	Ref.		Ref.	
Smoking				
Heavy	2.13 (1.68, 2.70)	<0.01	2.21 (1.69, 2.89)	<0.01
Recent Heavy	1.57 (1.20, 2.04)	0.06	1.44 (1.07, 1.94)	0.05
Non-Recent Heavy	1.31 (1.00, 1.74)	0.11	1.29 (0.97, 1.724	0.13
Light	1.34 (1.09, 1.65)	0.02	1.43 (1.08, 1.90)	0.03
Non-Smoker or Former Light	Ref.		Ref.	
Leisure Physical Inactivity				
High	2.31 (1.69, 3.16)	<0.01	2.21 (1.60, 3.04)	<0.01
Moderate	1.56 (1.18, 2.06)	<0.01	1.49 (1.13, 1.97)	0.02
Low	Ref.		Ref.	
Non-active Transport				
High	2.61 (1.82, 3.75)	<0.01	2.62 (1.81, 3.81)	<0.01
Moderate	2.20 (1.49, 3.24)	<0.01	2.24 (1.54, 3.27)	<0.01
Low	Ref.		Ref.	
Sedentary Activity				
High	1.78 (1.30, 2.45)	<0.01	1.90 (1.33, 2.73)	<0.01
Moderate	1.32 (1.05, 1.67)	0.05	1.36 (1.04, 1.79)	0.06
Low	Ref.		Ref.	
Unhealthy Diet Score				
Extremely Inadequate	1.15 (0.97, 1.36)	0.17	1.17 (0.97, 1.39)	0.16
Not Extremely Inadequate	Ref.		Ref.	
Body Mass Index				
Morbidly Obese (>=40)	1.81 (1.15, 2.84)	0.04	1.79 (1.10, 2.91)	0.05
Underweight (<20)	1.69 (1.30, 2.19)	0.03	1.53 (1.09, 2.14)	0.04
Other (20 to <40)	Ref.		Ref.	
Sense of Belonging				
Very Weak	1.65 (1.33, 2.05)	<0.01	1.63 (1.31, 2.03)	<0.01
Not Very Weak	Ref.		Ref.	
Food Insecurity				
Insecure	1.76 (1.31, 2.36)	<0.01	1.79 (1.29, 2.48)	<0.01
Secure	Ref.		Ref.	
Marital Status				
Single/Divorced	1.18 (0.99, 1.42)	0.13	1.13 (0.92, 1.39)	0.33

Married/Common-Law	Ref.		Ref.	
Home Ownership				
No	1.22 (1.01, 1.46)	0.08	1.3 (1.05, 1.61)	0.05
Yes	Ref.		Ref.	

Table 5.12: Odd Ratios for Transitions from Moderate to Severe Immobility

Exposures	Unweighted Model OR (90% CI)	Unweighted Model P-Value	Weighted Model OR (90% CI)	Weighted Model P-Value
Open vs. Closed Cohort	1.02 (0.95, 1.10)	0.61	0.99 (0.90, 1.08)	0.81
Year	0.94 (0.91, 0.96)	<0.01	0.93 (0.90, 0.97)	<0.01
Age	1.07 (1.05, 1.09)	<0.01	1.05 (1.03, 1.07)	<0.01
Age Above 75	1.13 (1.07, 1.19)	<0.01	1.19 (1.13, 1.26)	<0.01
Sex				
Female	1.72 (1.47, 2.02)	<0.01	1.72 (1.42, 2.08)	<0.01
Male	Ref.		Ref.	
Smoking				
Heavy	1.49 (1.11, 2.00)	0.03	1.41 (1.00, 1.98)	0.10
Light	1.23 (0.99, 1.54)	0.12	1.24 (0.95, 1.62)	0.18
Non-Smoker or Former	Ref.		Ref.	
Leisure Physical Inactivity				
High	1.91 (1.37, 2.65)	<0.01	2.07 (1.43, 3.01)	<0.01
Moderate	1.48 (1.1, 2.00)	0.03	1.50 (1.09, 2.08)	0.04
Low	Ref.		Ref.	
Non-active Transport				
High	2.54 (1.91, 3.39)	<0.01	2.63 (1.86, 3.72)	<0.01
Moderate	1.48 (1.09, 2.02)	0.04	1.49 (1.03, 2.17)	0.08
Low	Ref.		Ref.	
Unhealthy Diet Score				
Extremely Inadequate	1.36 (1.11, 1.65)	0.01	1.41 (1.13, 1.77)	0.01
Not Extremely Inadequate	Ref.		Ref.	
Immigrant Status				
Yes	1.35 (1.11, 1.65)	0.01	1.59 (1.27, 1.99)	0.01
No	Ref.		Ref.	
Body Mass Index				
Underweight (<20)	3.21 (2.17, 4.77)	<0.01	3.09 (2.03, 4.69)	<0.01
Not Underweight	Ref.		Ref.	
Sense of Belonging				
Very Weak	1.92 (1.45, 2.53)	<0.01	1.99 (1.47, 2.71)	<0.01
Not Very Weak	Ref.		Ref.	
Home Ownership				
No	1.42 (1.16, 1.75)	<0.01	1.53 (1.22, 1.91)	<0.01
Yes	Ref.		Ref.	
Pre Neighborhood Income				
Lowest Quintile	1.16 (0.98, 1.38)	0.15	1.12 (0.92, 1.37)	0.34
Non-lowest Quintile	Ref.		Ref.	

6. Representation of Canadian Community Health Survey respondents for health outcome linkage studies: 2003 to 2014

6.1 ABSTRACT

Background

Similar to other household surveys, the Canadian Community Health Survey (CCHS) has experienced decreasing response rates. There is concern that participation of unhealthy respondents in particular is declining and increasing health-related non-response biases within CCHS samples. This study compared the health of unweighted linked CCHS respondents with residents of the community-dwelling ("community") and general populations ("general"), and examined whether health differences were consistent with expectations and stable over time.

Data and Methods

Ontario subsamples of the CCHS (2.1, 3.1, 2007-2014) were linked to the Registered Persons Database (RPDB) and pooled. Two matched cohorts (community and general cohorts) were constructed from RPDB by matching (1:5) on birth year, sex, public health unit, and survey date. The community cohort excluded individuals who were long-term care (LTC) residents, which was ascertained from linkages to physician, lab, and drug administrative records. The matched cohorts were compared to the CCHS cohort using comorbidity measures (historical use of Ontario Disability Support Program, specific diseases, multimorbidity, presence of severe illness, and prior history of severe illness) ascertained through linkages to six administrative health databases, in addition to death and LTC admission (only community comparison). Standardized differences and Cox proportional hazards regression were used to compare cohorts.

Results

The CCHS cohort (n=197,722) included over 98% of linked respondents. The community cohort demonstrated higher prevalence of dementia than the CCHS cohort in age groups 85 and older (Cohen's d > 0.10), whereas the general cohort displayed higher prevalence of dementia and severe mental illness in age groups 75 and older (Cohen's d > 0.10), and higher prevalence of history of stroke, multimorbidity, and history of severe illness in age groups 85 and older (Cohen's d > 0.10). Both matched cohorts demonstrated higher mortality risk (community

Hazard Ratio (HR): 1.08 (1.06 - 1.11); general HR: 1.19 (1.17 - 1.22), and the community cohort also displayed higher risk of admission for LTC (HR: 1.22 (1.18 - 1.27)). Health outcome differences were more pronounced at older ages (p-value < 0.05), but the differences were stable over time (p-value > 0.30).

Interpretation

Health differences between linked CCHS respondents and two matched sampling frames are consistent with expectations and stable over time. Despite decreasing response rates, there are no indications of increasing health-related non-response biases within the CCHS. Future studies should evaluate the extent to which these biases exist within each survey cycle, and the adequacy of weighting procedures in addressing these biases.

6.2 INTRODUCTION

The Canadian Community Health Survey (CCHS) is a population-based health survey routinely administered by Statistics Canada.[130] The survey was designed to provide cross-sectional measures that are representative of the household population. This target population consists of individuals living in private residences, and therefore excludes individuals in long-term care (LTC) facilities, correctional facilities, retirement and group homes, shelters, and other non-private residences. Population-based health surveys are invaluable for governments and researchers because they can help assess leading health risks (e.g. lifestyle and sociodemographic information) and indicators of health status that are difficult to ascertain within the healthcare system or other data sources.[165–167] Linking CCHS respondents to administrative healthcare databases has expanded the survey's applicability, e.g., enabling the formation of longitudinal databases[125,168,169] These databases allow follow-up over time to assess future outcomes including the development of chronic diseases and health care use.[47,170–174] The linked uses of health surveys has enabled a wealth of unique and multi-layered research that would not have been possible otherwise, including the development of risk algorithms that can be leveraged to estimate the burden of health behaviours on life expectancy, healthcare utilization, and healthcare costs.[45,99,111] Research using the linked CCHS is bound to grow as additional cycles become available, and the types of administrative databases linked to the CCHS continue to diversify. Improving our understanding of the linked CCHS can help inform future research.

Similar to other household surveys around the world, the CCHS has experienced increases in non-response and non-consent to linkage since its introduction. In the Ontario subsamples, response rates have decreased from 82% in 2000 to 63% in 2014.[124] Over this same period, the proportion of Ontario respondents consenting to linkage and successfully linked has declined from 84% to 74%. The generalizability of linked household surveys is predicated on the ability of their survey weights to properly adjust for unequal selection probabilities, non-response, and non-consent to linkage. The falling response and consent (to linkage) rates in household surveys has raised concerns that samples are becoming increasingly biased and creating challenges for weighting procedures.[175–178] For some novel areas of research that utilize the linked CCHS, increasingly biased CCHS samples is already problematic in itself irrespective of the adequacy of weighting procedures. For example, there has been a proliferation of CCHS-based risk algorithms for linked chronic conditions and other rare outcomes in the literature.[47,170–174] It is not unusual for these algorithms to be derived from unweighted CCHS cohorts that are constructed by aggregating multiple cycles of the CCHS together.[45,99,111,179–181] This research may be vulnerable to non-response biases as a result, especially if the scale of bias is increasing across cycles. In many settings, survey respondents are healthier than non-respondents, and respondents who consent to linkage can also be healthier than respondents who decline linkage.[126,182–186] Unhealthy individuals may be increasingly underrepresented in linked CCHS samples because of increases in non-response and non-consent to linkage.

Linkage of the CCHS and population-based registries to administrative healthcare databases provide opportunities to explore health-related non-response biases through measures common to both CCHS respondents and the general population. Linkages to mortality records have been used to demonstrate that at very old ages linked CCHS respondents have lower rates of mortality than individuals in the general population.[125] These differences need not implicate health-related non-response biases, since household surveys do not include individuals residing in institutions. Institutionalized individuals generally have very limited access to the wider community and have their everyday essential needs provided by the institution in which they reside. Although only about 0.5% of the Canadian population reside within institutional settings, they are predominantly: elderly LTC (LTC) residents; in poor health; and high health care users.[149,187–189] The remainder of the institutionalized population mostly resides within correctional facilities, but are likely less of a contributing factor since this group is relatively

small in Canada and are typically non-seniors. Excluding institutionalized individuals from the general population leads to the creation of the community-dwelling population, which is comprised of individuals in private households — the target population of household surveys — but also individuals living in collective dwellings. Only about 1.5% of the overall Canadian population reside within collective dwellings.[189] The majority of these individuals are seniors in retirement and group homes. Their health level is expected to fall in between the private household dwellers of the CCHS cohort (healthiest) and the health of LTC residents (least healthy).[190] Unlike LTC residents and long-term residents of correctional facilities (situated within federal facilities), residents of retirement and group homes currently cannot accurately be excluded from population-based registries. As a result, a viable representation of the household population, which serves as the target population of the CCHS, cannot be derived from population-based registries.

Although neither the community-dwelling nor general population represents an ideal reference standard of the CCHS target population, they can still be indirectly utilized to evaluate whether unhealthy individuals are being increasingly underrepresented in CCHS samples as rates of non-response and non-consent to linkage increase. If unhealthy individuals are increasingly underrepresented in the CCHS samples, differences in health between linked CCHS respondents and corresponding individuals in either the community-dwelling or general populations would be expected to increase over time. On the other hand, if health differences are stable over time, it would suggest that the health of linked CCHS respondents is comparable over time regardless of increasing trends in non-response and non-consent. Under this scenario, survey weighting procedures should not be any less adequately presently than they have been in the past, and future research that pools linked respondents from multiple cycles into a single unweighted CCHS cohort can be confident that the CCHS samples are at least comparable.

This study had two key objectives. The first objective was to assess whether there are differences in comorbidity measures (historical use of Ontario Disability Support Program (ODSP), specific diseases, multi-comorbidity, severe illness, and prior history of severe illness) and health outcomes (death and admission to LTC) between linked CCHS respondents and comparable individuals in the community-dwelling ("community") and general populations ("general"). The second objective was to evaluate whether differences in health outcomes were

stable across Calendar time. The community and general populations are naturally older than the household population. At the same time, the CCHS is designed to produce robust estimates at the health region level. As a result, it samples at much higher rates in sparsely populated regions, and much lower rates in heavily populated regions. Health differences over time could be a product of different demographic trends between the populations or evolving CCHS sampling strategies. A matched study design was employed utilizing basic demographics (birth year and sex) and health region (public health unit) to ensure that health differences between linked CCHS respondents and their corresponding matches are comparable across cycles.

6.3 DATA AND METHODS

Study Frames

Registered Person Database of Ontario

The Registered Person Database (RPDB) is a population-based registry maintained by the Ministry of Health and Long-Term Care in Ontario. It contains all unique health numbers that have been issued for Ontario Health Insurance Plan (OHIP), and their corresponding periods of coverage. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. These groups are also excluded from the CCHS sampling frame.

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian population. The survey employs a complex multistage sampling strategy and is conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions and collective dwellings. The details of the survey methodology have been previously published and are available online.[130] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content were linked to RPDB at ICES. The linked Ontario subsamples from 2.1 (2003-04), 3.1(2005-06), and 2007 to 2014 were combined to create an initial linked CCHS cohort of 201,048 respondents for this study.

Health Services & Utilization Data

Multiple types of administrative data were linked to the study frames for the purpose of this study, which included records from the OHIP database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), Ontario Mental Health Reporting System (OMHRS) database, Same-Day Surgery (SDS) database, and National Ambulatory Care Reporting System (NACRS) database. These databases were previously described in Section 2.2.

Development of Cohorts

Community and general matched cohorts were created from RPDB to compare the health of different sampling frames to the CCHS cohort. The community cohort excluded LTC residents, whereas the general cohort made no exclusions based on resident type. To facilitate the selection of appropriate matches, a population-based Ontario cohort for each survey year in the CCHS cohort (2003 to 2014) was created from RPDB. Each annual cohort was designed to represent the population of Ontarians within its respective year. To create a cohort for a particular year, all individuals from RPDB were selected whose records indicated: (1) Ontario residency within the year, (2) contact with the healthcare system in the prior 12-year period, (3) OHIP coverage for a period of time within the year, and (4) a valid Ontario postal code for linkage to geographic indicators. Linked CCHS respondents who were not present within their corresponding annual Ontario cohort were excluded, reducing the linked CCHS cohort to 200,543 respondents, or 99.8% of the initial CCHS cohort.

Since an individual's residence type changes over time, an initial pool of potential matches was constructed for each CCHS respondent disregarding residence type. Once assigned to a CCHS respondent's pool, the residence type of a potential match could be determined at the time of the CCHS respondent's survey date. For each CCHS respondent, 25 individuals were randomly selected from their corresponding annual Ontario cohort who matched on sex, birth year, and public health unit. The postal code conversion file was used to identify each individual's corresponding public health unit. CCHS respondents were removed from their own pool of potential matches in the rare case where they had been matched to themselves. Each potential match was assigned the same index date as their corresponding CCHS respondent (i.e., CCHS survey date). CCHS respondents and each of their potential matches were considered

eligible for the matched cohorts if they demonstrated OHIP eligibility within 6 months of their index date, and contact with the health care system before (<=10 years) and after (<=6 years) their index date. Contact was defined as any form of healthcare use, registration or renewal of the OHIP card, or death. These criteria further reduced the linked CCHS cohort to 197,911 respondents, or 98.4% of the initial CCHS cohort. This process produced a potential pool of matches for each of these linked CCHS respondents that included both community-dwelling and LTC residents.

Each potential match's type of residence (i.e., community-dwelling or LTC) at the time of their index date was indirectly inferred from OHIP and ODB records in the 9 months before their index date. LTC residency could not be directly ascertained from CCRS since not all LTC facilities reported to CCRS until late 2009. The OHIP and ODB records contain indicators that designate whether an individual is a LTC resident. The first of two consecutive records with LTC indicators were considered a LTC admission, and the first of two consecutive records without LTC indicators was considered a LTC discharge. This approach was designed to limit miss-classification from a single false positive or false negative record. Each individual in a CCHS respondent's pool of potential matches was designated as either a community-dwelling or LTC resident using these estimated LTC admission and discharge dates. Potential matches from 2011 to 2014 whose residency status was directly known from the CCRS were used to validate this methodology. After reclassifying any individuals under the age of 35 as community-dwelling residents, the algorithm utilized to identify LTC residents demonstrated good performance metrics in the validation set (sensitivity: 98.56, specificity: 99.97).

The matched cohorts were generated from the population sample of potential matches. For the community-dwelling matched cohort, five matches were randomly selected for each CCHS respondent from among its corresponding pool of potential matches who were community-dwelling residents. For the general matched cohort, five matches were randomly selected for each CCHS respondent from among its corresponding pool of potential matches irrespective of residence type. Given the rarity of obtaining multiple community-dwelling matches among very elderly CCHS respondents, CCHS respondents who were 95 years of age or older on January 1st of their survey year were excluded from the matching processes. This

resulted in a final CCHS cohort of 197,722 respondents and matched cohorts of 988,610 individuals each.

Baseline Characteristics

The individuals in each cohort were linked to baseline characteristics not used in the matching process to ascertain the comparability of the matched cohorts across important sociodemographic and disease indicators. The postal code conversion file was used to identify each individual's corresponding neighbourhood income quintile and rurality index of Ontario (RIO) score. Historical use of ODSP was ascertained by evaluating the funding source of over-the-counter drugs prescriptions in the prior 10 years. Pre-existing diseases were identified using healthcare utilization records. ICES-derived validated disease algorithms were utilized to identify dementia, asthma, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and hypertension.[131–136] Other diseases were identified based upon an in-hospital diagnosis or multiple relevant OHIP billings occurring within a specified number of years of each other. These diseases included renal disease, history of stroke, osteoporosis, heart arrhythmias, ischemic heart disease, history of cancer, osteoarthritis, and mental illness. 10-year look back windows were used when evaluating the presence of pre-existing diseases without a validated algorithm. The sixteen identified diseases were summed and categorized (0-1, 2-3, 4-5, 6 or more) to provide an indicator of multi-comorbidity.

The Canadian Institute for Health Information's (CIHI) population grouping methodology was utilized to provide indicators of current severe illness and prior history of severe illness. This methodology was applied to records from the DAD, SDS database, NACRS database, and OHIP database in a two year observation window around the index date (9 months before to 15 months after) to assess for the existence of severe illness at baseline, as well as the type of severe illness profile (palliative, cancer illness, mental illness, chronic illness, acute illness, or no severe illness).[191] This grouping methodology was also applied to OHIP records in a ten year window ending 9 months before the index date to assess prior history of severe illness. Based on the number of intervals with severe illness, the prior history of severe illness for each individual was defined as extensive (4-5), moderate (2-3), minimal (1), or none.

Health Outcomes

Individuals in each cohort were followed up to a maximum of five years from their index date, and censored at the beginning of any time period 6 months or longer where they were OHIP ineligible. Deaths within each individual's follow-up period were obtained from RPDB. Given that the general matched cohort already included LTC residents, LTC admission was only ascertained for individuals within the CCHS cohort and community cohort. LTC admission was indirectly inferred from OHIP and ODB records using the same methodology that had been previously used to assess LTC residency at the index date of each individual.

Statistical Analysis

Baseline characteristics were compared between each matched cohort and the CCHS cohort. Standardized differences were computed for each baseline characteristic comparison. A p-value from a statistic test can provide insight into whether a significant difference exists, but with sufficiently large samples, it will almost always demonstrate a significant difference.[192] In contrast, the standardized difference takes into account the effect size and is not dependent on sample size. It is increasingly used in observational studies to evaluate whether cohorts differ along baseline characteristics. A standardized difference greater than 0.1 is considered evidence of an important difference.[193] These comparisons were also performed within age groups (<65, 65-74, 75-84, and 85-94) to evaluate whether differences were age-dependent.

Models were constructed to compare rates of mortality and LTC admission between matched cohorts and the CCHS sample. A Cox proportional hazard model, stratified on the matched sets, was used to analyze the differences. The outcome was time to death with cohort type as the independent variable. A separate competing risk Cox proportional hazard model, stratified on the matched sets, was used to compare differences in rates of LTC admission between the CCHS cohort and the community-dwelling matched cohort, where the outcome was time until LTC admission with death as the competing risk. An interaction term was entered between the cohort type and age group (<65, 65-74, 75-84, and 85-94) in each mortality and LTC entry model to evaluate whether differences were more prominent in older age groups. An interaction term was also entered between cohort type and time periods (2003-2006, 2007-2010, and 2011-2014) in each mortality and LTC admission model to evaluate whether differences

were changing over time. All pairwise differences between the CCHS cohort and each matched cohort by age group and time period were estimated using least-square means.

6.4 RESULTS

Baseline Indicators

Outside of expected differences in the proportion of LTC residents, the CCHS cohort demonstrated similar overall baseline distributions to each matched cohort (See Table 6.1). The general cohort had a relatively small percentage of LTC residents overall (0.9%), but this percentage was largely age-dependent and only significant within the senior age groups. With a few exceptions within the oldest age groups, the baseline distributions of the matched cohorts were also very similar within different age groups (See Tables 6.2-6.5). In comparisons with the community cohort, the one exception was dementia within the oldest age group (85-94), where the community cohort was over 50% higher (13.9% vs. 8.9%). In comparisons with the general cohort, differences became apparent already in the 2nd oldest age group (75-84), where the general cohort had nearly double the percentage of dementia (8.5% vs. 4.3%), and over 60% higher severe mental illness (7.1% vs. 4.3%). These differences became more pronounced in the oldest age group, where the general cohort demonstrated over 150% higher dementia (22.3% vs. 8.9%), and roughly 100% higher severe mental illness (17.3% vs. 8.6%). Within this same age group, the general cohort also demonstrated a larger proportion of individuals with stroke history (13.4% vs. 9.9%), mental illness (33.3% vs. 27.2%), 6 or more comorbidities (25.6% vs. 20.4%), an extensive history of prior severe illness (8.9% vs. 5.9%), and a moderate history of prior severe illness (24.5% vs. 19.4%). All differences had a standardized difference >0.1.

Health Outcomes

The matched cohorts demonstrated worse health through comparisons of health outcomes (See Table 6.6 & 6.7). The community cohort demonstrated significantly higher risk of LTC admission than the CCHS cohort (HR: 1.22 (1.18 - 1.27)). These differences were only significant within the senior age groups, where differences increased with age (HR: 1.11 (1.01 - 1.22) to 1.30 (1.22 - 1.39)). No significant differences were observed in LTC admissions between the community cohort and CCHS cohort over time (p=0.31). The community cohort also demonstrated significantly higher mortality rates than the CCHS cohort (HR: 1.08 (1.06 –

1.11)). These differences were only significant within the two oldest age groups, where differences increased with age (HR: 1.10 (1.06 - 1.14) to 1.22 (1.16 - 1.28)). No significant differences in mortality risk were observed between the community cohort and CCHS cohort over time. The general cohort demonstrated an even higher risk of mortality than the CCHS cohort (HR: 1.19 (1.17 - 1.22)). These differences were only significant within the senior age groups, with differences also increasing with age (HR: 1.09 (1.04 - 1.14) to 1.43 (1.36 - 1.50)). No significant differences in mortality risk were observed between the general cohort and CCHS cohort over time (p=0.74).

6.5 DISCUSSION

Baseline health measures and health outcomes provided an assessment of health differences between linked CCHS respondents and comparable individuals in two matched cohorts. Health differences follow a pattern indicative of differences in the proportion of unhealthy individuals between the sampling frames. As expected, the CCHS cohort was healthier than its counterpart cohorts at older ages, where the exclusions (by CCHS) of those living in collective dwellings or institutions are expected to manifest in health differences. Differences were most evident relative to the general matched cohort, a reflection of the residents in retirement homes and LTC facilities in the general matched cohort that are excluded from the CCHS sampling frame. Although LTC residents are likely the largest contributing factor to health differences, the absence of retirement home residents from the CCHS sampling frame is also a notable factor as demonstrated by health differences relative to the community cohort. The number of dimensions with differences and the magnitude of differences increased with age with respect to both sampling frames, mirroring the likelihood of individuals to reside within either a retirement home or LTC facility which is also largely age dependent. The differences between each sampling frame and the CCHS demonstrate that the CCHS should be used with caution or not used at all for population-wide inferences beyond the household population.

The stability of health outcome differences over time suggests that health-related non-response biases are not increasing over time in conjunction with the observed increases in non-response and non-consent to linkage. If the participation of unhealthy respondents in particular was declining, health differences would have increased beyond the expected health differences that arise from differences in the sampling frames. Health may not be an important driver of

CCHS response rates in general. Health is closely associated with age, yet CCHS response rates and their decline over time have been practically uniform across age groups. It is plausible that pressure on the long-term care system from a Canadian ageing population may have increased the likelihood of unhealthy individuals remaining in the community over time and complicated interpretation of trends. The impact of these demographic trends is a recent phenomenon, however, with the first baby boomers retiring in 2012 and not even 70 years of age by the end of their five-year observation window in this study. Analysis of health-related non-response biases in more recent CCHS cycles may have to account for these trends, but they are unlikely to influence the cycles utilized in this study.

The results of this study are in line with examinations of other large population-based health surveys that are routinely administered. The National Health Interview Survey (NHIS) which is a survey similar to the CCHS that is administered in the United States has also experienced declines in response rates and consent to linkage over time. It initially achieved response rates greater than 90% in the 1990s, but more recently has seen response rates as low as 64%.[194] These response rates and their declines over time have been consistent across age groups. The proportion of respondents consenting to linkage, however, has remained relatively high, approximately 90% in most years. Research has demonstrated that despite the declines in response rates, differences in mortality rates between linked NHIS respondents and individuals in the general population have been relatively stable over a period of 20 years.[195] The equivalent population-based health survey in Scotland (SHeS) has experienced a decline in response rates from over 80% in 1995 to less than 60% in 2010.[196] Similar to the NHIS and CCHS, response rates and their subsequent declines in the SHeS have not demonstrated any differences across age groups. The proportion of its respondents consenting to linkage has remained relatively high, over 85% in all years. Research has demonstrated that differences in mortality rates between linked SHeS respondents and individuals in the general population to also be relatively stable over a period of 15 years.[197] Concerns that declining response rates in national health surveys are leading to increases in non-response biases may be overstated.

There are several areas of interest for future research. The baseline health measures and health outcomes in this study are only common amongst the most elderly, providing little insight into health differences that may exist between younger linked CCHS respondents and the other sampling frames. Moreover, health-related differences are not the only type of differences that

would be of interest to researchers. Evaluating differences with regards to socioeconomic indicators and other areas of interest would provide a more comprehensive picture of differences between the CCHS sampling frame and others. The lack of a viable representation of the household population, the CCHS target population, was an underlying limitation that permeated this entire study. Health differences between linked CCHS respondents and individuals in each sampling frame may be a combination of expected differences between individuals in the household population and each sampling frame, as well as health-related non-response biases in linked CCHS respondents. Opportunities to further understand linked CCHS respondents with respect to different sampling frames will expand as linkages to different types of administrative databases continue to diversify. For example, recent studies include residents of provincial correctional facilities and retirement homes, two populations of interest that are not part of the household population but could not be excluded from the comparative sampling frames of this study.[190,198] Future research will have opportunities to examine linked CCHS respondents with respect to a sampling frame more representative of the household population which will enable direct assessments of health-related non-response biases and other potential biases. Provision of a viable representation of the household population will also provide opportunities to assess weighted estimates of both the unlinked and linked CCHS, ensuring that the weighting processes do not introduce bias for weighted analyses.

6.6 TABLES

Table 6.1: Characteristics of All Ages in Matched Cohorts

Characteristics	CCHS Respondents N=197,722	Community Ontarians N=988,610	General Ontarians N=988,610
Time Period			
2003-2006	33.4%	33.4%	33.4%
2007-2010	34.0%	34.0%	34.0%
2011-2014	32.6%	32.6%	32.6%
Neighborhood Income Quintile			
Q1 (lowest)	19.1%	19.2%	19.3%
Q2	20.0%	19.8%	19.8%
Q3	20.3%	20.4%	20.3%
Q4	20.7%	20.6%	20.6%
Q5 (highest)	19.9%	20.0%	20.0%
Rurality			
Rural	17.1%	17.3%	17.3%
Suburban	30.5%	30.5%	30.5%
Urban	52.4%	52.3%	52.3%
Historical Use of ODSP			
Yes	4.9%	5.2%	5.3%
No	95.1%	94.8%	94.7%
Long-term Care Resident			
Yes	0.0%	0.0%	***0.9%***
No	100.0%	100.0%	98.9%
Comorbidities			
Dementia	0.9%	1.2%	1.7%
Renal Disease	1.8%	1.8%	1.8%
History of Stroke	2.0%	2.0%	2.2%
Asthma	2.2%	2.2%	2.2%
Congestive Heart Failure	2.7%	2.8%	2.9%
Osteoporosis	3.8%	3.7%	3.7%
Heart Arrhythmias	5.1%	5.1%	5.1%
Chronic Obstructive Pulmonary Disease	9.2%	9.0%	9.1%
Ischemic Heart Disease	10.1%	10.0%	10.0%
Diabetes	10.2%	10.6%	10.6%
History of Cancer	23.8%	22.5%	22.4%
Hypertension	29.2%	29.0%	28.9%
Mental Illness	30.5%	30.1%	30.5%
Osteoarthritis	32.9%	32.5%	32.4%
Multiple Comorbidities			

6 or More	3.5%	3.6%	3.7%
4 to 5	10.8%	10.5%	10.5%
2 to 3	28.5%	28.1%	28.0%
1 or less	57.3%	57.9%	57.8%
Current Severe Illness Profile			
Palliative	0.4%	0.5%	0.5%
Cancer Illness	2.2%	2.1%	2.1%
Mental Illness	1.9%	2.0%	2.4%
Chronic Illness	5.1%	4.9%	4.9%
Acute Illness	3.6%	3.7%	3.6%
None	86.8%	86.8%	86.4%
Prior History of Severe Illness			
Extensive	1.7%	1.7%	1.9%
Moderate	5.3%	5.2%	5.4%
Minimal	10.3%	10.1%	10.1%
None	82.7%	83.0%	82.6%

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 6.2: Characteristics of Under 65 Year Olds in Matched Cohorts

Characteristics	CCHS Respondents N=146,164	Community Ontarians N=730,820	General Ontarians N=730,820
Time Period			
2003-2006	34.9%	34.9%	34.9%
2007-2010	34.2%	34.2%	34.2%
2011-2014	30.9%	30.9%	30.9%
Neighborhood Income Quintile			
Q1 (lowest)	18.7%	19.0%	19.0%
Q2	19.8%	19.5%	19.5%
Q3	20.3%	20.4%	20.4%
Q4	21.0%	21.0%	20.9%
Q5 (highest)	20.1%	20.1%	20.1%
Rurality			
Rural	16.4%	16.5%	16.5%
Suburban	29.8%	29.7%	29.7%
Urban	53.8%	53.8%	53.8%
Historical Use of ODSP			
Yes	5.8%	5.9%	6.0%
No	94.2%	94.1%	94.0%
Long-term Care Resident			
Yes	0.0%	0.0%	0.1%
No	100.0%	100.0%	99.9%
Comorbidities			
Dementia	0.1%	0.1%	0.1%
Renal Disease	0.7%	0.6%	0.6%
History of Stroke	0.7%	0.6%	0.7%
Asthma	2.1%	2.1%	2.1%
Congestive Heart Failure	0.6%	0.6%	0.6%
Osteoporosis	1.7%	1.6%	1.6%
Heart Arrhythmias	2.1%	2.0%	2.0%
Chronic Obstructive Pulmonary Disease	4.8%	4.5%	4.6%
Ischemic Heart Disease	4.1%	3.8%	3.9%
Diabetes	5.8%	5.9%	5.9%
History of Cancer	16.7%	15.9%	15.8%
Hypertension	15.7%	15.2%	15.2%
Mental Illness	30.6%	30.0%	30.1%
Osteoarthritis	25.4%	25.1%	25.0%
Multiple Comorbidities			
6 or More	0.8%	0.8%	0.7%
4 to 5	4.8%	4.5%	4.5%

2 to 3	23.6%	23.4%	23.4%
1 or less	70.7%	71.4%	71.3%
Current Severe Illness Profile			
Palliative	0.1%	0.1%	0.1%
Cancer Illness	1.2%	1.1%	1.1%
Mental Illness	1.4%	1.3%	1.4%
Chronic Illness	2.5%	2.3%	2.3%
Acute Illness	2.3%	2.3%	2.3%
None	92.5%	92.9%	92.8%
Prior History of Severe Illness			
Extensive	1.0%	0.8%	0.9%
Moderate	2.9%	2.7%	2.7%
Minimal	7.6%	7.4%	7.3%
None	88.6%	89.1%	89.2%

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 6.3: Characteristics of 65-74 Year Olds in Matched Cohorts

Characteristics	CCHS Respondents N=27,751	Community Ontarians N=138,755	General Ontarians N=138,755
Time Period			
2003-2006	29.2%	29.2%	29.2%
2007-2010	32.5%	32.5%	32.5%
2011-2014	38.3%	38.3%	38.3%
Neighborhood Income Quintile			
Q1 (lowest)	19.0%	18.6%	18.6%
Q2	20.2%	20.1%	20.2%
Q3	20.2%	20.2%	20.1%
Q4	20.4%	20.5%	20.4%
Q5 (highest)	20.2%	20.6%	20.7%
Rurality			
Rural	20.3%	20.4%	20.4%
Suburban	32.7%	32.8%	32.8%
Urban	47.1%	46.8%	46.8%
Historical Use of ODSP			
Yes	4.4%	4.8%	5.1%
No	95.6%	95.2%	94.9%
Long-term Care Resident			
Yes	0.0%	0.0%	***0.8%***
No	100.0%	100.0%	99.2%
Comorbidities			
Dementia	1.1%	1.3%	1.8%
Renal Disease	3.3%	3.3%	3.3%
History of Stroke	3.7%	3.8%	4.0%
Asthma	2.1%	2.1%	2.2%
Congestive Heart Failure	5.2%	5.2%	5.3%
Osteoporosis	8.4%	8.1%	8.2%
Heart Arrhythmias	9.2%	9.2%	9.2%
Chronic Obstructive Pulmonary Disease	18.9%	18.4%	18.5%
Ischemic Heart Disease	21.4%	21.7%	21.6%
Diabetes	21.5%	22.7%	22.5%
History of Cancer	39.3%	37.0%	37.3%
Hypertension	60.2%	60.3%	60.3%
Mental Illness	31.6%	30.5%	31.1%
Osteoarthritis	50.4%	49.6%	49.5%
Multiple Comorbidities			
6 or More	6.8%	7.1%	7.1%
4 to 5	23.4%	22.7%	22.9%

2 to 3	44.8%	44.4%	44.5%
1 or less	24.9%	25.8%	25.5%
Current Severe Illness Profile			
Palliative	0.7%	0.8%	0.8%
Cancer Illness	4.6%	4.8%	4.7%
Mental Illness	1.7%	1.8%	2.1%
Chronic Illness	9.3%	9.2%	9.3%
Acute Illness	5.8%	6.0%	6.0%
None	77.9%	77.4%	77.0%
Prior History of Severe Illness			
Extensive	3.0%	3.0%	3.2%
Moderate	9.1%%	8.8%	9.1%
Minimal	15.3%	15.2%	15.2%
None	72.6%	73.1%	72.5%

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 6.4: Characteristics of 75-84 Year Olds in Matched Cohorts

Characteristics	CCHS Respondents N=18,389	Community Ontarians N=91,945	General Ontarians N=91,945
Time Period			
2003-2006	30.1%	30.1%	30.1%
2007-2010	34.6%	34.6%	34.6%
2011-2014	35.3%	35.3%	35.3%
Neighborhood Income Quintile			
Q1 (lowest)	21.0%	20.5%	20.8%
Q2	20.9%	21.2%	21.1%
Q3	20.1%	20.5%	20.2%
Q4	19.1%	19.2%	19.2%
Q5 (highest)	18.9%	18.7%	18.7%
Rurality			
Rural	18.2%	18.7%	18.9%
Suburban	32.5%	32.4%	32.5%
Urban	49.2%	48.8%	48.6%
Historical Use of ODSP			
Yes	0.4%	0.9%	0.8%
No	99.6%	99.1%	99.2%
Long-term Care Resident			
Yes	0.0%	0.0%	***4.0%***
No	100.0%	100.0%	96.0%
Comorbidities			
Dementia	4.3%	5.8%	***8.5%***
Renal Disease	6.2%	6.3%	6.5%
History of Stroke	7.2%	7.6%	8.4%
Asthma	2.7%	2.7%	2.8%
Congestive Heart Failure	11.2%	11.7%	12.3%
Osteoporosis	11.1%	11.0%	10.8%
Heart Arrhythmias	17.1%	17.5%	17.5%
Chronic Obstructive Pulmonary Disease	24.3%	24.6%	25.1%
Ischemic Heart Disease	32.8%	32.8%	32.8%
Diabetes	24.7%	25.7%	26.4%
History of Cancer	48.3%	45.3%	45.0%
Hypertension	74.4%	75.4%	75.0%
Mental Illness	29.0%	29.8%	31.2%
Osteoarthritis	57.8%	57.1%	56.8%
Multiple Comorbidities			
6 or More	14.5%	15.0%	16.1%
4 to 5	31.9%	32.0%	32.0%

2 to 3	40.0%	39.5%	38.6%
1 or less	13.6%	13.5%	13.3%
Current Severe Illness Profile			
Palliative	1.4%	1.8%	1.9%
Cancer Illness	5.7%	5.7%	5.4%
Mental Illness	4.3%	5.3%	***7.1%***
Chronic Illness	15.5%	15.1%	15.2%
Acute Illness	8.6%	8.7%	8.7%
None	64.5%	63.5%	61.7%
Prior History of Severe Illness			
Extensive	4.8%	5.1%	5.9%
Moderate	15.1%	14.9%	16.0%
Minimal	19.9%	20.3%	20.4%
None	60.3%	59.7%	57.6%

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 6.5: Characteristics of 85-94 Year Olds in Matched Cohorts

Characteristics	CCHS Respondents	Community Ontarians	General Ontarians
	N=5,418	N=27,090	N=27,090
Time Period			
2003-2006	24.2%	24.2%	24.2%
2007-2010	34.3%	34.3%	34.3%
2011-2014	41.4%	41.4%	41.4%
Neighborhood Income Quintile			
Q1 (lowest)	23.8%	22.8%	23.3%
Q2	21.5%	21.6%	21.4%
Q3	20.9%	20.0%	19.5%
Q4	16.9%	18.2%	18.5%
Q5 (highest)	17.0%	17.4%	17.4%
Rurality			
Rural	17.0%	16.8%	16.8%
Suburban	31.0%	32.2%	32.2%
Urban	52.0%	51.0%	50.9%
Historical Use of ODSP			
Yes	0.1%	0.3%	0.3%
No	99.9%	99.7%	99.7%
Long-term Care Resident			
Yes	0.0%	0.0%	***14.4%***
No	100.0%	100.0%	85.6%
Comorbidities			
Dementia	8.9%	***13.9%***	***22.3%***
Renal Disease	8.9%	9.1%	9.4%
History of Stroke	9.9%	11.0%	***13.4%***
Asthma	3.1%	3.2%	3.3%
Congestive Heart Failure	19.3%	20.9%	22.3%
Osteoporosis	12.5%	12.4%	12.7%
Heart Arrhythmias	24.0%	25.3%	25.1%
Chronic Obstructive Pulmonary Disease	26.6%	27.4%	28.7%
Ischemic Heart Disease	37.7%	38.5%	38.6%
Diabetes	22.0%	23.2%	23.8%
History of Cancer	52.0%	48.8%	47.3%
Hypertension	81.6%	82.4%	81.6%
Mental Illness	27.2%	29.9%	***33.3%***
Osteoarthritis	60.3%	61.0%	60.8%
Multiple Comorbidities			
6 or More	14.5%	15.0%	***16.1%***
4 to 5	31.9%	32.0%	32.0%

2 to 3	40.0%	39.5%	38.6%
1 or less	13.6%	13.5%	13.3%
Current Severe Illness Profile			
Palliative	2.8%	3.8%	3.7%
Cancer Illness	4.6%	4.5%	4.2%
Mental Illness	8.6%	11.7%	***17.3%***
Chronic Illness	20.3%	20.0%	19.2%
Acute Illness	11.0%	10.9%	10.1%
None	52.8%	49.0%	***45.4%***
Prior History of Severe Illness			
Extensive	5.9%	7.0%	***8.9%***
Moderate	19.4%	21.2%	***24.5%***
Minimal	24.0%	24.4%	24.6%
None	50.7%	47.5%	***42.0%***

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 6.6: Hazard Ratios for Long-term Care Entry

Outcomes	Community Ontarians versus CCHS Respondents
Overall	1.22 (1.18, 1.27)
Age Interaction (p=0.03)	
Under 65	1.15 (0.98, 1.34)
65 to 74	1.11 (1.01, 1.22)
75 to 84	1.21 (1.14, 1.28)
85 to 94	1.30 (1.22, 1.39)
Period Interaction (p=0.31)	
2003-2006	1.17 (1.10, 1.25)
2007-2010	1.22 (1.15, 1.30)
2011-2014	1.26 (1.18, 1.34)

Table 6.7. Cox-proportional Hazard Ratios for Mortality

Outcomes	Community Ontarians versus CCHS Respondents	General Ontarians versus CCHS Respondents
Overall	1.08 (1.06, 1.11)	1.19 (1.17, 1.22)
Age Interaction (p<0.01)		
Under 65	0.99 (0.94, 1.03)	0.99 (0.94, 1.04)
65 to 74	1.03 (0.99, 1.08)	1.09 (1.04, 1.14)
75 to 84	1.10 (1.06, 1.14)	1.25 (1.20, 1.29)
85 to 94	1.22 (1.16, 1.28)	1.43 (1.36, 1.50)
Period Interaction (p=0.74)		
2003-2006	1.09 (1.05, 1.14)	1.20 (1.16, 1.25)
2007-2010	1.08 (1.04, 1.12)	1.18 (1.14, 1.22)
2011-2014	1.08 (1.05, 1.12)	1.19 (1.15, 1.24)

7. Representation of Canadian Community Health Survey respondents for healthcare utilization and cost outcome linkage studies: 2007 to 2014

7.1 ABSTRACT

Background

Publicly funded healthcare costs associated with each Canadian Community Health Survey (CCHS) respondent can be estimated through linkages to healthcare databases. Canadians who are in poor health and high-health care users may not reside within its intended sampling frame, the household population. For example, people who live in retirement homes and long-term (LTC) facilities are not included in the CCHS sampling frame. We hypothesized that individuals in the community-dwelling ("community") and general population ("general") would accumulate costs more rapidly than linked CCHS respondents at older ages.

Data and Methods

Ontario subsamples of the CCHS (2007-2014) were linked to the Registered Persons Database (RPDB) and pooled. Two matched cohorts (community and general cohorts) were constructed from RPDB by matching (1:5) on birth year, sex, public health unit, and survey date. The community cohort excluded individuals who were LTC residents, which was inferred from linkages to physician, lab, and drug administrative records. Utilization for each cohort was ascertained through linkages to nine administrative healthcare databases. Negative binomial regressions were used to compare matched cohorts to the CCHS cohort with regard to the different types of utilization and sector-specific utilization (chronic, inpatient, secondary, primary, and other care). Healthcare costs were estimated for each type of utilization using established costing methodologies. Gamma regressions were used to compare the matched cohorts by total healthcare costs.

Results

The CCHS cohort (n=131,712) included over 98% of linked respondents. Both matched cohorts demonstrated greater overall costs (community relative rate (RR): 1.08 (1.06-1.10), general RR: 1.17 (1.15-1.19)) than the CCHS cohort. These differences occurred only in the elderly (≥65), where they increased with age, and were greater in general cohort comparisons.

This was primarily the product of greater utilization in chronic care (community RR: 1.40 (1.35, 1.45), general RR: 2.88 (2.83, 2.93)) observed across all ages, and in inpatient care (community RR: 1.12 (1.09, 1.15), general RR: 1.12 (1.09, 1.15)) observed only within the elderly (≥65). In addition, both matched cohorts displayed higher utilization in primary care (community RR: 1.03 (1.02, 1.04), general RR: 1.06 (1.05, 1.07)) within the oldest ages (≥75), and in other care (community RR: 1.05 (1.04, 1.06), general RR: 1.14 (1.13, 1.15)) within the elderly (≥65). One exception to these trends was secondary care, where matched cohorts demonstrated lower utilization (Both RR: 0.96 (0.95, 0.97).

Interpretation

In sectors associated with severe illness, linked CCHS respondents accumulate healthcare costs and utilization slower than individuals in the community-dwelling and general populations. The most ill and costly segments of the population typically do not reside within private household dwellings.

7.2 INTRODUCTION

Linkage of population-based health surveys to healthcare administrative data provide opportunities to assess healthcare costs for people based on sociodemographics, health behaviours and other characteristics. Healthcare administrative data is generated from various types of healthcare utilization with an underlying focus on medical conditions.[199] For example, hospital records indicate which relevant conditions were present upon admission or contributed to the length of the stay. Leading risk factors for many of these conditions, such as lifestyle and sociodemographic factors, are not captured within hospitalization records. In contrast, population-based health surveys are designed with the primary intent of capturing these key lifestyle and sociodemographic factors.[165–167,200] Although these surveys also habitually ascertain measures of healthcare utilization, they are limited, self-reported, and retrospective[181,201–203] The linkage of population-based health surveys to healthcare administrative data allows researchers to directly study the impact of lifestyle and sociodemographic factors on different types of healthcare utilization.[99,204,205] Using costing methods specific to each type of utilization, corresponding costs can be generated which take into account utilization intensity, and allow for aggregation of different utilization along a common unit of measurement (namely CN$).[127]

Given the rising costs of healthcare systems in many jurisdictions, there is growing interest in understanding the underlying drivers of healthcare costs.[206–208] Linked household surveys represent an invaluable resource.[96,209,210]

The Canadian Community Health Survey (CCHS) is a population-based health survey administered by Statistics Canada that has been linked to many databases.[130] In particular, the Ontario subsamples have been linked at ICES to a wide range of healthcare administrative data associated with the publicly funded universal healthcare system.[45,99,111] Researchers at ICES have developed corresponding costing methodologies for each type of utilization within the various sectors of care.[127] This includes chronic care (complex continuing care hospitalizations, and long-term care stays), inpatient care (acute, mental health, and rehabilitation hospitalizations), secondary care (hospital clinic visits, specialist care visits, and same-day procedures), primary care (general care visits, home care visits, and emergency care visits), as well as other care (laboratory testing, and drug prescriptions). The costing methodologies are able to allocate the vast majority of the province's healthcare costs to individual Ontarians using their Ontario Health Insurance Plan (OHIP) Number. The individual level cost of any linked CCHS respondent can therefore be estimated for any period of time in which utilization data is available. This ability to estimate healthcare costs for linked CCHS respondents within the Ontario subsamples has led to several novel research studies examining the impact of health behaviours and sociodemographic indicators on healthcare costs.[45,99,111]

The extent to which individuals in the general population accumulate utilization and cost more rapidly than respondents in the CCHS is unclear. Previous work (Chapter 6) has demonstrated that individuals in the general population have a higher risk of mortality than comparable individuals in the CCHS. The rate at which healthcare utilization and cost accumulate typically increases as people approach their end of their life.[145,146,211] An individual in their last year of life costs multiple times more than an individual of the same age who is not in their last year of life.[212] End of life costs consume 10% of the healthcare budget in Ontario.[149] Many individuals spend their last stages of life within long-term care facilities. The household population, however, does not include institutional settings such as long-term care facilities. Individuals in long-term care receive 24-hour nursing and personal care, with the vast majority dying within a few years of their admission.[213–215] By residing within government subsidized

institutions full-time, long-term care residents naturally require more healthcare than individuals in the community.[216] Linked population-based registries at ICES can be utilized to compare individuals in the general population with comparable linked CCHS respondents.

It is unlikely that all high-cost healthcare users outside of the household population are solely situated within institutional settings. Previous work (Chapter 6) has demonstrated that individuals in the community-dwelling population have a higher risk of mortality and long-term care admission than comparable individuals in the CCHS. This is likely due, in part, to retirement homes, which are a type of community-dwelling that provides supportive and lifestyle services to older segments of the population.[217] The degree to which retirement home residents are (ceteris paribus) more ill than other segments of the population is difficult to ascertain. In contrast to long-term care facilities, retirement homes are not typically registered or licensed, and there are no standardized assessments applied to their residents. Moreover, it is not possible to distinguish between residents of retirement homes and private dwellings within most population-based registries. It is generally assumed that these residents are not as ill as long-term care residents, since many eventually transition into long-term care once their needs can no longer be addressed by their retirement home.[190,218] The extent to which individuals in the community-dwelling population accumulate utilization and cost more rapidly than respondents in the CCHS is unclear. Linked population-based registries at ICES can also be utilized to compare individuals in the community-dwelling population with comparable linked CCHS respondents.

This study had two key objectives. The first objective was to assess whether individuals in the community-dwelling and general population demonstrated higher rates of utilization than comparable individuals in the CCHS across various different types of care. The second objective was to assess whether individuals in the community-dwelling and general population had higher rates of overall cost than comparable individuals in the CCHS. The community and general populations are naturally older than the household population. At the same time, the CCHS is designed to produce robust estimates at the health region level. As a result, it samples at much higher rates in sparsely populated regions, and much lower rates in heavily populated regions. A matched study design was employed utilizing basic demographics (birth year and sex) and health region (public health unit) to ensure that health differences between linked CCHS respondents

and their corresponding matches are comparable across cycles and not simply the product of demographics and geography.

7.3 DATA AND METHODS

Study Frames

Registered Person Database of Ontario

The Registered Person Database (RPDB) is a population-based registry maintained by the Ministry of Health and Long-Term Care in Ontario. It contains all unique health numbers that have been issued for Ontario Health Insurance Plan (OHIP), and their corresponding periods of coverage. All residents of Ontario are eligible for OHIP except for individuals living on First Nation Reserves, full-time members of the Canadian forces, and inmates of federal correctional facilities. These groups are also excluded from the CCHS sampling frame.

Linked Canadian Community Health Surveys

The CCHS is a cross-sectional population-based health survey routinely administered by Statistics Canada. The surveys collect information about health status, and health determinants for the household population 12 years of age and older, which represents 97% of the Canadian population. The survey employs a complex multistage sampling strategy and is conducted through telephone and in-person interviews. Excluded from the sampling frame are people living on First Nation Reserves, full-time members of the Canadian Forces, as well as residents of institutions and collective dwellings. The details of the survey methodology have been previously published and are available online.[130] Several of the Ontario subsamples representing respondents who agreed to share and link their survey content were linked to RPDB at ICES. The linked Ontario subsamples from 2007 to 2014 were combined to create an initial linked CCHS cohort of 133,983 respondents.

Health Services & Utilization Data

Multiple types of administrative data were linked to the study frames for the purpose of this study, which included records from the OHIP database, Ontario Drug Benefit Claims (ODB) database, Continuing Care Reporting System (CCRS) database, Discharge Abstract Database (DAD), Ontario Mental Health Reporting System (OMHRS) database, National Rehabilitation

Reporting System (NRS) database, Same-Day Surgery (SDS) database, National Ambulatory Care Reporting System (NACRS) database, and Home Care Database (HCD). The OHIP database contains claims to Ontario's publicly funded health insurance system by health care practitioners for inpatient and outpatient services. The ODB database contains prescription claims covered by the provincial drug program, which is restricted to those aged 65 and older, LTC residents, or service recipients of home care, social assistance and special drug programs. The CCRS database contains information for episodes of complex continuing care within hospitals and long-term care facilities, where individuals receive 24-hour nursing services. Reporting to CCRS database was only mandated for all long-term care facilities in 2009. The DAD contains information for all admissions and transfers to inpatient care within acute, rehabilitation, and chronic care institutions. The OMHRS database contains information for all admissions to adult inpatient hospital care beds designated for mental health.

Development of Cohorts

Community and general matched cohorts were created from RPDB to compare the health of different sampling frames to the CCHS cohort. The community cohort excluded long-term care residents, whereas the general cohort made no exclusions based on resident type. To facilitate the selection of appropriate matches, a population-based Ontario cohort for each survey year in the CCHS cohort (2003 to 2014) was created from RPDB. Each annual cohort was designed to represent the population of Ontarians 10 years of age or older within its respective year. To create a cohort for a particular year, all individuals from RPDB were selected whose records indicated: (1) An age of 10 years or older on January 1st of the year in question, (2) Ontario residency within the year, (3) contact with the healthcare system in the prior 12-year period, (4) OHIP coverage for a period of time within the year, and (5) a valid Ontario postal code for linkage to geographic indicators. Linked CCHS respondents who were not present within their corresponding annual Ontario cohort were excluded, reducing the linked CCHS cohort to 133,641 respondents, or 99.7% of the initial CCHS cohort. The annual Ontario cohorts ranged in size from 11,825,528 in 2007 to 12,737,298 in 2014.

Since an individual's residence type changes, an initial pool of potential matches was constructed for each CCHS respondent disregarding residence type. Once assigned to a CCHS respondent's pool, the residence type of a potential match could be determined at the time of the

CCHS respondent's survey date. For each CCHS respondent, 25 individuals were randomly selected from their corresponding annual Ontario cohort who matched on sex, birth year, and public health unit. The postal code conversion file was used to identify each individual's corresponding public health unit. CCHS respondents were removed from their own pool of potential matches in the rare case where they had been matched to themselves. Each potential match was assigned the same index date as their corresponding CCHS respondent (i.e., CCHS survey date). CCHS respondents and each of their potential matches were considered eligible for the matched cohorts if they demonstrated OHIP eligibility within 6 months of their index date, and contact with the health care system before (<=10 years) and after (<=6 years) their index date. Contact was defined as any form of healthcare use, registration or renewal of the OHIP card, or death. These criteria further reduced the linked CCHS cohort to 131,712 respondents, or 98.3% of the initial CCHS cohort. This process produced a potential pool of matches for each of these linked CCHS respondents that included both community-dwelling and long-term care residents.

Each potential match's type of residence (i.e., community-dwelling or long-term care) at the time of their index date was indirectly inferred from OHIP and ODB records in the 9 months before their index date. Long-term care residency could not be directly ascertained from CCRS since not all long-term care facilities reported to CCRS until late 2009. The OHIP and ODB records contain indicators that designate whether an individual is a long-term care resident. The first of two consecutive records with long-term care indicators were considered a long-term care admission, and the first of two consecutive records without long-term care indicators was considered a long-term care discharge. This approach was designed to limit miss-classification from a single false positive or false negative record. Each individual in a CCHS respondent's pool of potential matches was designated as either a community-dwelling or long-term care resident using these estimated long-term care admission and discharge dates. Potential matches from 2011 to 2014 whose residency status was directly known from the CCRS were used to validate this methodology. After reclassifying any individuals under the age of 35 as community-dwelling residents, the algorithm utilized to identify long-term care residents demonstrated good performance metrics in the validation set (sensitivity: 98.56, specificity: 99.97).

The matched cohorts were generated from the population sample of potential matches. For the community-dwelling matched cohort, five matches were randomly selected for each CCHS respondent from among its corresponding pool of potential matches who were community-dwelling residents. For the general matched cohort, five matches were randomly selected for each CCHS respondent from among its corresponding pool of potential matches irrespective of residence type. Given the rarity of obtaining multiple community-dwelling matches among very elderly CCHS respondents, CCHS respondents who were 95 years of age or older on January 1st of their survey year were excluded from the matching processes. This resulted in a final CCHS cohort of 131,712 respondents and matched cohorts of 658,560 individuals each. Individuals in each cohort were followed up to a maximum of five years and censored at either the beginning of any time period of 6 months or longer where they were OHIP ineligible, or death.

Baseline Characteristics

The individuals in each cohort were linked to baseline characteristics not used in the matching process to ascertain the comparability of the matched cohorts across important sociodemographic and disease indicators. The Postal Code Conversion File was used to identify each individual's corresponding neighbourhood income quintile and rurality index of Ontario (RIO) score. Pre-existing diseases were identified using healthcare utilization records. ICES-derived validated disease algorithms were utilized where available. The diseases with corresponding algorithms included dementia, asthma, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and hypertension.[131–136] Other diseases were identified based upon an in-hospital diagnosis or multiple relevant OHIP billings occurring within a specified number of years of each other. These diseases included renal disease, history of stroke, osteoporosis, heart arrhythmias, ischemic heart disease, history of cancer, osteoarthritis, and mental illness. Ten-year look back windows were used when evaluating the presence of pre-existing diseases without a validated algorithm. The sixteen identified diseases were summed and categorized (0-1, 2-3, 4-5, 6 or more) to provide an indicator of multi-comorbidity.

The Canadian Institute for Health Information's (CIHI) population grouping methodology was utilized to provide indicators of current severe illness and prior history of severe illness. This methodology was applied to records from the DAD, SDS database, NACRS

database, and OHIP database in a two year observation window around the index date (9 months before to 15 months after) to assess for the existence of severe illness at baseline, as well as the type of severe illness profile (palliative, cancer illness, mental illness, chronic illness, acute illness, or no severe illness).[191] This grouping methodology was also applied to OHIP records in a 10-year window ending 9 months before the index date to assess prior history of severe illness. Since the program for this methodology can only be applied to two-year intervals, it was applied in each two-year interval within the 10-year interval. Based on the number of intervals with severe illness, the prior history of severe illness for each individual was defined as extensive (4-5), moderate (2-3), minimal (1), or none.

Health Outcomes

Healthcare Utilization

Utilization associated with institutional care was assessed through the number of bed days in each institutional setting. This metric provided a more accurate measure of healthcare use in these settings than the number of admissions, since the average length of stay of admissions could vary immensely between cohorts. Chronic care utilization included long-term care bed days and complex continuing care bed days. Both forms of utilization are typically identified through the CCRS database. Since not all long-term care facilities reported to the CCRS database until the end of 2009, however, long-term care be days were estimated indirectly using OHIP and ODB records. Inpatient care utilization included acute, rehabilitation, and mental health hospital bed days. These forms of utilization were identified through DAD, the NRS database, and the OMHRS database respectively.

Utilization associated with non-institutional care was mainly assessed through the number of interactions with the healthcare system in each setting. Secondary care utilization included hospital clinic visits, specialist physician visits, and same-day procedures. These forms of utilization were identified through the DAD (i.e., general hospital clinic visits), NACRS database (i.e., cancer and dialysis hospital clinic visits), OHIP database, and SDS database. Each set of OHIP billings between an individual and a specialist on a particular day was considered a single specialist visit. Primary care utilization included primary care visits, home care visits, and emergency care visits. These forms of utilization were identified through the OHIP database, the HCD, and the NACRS database. Each set of OHIP billings between an individual and a primary

care professional (family physician, pediatrician, geriatrician, and other healthcare professional) on a particular day was considered a single primary care visit. Other utilization examined included drug prescription and laboratory testing, which were identified through the ODB and OHIP databases. OHIP billings of primary care professionals and specialists that occurred within another setting (e.g., acute hospitalization) did not contribute to primary care or specialist visits. The only exception was billings within long-term care since these visits would be beyond what is considered routine care in these facilities.

Healthcare Costs

ICES-derived costing methodologies were applied to these various types of utilization to generate corresponding overall healthcare costs.[127] For healthcare utilization funded by the global budgets of institutions, healthcare costs per utilization unit is estimated using a top-down approach that employs a case-mix methodology. In this scenario, the total annual cost of this system of a particular type of utilization is divided by its total annual weighted days to generate a corresponding cost per weighted day. An individual's annual cost with respect to a particular utilization is the product of its weighted days and cost per weighted days. The costs for utilization captured through the CCRS, DAD, NRS, OMHRS, NACRS, and SDS were estimated using this approach. Long-term care utilization indirectly estimated through OHIP and ODB were also estimated using this approach. The costs for utilization from the HCD, OHIP, and ODB were directly estimated from explicit fee payments, unless the healthcare professional involved in the encounter was compensated through a capitation system. In this latter case, the cost was estimated by substituting the fee schedule of payments (i.e., the fee utilized by professionals with explicit payments) or by the median fee paid for the particulate type of utilization. The costs associated with OHIP billings that were previously excluded from primary care professional and specialist visits were attributed to the setting in which they occurred (e.g., acute hospitalization). Costs were inflation-adjusted to 2018 and expressed in annual cost per capita terms.

Statistical Analysis

Baseline characteristics were compared between each matched cohort and the CCHS cohort. Standardized differences were computed for each baseline characteristic comparison. A p-value from a statistic test can provide insight into whether a significant difference exists, but

with sufficiently large samples, it will almost always demonstrate a significant difference.[192] In contrast, the standardized difference takes into account the effect size and is not dependent on sample size. It is increasingly used in observation studies to evaluate whether cohorts differ along baseline characteristics. A standardized difference greater than 0.1 is considered evidence of a significant difference.[193]

5-year rates of healthcare utilization were estimated for the CCHS cohort and each matched cohort using negative binomial distributions. Rate ratios of healthcare utilization between the CCHS cohort and each matched cohort were subsequently examined using negative binomial generalized linear models. A model was generated for each cohort comparison (i.e., CCHS vs. community-dwelling matched cohort, CCHS vs. general matched cohort) with regards to each individual type of utilization, and sector-specific utilization (chronic, inpatient, secondary, primary, and other). The cohort type was employed as the independent variable, and the regression was stratified for matched sets. An offset was employed to adjust for differences in follow-up time between individuals. 5-year rates of overall healthcare costs were estimated for the CCHS cohort and each matched cohort using gamma distributions. Rate ratios of overall healthcare costs between the CCHS cohort and each matched cohort were subsequently examined using gamma generalized linear models. The cohort type was employed as the independent variable, and the regression was stratified for matched sets. An offset was employed to adjust for differences in follow-up time between individuals. This latter analysis was repeated stratified by age subgroups (<65, 65-74, 75-84, 85-94) to examine whether differences were more prominent in older age groups.

7.4 RESULTS

Outside of expected differences in the proportion of long-term care residents, the CCHS cohort demonstrated similar overall baseline distributions to each matched cohort (See Table 7.1). The general cohort had a relatively small proportion of long-term care residents overall (1.0%). Rurality and neighbourhood income quintile distributions were not significantly different across the cohorts despite not being included in the matching process. The cohorts had good health, a reflection of their young age, which was not unexpected given that the matched cohorts were relatively young (72% under 65 years of age). Roughly 56% of each cohort had 1 or less

comorbidities, with hypertension consistently demonstrating the highest prevalence and dementia the lowest. Only about 14% of each cohort demonstrated severe illness at baseline, and roughly 18% of each cohort demonstrated any prior history of severe illness.

The estimated 5-year rates for individual types of healthcare utilization, as well as the corresponding model comparisons between the CCHS cohort and each matched cohort are presented in Tables 7.2 and 7.3. Within chronic care, the community and general cohorts demonstrated higher rates of bed days with regard to both long-term care and complex continuing care. Long-term care bed days were much higher in the general cohort, which was expected since this cohort included long-term care residents at baseline. Within inpatient care, the community and general cohorts demonstrated higher rates of bed days with regard to acute and mental health inpatient care. No differences were observed in rehabilitation inpatient care bed days. Within secondary care, all rates of utilization were lower among the matched cohorts. This was not unexpected since the most common same-day procedures are not related to illness but rather indicators of good health (e.g. child birth), or preventative health measures (e.g. colorectal screening).[219] In addition, the greater amount of time spent within chronic and inpatient care by individuals in the matched cohorts would have reduced opportunities to visit specialists and hospital clinics. Within primary care, the community and general cohorts demonstrated higher rates of utilization with regard to general and emergency care visits. General care visits and their costs were much higher in the general cohort. There was no difference with regard to home care visits. This was not unexpected since long-term care residents within the general cohort would have not required home care. Within other care, the community and general cohorts demonstrated higher rates of drug prescriptions, with the prescription rate much higher in the general cohort. There was no difference with regard to laboratory tests.

The estimated 5-year rates for sector-specific healthcare utilization, as well as the corresponding model comparisons between the CCHS cohort and each matched cohort are presented in Tables 7.2 and 7.3. Chronic care utilization was higher in both matched cohorts, with differences more pronounced in the general cohort. These differences were observed across all ages, including the non-elderly (<65). This suggests that there are a significant number of individuals in the matched cohorts who are more ill than CCHS respondents but not within

retirement homes or long-term care facilities. Inpatient care and other care utilization were higher in both matched cohorts, but only among elderly age groups (>65). The matched cohorts also demonstrated higher primary care utilization within the two oldest age groups (≥75). Contrary to these trends but mirroring prior analysis of individuals secondary utilization types, the matched cohorts demonstrated lower secondary care utilization than the CCS cohort.

Analysis of overall healthcare costs are presented in Table 7.6 and 7.7. Higher rates of overall healthcare costs were observed in both matched cohorts, with differences more pronounced in the general cohort. There were no significant differences among non-elderly (<65), which could be expected since the chance of being in a retirement home or long-term care facility at those ages is incredibly rare. Within the elderly age groups, differences increased with age. These overall trends coincided with the observed sector-specific utilization differences within chronic, inpatient, primary, and other care.

7.5 DISCUSSION

Analysis of utilization and healthcare costs within this study provided as assessment of health differences between linked CCHS respondents and comparable individuals in two matched cohorts. Differences follow a pattern indicative of health differences between the sampling frames. Linked CCHS respondents accumulated utilization and healthcare costs less rapidly at older ages where we would expect residents of retirement homes and long-term care facilities to be concentrated within the other sampling frames. Differences were most evident relative to the general matched cohort, which reflects the inclusion of residents in retirement and long-term care facilities that were excluded from the CCHS sampling frame. The differences between the community cohort and CCHS cohort imply that long-term care residents are likely not the only contributing factor to these differences, and that retirement home residents also play a role. Differences among younger people (under 65 yrs.) suggest that there are also ill and costly segments of the population not captured within the CCHS that cannot be explained by retirement home and long-term care residents. This could be a product of individuals who are incarcerated, residents of group homes (i.e., for the disabled, addicts, etc.), or homeless; all segments of the wider population that are excluded from the household population. It is also plausible that this may be a product of health-related non-response biases among the younger respondents. Caution

should be exercised when using the CCHS to make population-wide references that go beyond its target population, including the non-elderly.

The comprehensive assessment of utilization and healthcare cost differences between linked CCHS respondents and other sampling frames in this study goes beyond existing examinations of other large population-based health surveys that are routinely administered, or even within other provincial subsamples of the CCHS. Linkages of the National CCHS to hospitalization records has demonstrated similar differences to the general population at older ages, but widespread linkages to multiple administrative healthcare databases is limited within other provinces.[220] Without more extensive linkages, estimating overall healthcare costs or even the healthcare costs of large sectors is a challenge. These challenges are even more notable in other large national population-based health surveys, where the lack of linkages has often limited comparisons to a few select outcomes (e.g., mortality and hospitalization). Extensive linkage of national population healthy surveys is still not widely practiced.[221] Moreover, all the data holdings required to undergo such an analysis are not typically available within the same repository.[222] Given differences in healthcare systems between nations, the extent to which differences in utilization and healthcare costs exist between household survey respondents and wider population frames exist, and what drive those differences, may vary. The Universal health-care model followed in Canada may opportunistically allow for a more well posed examination of the role of non-response bias.

There are several areas of interest for future research, which are primarily a product of the limitations within this study. The lack of a viable representation of the household population, the CCHS target population, was a central limitation. Differences in utilization and healthcare costs between linked CCHS respondents and individuals in each sampling frame may be a combination of expected differences between individuals in each sampling frame, as well as health-related response non-biases in linked CCHS respondents. As linkages to different types of administrative databases continue to expand, our ability to properly isolate the household population within population registries will improve. Several studies have already successfully identified residents of correctional facilities and retirement homes within population registries, two populations that are not part of the household population but could not be identified and excluded from this study's comparative sampling frames.

The nature of these differences in utilization and healthcare costs may also be more complicated than presented in this study. Differences in absolute utilization or healthcare costs may vary according to follow-up. This study did not examine absolute differences but rather rates of accumulation that adjusted for differences in follow-up. Given the observed differences in mortality and accumulation of healthcare costs between linked CCHs respondents and individuals in the wider sampling frames, the degree to which utilization and healthcare costs differ within a given interval of follow-up is likely to vary. Researchers using a longitudinal framework (e.g. lifetime) need to take into account that for many CCHS respondents, the most costly stages of their lifespan are unlikely to occur within the observation window available, nor will they all remain part of the household population in perpetuity.[169] In order to accurately estimate lifetime costs, more advanced methods are required that take into account changes in mortality and cost risks over time.[148,153,223] Understanding the trajectory of utilization and healthcare costs of linked CCHS respondents relative to wider sampling frames could help inform these areas of research. After enough follow-up, CCHS linked respondents may actually be a better representation of the general population than the household population.

Overall, this study provided unique insights into differences in utilization and healthcare costs between linked CCHS respondents and wider population frames, and demonstrated that they closely align with previously presented health differences. Since most health outcomes are typically rare and heavily concentrated among the elderly, utilization and healthcare costs may provide better insights into population differences that may be otherwise undetectable.

7.6 TABLES

Table 7.1. Characteristics of CCHS cohort and Matched cohorts

Characteristics	CCHS	Community	General
	N=131,712	N=658,560	N=658,560
Survey Cycle			
2003-2004	25.8%	25.8%	25.8%
2005-2006	25.1%	25.1%	25.1%
2007-2008	24.4%	24.4%	24.4%
2009-2010	24.7%	24.7%	24.7%
Sex			
Male	45.0	45.0	45.0
Female	55.0	55.0	55.0
Age Group			
Under 65	72.2%	72.2%	72.2%
65 to 74	14.9%	14.9%	14.9%
75 to 84	9.8%	9.8%	9.8%
85 to 94	3.1%	3.1%	3.1%
Neighborhood Income Quintile			
Q1 (lowest)	18.9%	19.1%	19.1%
Q2	19.8%	19.7%	19.7%
Q3	20.3%	20.4%	20.4%
Q4	20.9%	20.8%	20.7%
Q5 (highest)	20.2%	20.1%	20.1%
Rurality Index of Ontario			
Rural	16.9%	17.1%	17.2%
Suburban	30.5%	30.2%	30.2%
Urban	52.6%	52.7%	52.7%
Long-term Care Resident			
Yes	0.0%	0.0%	***1.0%***
No	100.0%	100.0%	99.0%
Comorbidities			
Dementia	1.0%	1.4%	2.0%
Renal Disease	2.2%	2.1%	2.2%
History of Stroke	2.0%	2.1%	2.3%
Asthma	2.3%	2.2%	2.3%
Congestive Heart Failure	2.9%	2.9%	3.1%
Osteoporosis	3.8%	3.7%	3.7%
Heart Arrhythmias	5.3%	5.3%	5.4%
Chronic Obstructive Pulmonary Disease	9.9%	9.6%	9.8%
Ischemic Heart Disease	10.2%	10.1%	10.1%
Diabetes	11.5%	11.8%	11.8%

History of Cancer	24.2%	22.9%	22.7%
Hypertension	31.2%	31.0%	30.9%
Mental Illness	30.5%	30.0%	30.3%
Osteoarthritis	32.9%	32.5%	32.4%
Multiple Comorbidities			
6 or More	3.7%	3.8%	4.0%
4 to 5	11.4%	11.1%	11.1%
2 to 3	29.0%	28.6%	28.4%
1 or less	55.9%	56.5%	56.4%
Current Severe Illness Profile			
Palliative	0.5%	0.6%	0.6%
Cancer Illness	2.3%	2.2%	2.2%
Mental Illness	2.0%	2.2%	2.6%
Chronic Illness	5.4%	5.2%	5.2%
Acute Illness	3.8%	3.8%	3.8%
None	86.0%	86.0%	85.7%
Prior History of Severe Illness			
Extensive	1.9%	1.9%	2.0%
Moderate	5.7%	5.4%	5.7%
Minimal	10.4%	10.3%	10.3%
None	82.0%	82.4%	81.9%

Italicized Bold: Standardized Difference >0.1 in comparison to CCHS cohort

Table 7.2. Estimated 5-Year Healthcare Utilization Rates among the CCHS cohort and each Matched Cohort using Negative Binomial Distributions

Type of Utilization	CCHS	Community	General
Chronic Care Bed Days			
Long-term Care	9.31 (8.81 - 9.83)	12.91 (12.58 - 13.24)	28.88 (28.25 - 29.51)
Complex Continuing Care	1.45 (1.33 - 1.59)	2.09 (2.00 - 2.19)	2.10 (2.01 - 2.21)
Inpatient Care Bed Days			
Acute Hospitals	6.69 (6.50 - 6.88)	7.35 (7.23 - 7.46)	7.32 (7.21 - 7.44)
Mental Health Hospitals	0.55 (0.48 - 0.63)	0.83 (0.77 - 0.88)	0.91 (0.85 - 0.97)
Rehabilitation Hospitals	0.68 (0.63 - 0.72)	0.67 (0.64 - 0.69)	0.65 (0.62 - 0.67)
Secondary Care Encounters			
Specialist Care Visits	22.36 (22.13 - 22.59)	21.44 (21.31 - 21.55)	21.32 (21.20 - 21.44)
Hospital Clinic Visits	6.15 (6.01 - 6.29)	5.95 (5.88 - 6.02)	5.92 (5.85 - 5.99)
Same-day Procedures	0.75 (0.75 - 0.76)	0.72 (0.71 - 0.72)	0.71 (0.71 - 0.72)
Primary Care Encounters			
General Care Visits	31.16 (30.91 - 31.41)	32.20 (32.04 - 32.37)	34.51 (34.30 - 34.72)
Home Care Visits	20.25 (19.63 - 20.89)	20.95 (20.60 - 21.31)	19.97 (19.64 - 20.31)
Emergency Care Visits	2.81 (2.78 - 2.84)	2.89 (2.87 - 2.90)	2.88 (2.87 - 2.90)
Other Encounters			
Drug Prescriptions	65.82 (64.55 - 67.13)	73.41 (72.48 - 74.35)	83.53 (82.46 - 84.62)
Laboratory Tests	63.87 (63.42 - 64.32)	63.20 (62.93 - 63.46)	64.09 (63.81 - 64.37)

Table 7.3. Estimated Relative Rates of Utilization between each Matched cohort and the CCHS cohort using Negative Binomial Regressions

Type of Utilization	Community versus CCHS	General versus CCHS
Chronic Care Bed Days		
Long-term Care	1.39 (1.33, 1.45)	3.10 (3.04, 3.16)
Complex Continuing Care	1.44 (1.34, 1.54)	1.45 (1.35, 1.55)
Inpatient Care Bed Days		
Acute Hospitals	1.10 (1.07, 1.13)	1.09 (1.06, 1.12)
Mental Health Hospitals	1.51 (1.36, 1.66)	1.65 (1.50, 1.80)
Rehabilitation Hospitals	0.99 (0.91, 1.07)	0.96 (0.88, 1.04)
Secondary Care Encounters		
Specialist Care Visits	0.96 (0.95, 0.97)	0.95 (0.94, 0.96)
Hospital Clinic Visits	0.94 (0.93, 0.95)	0.94 (0.93, 0.95)
Same-day Procedures	0.95 (0.94, 0.96)	0.95 (0.94, 0.96)
Primary Care Encounters		
General Care Visits	1.03 (1.02, 1.04)	1.11 (1.10, 1.12)
Home Care Visits	1.04 (1.01, 1.07)	0.99 (0.96, 1.02)
Emergency Care Visits	1.03 (1.02, 1.04)	1.03 (1.02, 1.04)
Other Care Encounters		
Drug Prescriptions	1.12 (1.10, 1.14)	1.27 (1.25, 1.29)
Laboratory Tests	0.99 (0.98, 1.00)	1.00 (0.99, 1.01)

Table 7.4. Estimated Sector-specific 5-Year Healthcare Utilization Rates among the CCHS cohort and each Matched Cohort using Negative Binomial Distributions and Stratified by Age Group

Sector-specific Utilization	CCHS	Community	General
Chronic Care Bed Days			
All Ages	10.8 (10.2 - 11.3)	15.0 (14.7 - 15.4)	31.0 (30.4 - 31.7)
Under 65	1.0 (0.8 - 1.2)	1.4 (1.3 - 1.5)	2.6 (2.4 - 2.7)
65 to 74	10.5 (9.3 - 11.9)	13.1 (12.4 - 13.9)	26.2 (25.0 - 27.4)
75 to 84	45.1 (41.8 - 48.7)	63.6 (61.7 - 65.7)	129.0 (125.7 - 132.4)
85 to 94	131.6 (121.6 - 142.4)	187.3 (181.5 - 193.2)	406.2 (396.2 - 416.4)
Inpatient Care Bed Days			
All Ages	8.0 (7.7 - 8.2)	8.9 (8.8 - 9.0)	8.9 (8.8 - 9.1)
Under 65	3.3 (3.1 - 3.4)	3.4 (3.3 - 3.5)	3.5 (3.4 - 3.6)
65 to 74	12.3 (11.6 - 13.0)	13.3 (12.9 - 13.6)	13.8 (13.4 - 14.2)
75 to 84	25.0 (23.9 - 26.2)	29.7 (28.9 - 30.4)	29.9 (29.1 - 30.6)
85 to 94	44.1 (41.4 - 47.0)	53.78(52.0 - 55.6)	49.3 (47.6 - 51.0)
Secondary Care Encounters			
All Ages	29.4 (29.1 - 29.8)	28.3 (28.1 - 28.5)	28.1 (28.0 - 28.3)
Under 65	19.5 (19.2 - 19.8)	18.5 (18.4 - 18.7)	18.6 (18.4 - 18.7)
65 to 74	50.8 (49.6 - 52.0)	49.9 (49.4 - 50.5)	49.6 (49.1 - 50.1)
75 to 84	62.4 (60.8 - 64.0)	59.7 (59.0 - 60.4)	59.6 (58.9 - 60.4)
85 to 94	56.5 (54.2 - 59.0)	55.9 (54.7 - 57.2)	52.3 (51.1 - 53.6)
Primary Care Encounters			
All Ages	54.1 (53.4 - 54.9)	55.9 (55.4 - 56.3)	57.4 (56.9 - 57.8)
Under 65	29.8 (29.4 - 30.3)	29.4 (29.3 - 29.6)	29.9 (29.7 - 30.1)
65 to 74	70.6 (68.6 - 72.7)	70.2 (69.3 - 71.2)	72.0 (71.1 - 73.0)
75 to 84	142.3 (138.2 - 146.5)	150.7 (148.6 - 152.8)	159.9 (157.7 - 162.1)
85 to 94	273.8 (263.1 - 284.9)	317.2 (311.3 - 323.1)	321.6 (316.0 - 327.3)
Other Care Encounters			
All Ages	129.6 (128.1 - 131.1)	136.5 (135.4 - 137.6)	147.4 (146.1 - 148.6)
Under 65	46.7 (46.3 - 47.1)	45.8 (45.5 - 46.0)	45.8 (45.6 - 46.0)
65 to 74	256.6 (251.8 - 261.5)	265.4 (263.1 - 267.8)	277.4 (274.8 - 280.0)
75 to 84	405.5 (397.1 - 414.0)	436.6 (432.3 - 441.0)	491.5 (486.5 - 496.5)
85 to 94	585.0 (566.8 - 603.8)	687.7 (677.6 - 698.0)	812.0 (800.6 - 823.7)

Table 7.5. Estimated Relative Rates of Sector-specific Utilization between each Matched cohort and the CCHS cohort using Negative Binomial Regressions and Stratified by Age Group

Sector-specific Utilization	Community versus CCHS	General versus CCHS
Chronic Care Bed Days		
All Ages	1.40 (1.35, 1.45)	2.88 (2.83, 2.93)
Under 65	1.47 (1.26, 1.68)	2.68 (2.48, 2.88)
65 to 74	1.25 (1.11, 1.39)	2.48 (2.35, 2.61)
75 to 84	1.41 (1.33, 1.49)	2.86 (2.78, 2.94)
85 to 94	1.42 (1.34, 1.50)	3.09 (3.01, 3.17)
Inpatient Care Bed Days		
All Ages	1.12 (1.09, 1.15)	1.12 (1.09, 1.15)
Under 65	1.04 (0.98, 1.10)	1.07 (1.01, 1.13)
65 to 74	1.08 (1.02, 1.14)	1.12 (1.06, 1.18)
75 to 84	1.19 (1.14, 1.24)	1.19 (1.14, 1.24)
85 to 94	1.22 (1.15, 1.29)	1.12 (1.05, 1.19)
Secondary Care Encounters		
All Ages	0.96 (0.95, 0.97)	0.96 (0.95, 0.97)
Under 65	0.95 (0.93, 0.97)	0.95 (0.93, 0.97)
65 to 74	0.98 (0.95, 1.01)	0.98 (0.95, 1.01)
75 to 84	0.96 (0.93, 0.99)	0.96 (0.93, 0.99)
85 to 94	0.99 (0.94, 1.04)	0.93 (0.88, 0.98)
Primary Care Encounters		
All Ages	1.03 (1.02, 1.04)	1.06 (1.05, 1.07)
Under 65	0.99 (0.97, 1.01)	1.00 (0.98, 1.02)
65 to 74	0.99 (0.96, 1.02)	1.02 (0.99, 1.05)
75 to 84	1.06 (1.03, 1.09)	1.12 (1.09, 1.15)
85 to 94	1.16 (1.12, 1.20)	1.17 (1.13, 1.21)
Other Care Encounters		
All Ages	1.05 (1.04, 1.06)	1.14 (1.13, 1.15)
Under 65	0.98 (0.97, 0.99)	0.98 (0.97, 0.99)
65 to 74	1.03 (1.01, 1.05)	1.08 (1.06, 1.10)
75 to 84	1.08 (1.06, 1.10)	1.21 (1.19, 1.23)
85 to 94	1.18 (1.15, 1.21)	1.39 (1.36, 1.42)

Table 7.6. Estimated 5-Year Total Healthcare Cost Rates among the CCHS cohort and each Matched Cohort using Gamma Distributions and Stratified by Age Group

Age Group	CCHS	Community	General
All Ages	26,900 (26,427 - 27,382)	29,132 (28,813 - 29,455)	31,435 (31,098 - 31,776)
Under 65	12,783 (12,399 - 13,179)	12,687 (12,494 - 12,884)	12,959 (12,773 - 13,148)
65 to 74	45,692 (44,201 - 47,234)	48,745 (47,818 - 49,691)	51,590 (50,582 - 52,618)
75 to 84	75,120 (72729 - 77,590)	85,874 (84,214 - 87,567)	94,771 (93,169 - 96,400)
85 to 94	112,140 (107,492 - 116,989)	136,008 (132,944 - 139,142)	161,760 (158,283 - 165,313)

Table 7.7. Estimated Relative Rates of Total Healthcare Costs between each Matched cohort and the CCHS cohort using Gamma Regressions and Stratified by Age Group

Age Group	Community versus CCHS	General versus CCHS
All Ages	1.08 (1.06, 1.10)	1.17 (1.15, 1.19)
Under 65	0.99 (0.96, 1.02)	1.01 (0.98, 1.04)
65 to 74	1.07 (1.03, 1.11)	1.13 (1.09, 1.17)
75 to 84	1.14 (1.10, 1.18)	1.26 (1.22, 1.30)
85 to 94	1.21 (1.16, 1.26)	1.44 (1.39, 1.49)

8. Conclusions, Implications, and Future Directions

This thesis represents an important initial step in developing a framework that can evaluate the burden of unhealthy behaviours over a lifetime through linked population-based health surveys. Unhealthy behaviours simultaneously influence many different facets in an individual's life, such as their risk of mortality, disability, and healthcare costs. As a result, disentangling the impact of unhealthy behaviours over a lifetime is a challenge. Understanding the degree to which life expectancy, lifetime disability and lifetime healthcare costs vary as a function of unhealthy behaviours could help us understand the conditions under which compression and expansion of morbidity occur.

The first two studies leveraged linkages of a routinely administered national health survey (i.e., the CCHS) to administrative healthcare databases to evaluate the impact of unhealthy behaviours directly on mortality and healthcare costs. In the first study, multivariable risk algorithms for mortality were developed to estimate the burden of unhealthy behaviours on period life expectancy through a cause-deleted approach. As expected, unhealthy behaviours predicted higher mortality risk, and lowered the average life expectancy of the population. In the second study, multivariable risk algorithms for phase-based healthcare costs based on proximity to death were developed that could be combined with the aforementioned risk algorithms for mortality. The cause-deleted approach in this study provided counterfactual estimates where the impact of unhealthy behaviours had been removed from life expectancy and healthcare costs within each phase of life. The results of this study were more complex, as the impact of unhealthy behaviours on lifetime healthcare costs varied; some increased lifetime healthcare costs and other decreased them. The results of these studies, however, were far from conclusive given the limitations associated with measures from cross-sectional health surveys and period life table approaches. Without accounting for temporal relationships between unhealthy behaviours and disability over time, or incorporation of realistic trajectories for unhealthy behaviours and other factors, these cause-deleted approaches are over simplistic.

The third study introduced a methodology by which to generate a quasi-longitudinal CCHS cohort that could be leveraged for transition risk algorithms in an effort to move beyond period life tables. Linked CCHS samples roughly four years apart were utilized to approximate transition cohorts for five different time periods. Imputation models were utilized to generate

missing survey information either at the beginning or end of each respondent's four-year window. The resulting quasi-longitudinal CCHS cohort was utilized to develop multivariable transition risk algorithms for transitions towards immobility. Unhealthy behaviours strongly predicted transitions towards moderate and severe immobility. The quasi-longitudinal CCHS cohort also allowed for transitions of health behaviours and several sociodemographic factors. As a result, multivariable transition risk algorithms could also potentially be developed for these other exposures. This provides opportunities to move beyond period life tables and towards cohort life tables. Rather than utilizing the experience of different birth cohorts to generate period life tables, each birth cohort can have exposure trajectories reflective of its own baseline exposures. Applying the multivariable transition algorithms to the baseline exposures of a birth cohort would generate distributions of theses exposures at a future point in time. Iteratively applying these algorithms to each subsequent forecasted distribution of exposures would generate lifetime trajectories for these exposures that are more realistic for the birth cohort in question.

The last two studies evaluated the representativeness of linked CCHS respondents for population-based studies. Response and consent (to linkage) rates in the CCHS have been declining since its introduction raising concerns surrounding the comparability of CCHS samples over time. Similar to other population-based surveys, survey weights are provided that are designed address biases that may arise from non-response and non-consent to linkage. CCHS-base multivariable risk algorithms (as in this thesis), however, are frequently derived from pooled unweighted CCHS samples. Relative to wider sampling frames, unweighted linked CCHS samples were observed to be comparable over time. Nevertheless, linked CCHS respondents were observed to be healthier than comparable individuals in the community-dwelling and general populations at older ages, where they demonstrated lower risks of mortality, long-term care admission, and healthcare costs. This was not unexpected given that important segments of the population (e.g., residents of retirement homes and long-term care care) are excluded from the CCHS sampling frame. These studies highlighted the difficulties of estimating life expectancy and corresponding lifetime healthcare costs from the household population, and the necessity to ensure estimates realistically incorporate the time individuals may live outside of the household population over a lifetime. Incorporating simple calibrations of

linked outcomes to account for the sampling frame is likely an over simplistic and inaccurate approach.

Future research can improve the framework outlined in this thesis to generate lifetime estimates by developing multivariable risk algorithms for transitions to important dwelling types outside of the household population. Admission into long-term care facilities for CCHS respondents can be readily identified through linked administrative healthcare databases, as evidenced by the fourth study in this thesis. Dementia and functional disability are common predictors of long-term care admission, both of which are predicted by unhealthy behaviours.[82,169,224] Very recently, research studies at ICES have been able to examine residents of retirement homes in Ontario through linkages of RPDB to the Retirement Homes Regulatory Authority of Ontario's public register of licensed retirement homes.[225] This implies that admission into retirement homes for CCHS respondents can potentially also be identified. Given the similarities of residents in retirement homes and long-term care, it would be not be surprising if unhealthy behaviours are also predictive of admission into retirement homes. Multivariable risk algorithms for admission into long-term care and retirement homes would improve the accuracy of cohort life table estimates by providing adjusted age-specific probabilities of admission that could be incorporated into the life tables. CCHS-based mortality risk algorithms could then incorporate time-varying exposures for long-term care and retirement home residency, and phase-based multivariable risk algorithms for healthcare costs could be developed for each dwelling type separately. Estimates of life expectancy and lifetime healthcare costs would therefore better reflect actual lifetime estimates where many individuals naturally spend a portion of their lifespans outside of the household population.

Linkages to retirement home and long-term care admission could also help improve the methodology utilized to create the quasi-longitudinal CCHS cohorts in the third study of this thesis. The quasi-longitudinal CCHS cohorts were generated through imputation without any external linked information and assuming a closed household population. For imputations that generated future exposures, this is unlikely to be true. In reality, many survey respondents would have transitioned to retirement homes and long-term care facilities by the end of the transition periods being observed. As a result, CCHS samples that were utilized to represent individuals at the end of transition periods would have been poor representations of reality. The imputations

projecting future exposures likely underestimated transitions that were associated with poor health (e.g., immobility); failing to take into account that the projected base CCHS sample should be unhealthier than the targeted CCHS sample. Imputations projecting past exposures are more realistic since survey respondents are unlikely to re-enter the household population from retirement homes and long-term care facilities. Linkages to retirement homes and long-term care facilities would allow individuals who are no longer part of the household population at the end of the period of observation to be removed. This would lead to multivariable transition risk algorithms that are more reflective of an actual household population.

Inclusion of linked information such as comorbidities (e.g., individual diseases), summary comorbidity scores (e.g., CIHI population grouping methodology), and healthcare costs have the potential to improve the specificity of imputations to generate quasi-longitudinal CCHS cohorts. Naturally, transitions are likely associated with the health of survey respondents which can be captured through the linked information. For example, individuals with a multitude of comorbidities are more likely to become immobile over time than individuals with few or no comorbidities.[80] The linked information can also help imputation models to generate future exposures that are reflective of differences between the forecasted base CCHS samples and the targeted CCHS samples. After all, it may not be always possible to remove all CCHS respondents who are ineligible to participate in subsequent CCHS samples from the hypothetical CCHS transition cohorts. Given the intense computing requirements of such imputation methods, a higher performing computing environment will be required. The quasi-longitudinal CCHS cohort developed in this thesis required approximately 1000 hours to be generated within the standard SAS environment at ICES. A more advanced version which incorporates transitions across multiple different periods of time (e.g., 2, 4, 6, 8 years etc.), and that utilizes linked longitudinal information is impractical within this environment. ICES now has a high-performance computing environment where artificial intelligence can be utilized to conduct data analysis. This environment would allow for exploration of these more advanced methods. Given the sheer multitude of available linked data that can be selected and the variety of options, machine learning approaches may be required to optimize these processes.

Once all the components of this framework are complete – multivariable risk algorithms for mortality, transitions, and healthcare costs etc. – explicit methods by which to combine and

validate these components will be required. Unlike period life table approaches, this framework should in theory be able to project mortality and healthcare costs beyond the observation window utilized to develop the risk algorithms. It is typical to derive and validate linked CCHS-based algorithms by utilizing a split sample of the linked CCHS samples, or by deriving on one set of linked CCHS samples and validating on another set of linked CCHS samples. These multivariable risk algorithms are never evaluated beyond the defined follow-up window utilized in their derivation, nor are they expected to work beyond this follow-up window. These algorithms can only predict risk based on the set of exposures available at baseline for a specified window that is typically rather short (e.g., 5 years) and not indicative of a lifetime. It is these limitations that necessitate the use of period life table approaches which use different birth cohorts to describe the trajectory of baseline exposures over a lifetime. If the multivariable risk algorithms for different transitions are able to update the baseline exposures across the follow-up window, then in theory, the multivariable risk algorithms for mortality and healthcare costs should be able to predict the associated risks in the next observation window. Moreover, multivariable transition risk algorithms for admission into retirement homes and long-term care facilities should allow the household population to project into different dwelling types over time, which can then be addressed independently of the household population. Once this framework is able to properly project mortality and healthcare costs, cause-deleted approaches can be applied to generate counterfactuals estimates where the impact of unhealthy behaviours on mortality, transitions, and healthcare costs is theoretically removed.

Given the amount of work associated with updating this entire framework, many processes will need to be automated to make future implementations of this framework feasible. There are several important exposures that warrant future examination and could be included in future updates. The impact of diet was likely underestimated in all the studies within this thesis since it was only captured by the CCHS through frequency intake of fruits and vegetables. Two specialized cycles of the CCHS are available that specifically target dietary consumption and could provide a better assessment of poor diet on mortality, transitions, and healthcare costs. Only one of these specialized cycles are linked at ICES, but the unlinked specialized cycle should be linked relatively soon. The one unhealthy behaviour that is available in the CCHS but was excluded completely from this thesis was alcohol consumption. Previous link CCHS-based studies have demonstrated counter-intuitive associations between alcohol consumption and

linked outcomes such as mortality, likely the product of challenges with reverse-causality.[45] As individuals age and become ill, alcohol consumption typically decreases and many individuals who would have exhibited drinking behaviours earlier in life become non-drinkers.[226,227] As a result, non-drinkers appear by far the unhealthiest drinking exposure when examining linked health outcomes. The use of linked information in the development in the quasi-longitudinal CCHS cohorts and incorporation of health states (e.g., disability and illness) into the framework outlined in this thesis may be able to address these challenges in the future. Household income was not included in any of the risk algorithms in this thesis because measures of household income that are adjusted for inflation are not currently available across all the linked CCHS at ICES. It may also be more feasible and accurate to conduct such analysis at Statistics Canada which has the CCHS linked to income tax files over time. As the CCHS changes over time and additional exposures of interest are ascertained, there will be opportunities to incorporate more exposures. Given the self-reported nature of CCHS information, extensive validation of exposures currently employed in this thesis also requires further consideration.

Although this thesis focused on developing a framework to generate lifetime estimates from the linked CCHS, the methodologies outlined here are also relevant to other linked national health surveys. Routinely administered national health surveys are common in a multitude of countries. Extensive linkage of these national health surveys is still a rather novel advancement that is only common in various developed countries, but this is likely to change with time as the infrastructure for linkage becomes more commonplace and affordable. Generating lifetime estimates through a framework that leverages these linkages to assess the impact of survey exposures (e.g., unhealthy behaviours and sociodemographic factors) on health outcomes, and uses linked information to convert linked national health surveys into quasi-longitudinal cohorts is a complex but worthwhile endeavour. It is unlikely that longitudinal studies that capture health behaviours, sociodemographic factors, and health outcomes will ever become commonplace given the vast resources required to conduct such studies. Moreover, the sample size of longitudinal studies will always pale in comparison to pooled linked national health surveys that over time will eventually include millions of individuals. Continued linkage of these surveys and the rapid expansion of high-computing environments suggest that this area of research has the potential to generate findings that historically would have not been feasible.

9. REFERENCES

1. Roser M. Life Expectancy — Our World in Data [Internet]. 2015 [cited 2020 Jan 14]. Available from: https://ourworldindata.org/life-expectancy

2. Riley JC. Estimates of regional and global life expectancy, 1800-2001. Popul Dev Rev. 2005 Sep;31(3):537–43.

3. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Burden by Risk 1990-2017 [Internet]. 2018 [cited 2020 Jan 17]. Available from: http://ghdx.healthdata.org/record/ihme-data/gbd-2017-burden-risk-1990-2017

4. Armstrong GL, Conn LA, Pinner RW. Trends in infectious disease mortality in the United States during the 20th century. J Am Med Assoc. 1999 Jan 6;281(1):61–6.

5. Wegman ME. Infant Mortality in the 20th Century, Dramatic but Uneven Progress. J Nutr. 2001 Feb 1;131(2):401S-408S.

6. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: A systematic review of the literature. Vol. 10, Ageing Research Reviews. 2011. p. 430–9.

7. Kingston A, Robinson L, Booth H, ageing MK-… and, 2018 undefined. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. academic.oup.com [Internet]. [cited 2020 Feb 29]; Available from: https://academic.oup.com/ageing/article-abstract/47/3/374/4815738

8. Guilbert JJ. The world health report 2002 - Reducing risks, promoting healthy life [2]. Vol. 16, Education for Health. World Health Organization; 2003. p. 230.

9. Brown GC. Living too long. EMBO Rep. 2015 Feb;16(2):137–41.

10. Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. Vol. 12, Preventive Medicine Reports. Elsevier Inc.; 2018. p. 284–93.

11. Rice DP, Feldman JJ. Living longer in the United States: demographic changes and health needs of the elderly. Milbank Meml Fund Quarterly, Heal Soc. 1983 Jun;61(3):362–96.

12. Fries JF. Aging, Natural Death, and the Compression of Morbidity. N Engl J Med. 1980 Jul 17;303(3):130–5.

13. Oeppen J, Vaupel JW. Demography: Broken limits to life expectancy. Vol. 296, Science. 2002. p. 1029–31.

14. Gruenberg EM. The Failures of Success. Milbank Mem Fund Q Health Soc [Internet]. 1977 [cited 2020 Jan 19];55(1):3. Available from: https://www.jstor.org/stable/3349592?origin=crossref

15. Kramer M. The rising pandemic of mental disorders and associated chronic diseases and disabilities. Acta Psychiatr Scand [Internet]. 1980 Sep [cited 2020 Jan 19];62(285 S):382–97. Available from: http://doi.wiley.com/10.1111/j.1600-0447.1980.tb07714.x

16. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin [Internet]. 2020 Jan 8 [cited 2020 Jan 19];70(1):7–30. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21590

17. Mensah GA, Wei GS, Sorlie PD, Fine LJ, Rosenberg Y, Kaufmann PG, et al. Decline in Cardiovascular Mortality: Possible Causes and Implications. Vol. 120, Circulation Research. Lippincott Williams and Wilkins; 2017. p. 366–80.

18. Lichtenberg FR. The impact of new drug launches on longevity: Evidence from longitudinal, disease-level data from 52 countries, 1982-2001. Int J Health Care Finance Econ [Internet]. 2005 Jan [cited 2020 Mar 1];5(1):47–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15714263

19. Manton KG. Changing concepts of morbidity and mortality in the elderly population. Milbank Meml Fund Quarterly, Heal Soc. 1982 Mar;60(2):183–244.

20. He W, Goodkind D, Kowal P. An Aging World [Internet]. International Population Reports. 2016 [cited 2020 Feb 1]. Available from: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf

21. Canada G of. Government of Canada — Action for Seniors report [Internet]. Government of Canada. 2014 [cited 2020 Feb 1]. Available from: https://www.canada.ca/en/employment-social-development/programs/seniors-action-report.html

22. Health at a Glance 2019 [Internet]. OECD; 2019 [cited 2020 Jan 20]. (Health at a Glance). Available from: https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2019_4dd50c09-en

23. Canadian Institute for Health Information. National Health Expenditure Trends , 1975 to 2016 [Internet]. Vol. 36, Canadian Institute for Health Information Report. 2016 [cited 2020 Jan 20]. 423–5 p. Available from: https://secure.cihi.ca/free_products/NHEX-Trends-Narrative-Report_2016_EN.pdf

24. Sawyer B, Claxton G. How do health expenditures vary across the population? - Peterson-Kaiser Health System Tracker [Internet]. Peterson-Kaiser Health System Tracker. 2019 [cited 2020 Jan 21]. Available from: https://www.healthsystemtracker.org/chart-collection/health-expenditures-vary-across-population/#item-people-age-55-and-over-account-for-over-half-of-total-health-spending_2016

25. Sinha SK. Why the Elderly Could bankrupt Canada and how demographic imperatives will force the redesign of acute care service delivery. Vol. 11, Healthcare Papers. 2011. p. 46–51.

26. Henningsen N, McAlister M. Paradigm shift: Shaping policy to meet the needs of our aging population [Internet]. Vol. 11, Healthcare Papers. 2011 [cited 2020 Mar 8]. p. 20–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21464623

27. Chappell NL, Hollander MJ. An evidence-based policy prescription for an aging population. Healthc Pap. 2011 Mar;11(1):8–18.

28. Ward MM. Estimating disease prevalence and incidence using administrative data: Some assembly required. Vol. 40, Journal of Rheumatology. NIH Public Access; 2013. p. 1241–3.

29. Bagley SC, Altman RB. Computing disease incidence, prevalence and comorbidity from electronic medical records. Vol. 63, Journal of Biomedical Informatics. Academic Press Inc.; 2016. p. 108–11.

30. Sullivan DF. A single index of mortality and morbidity. HSMHA Health Rep. 1971 Apr;86(4):347–54.

31. Colvez A, Blanchet M. Disability trends in the United States population 1966-76: Analysis of reported causes. Vol. 71, American Journal of Public Health. 1981.

32. Bebbington AC. The expectation of life without disability in England and Wales. Soc Sci Med. 1988;27(4):321–6.

33. Crimmins EM, Reynolds SL, Saito Y. Trends in Health and Ability to Work Among the Older Working-Age Population [Internet]. Vol. 54, Journal of Gerontology: SOCIAL SCIENCES. 1999 [cited 2020 Jan 20]. Available from: https://academic.oup.com/psychsocgerontology/article-abstract/54B/1/S31/592851

34. Verbrugge LM. Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. Milbank Meml Fund Quarterly, Heal Soc. 1984 Sep 9;62(3):475–519.

35. Manton KG. National Long Term Care Survey [Internet]. Encyclopedia of Health & Aging. 2012 [cited 2020 Mar 8]. Available from: https://www.nia.nih.gov/research/resource/national-long-term-care-survey-nltcs

36. Bureau UC. National Health Interview Survey. Med Sci Sports Exerc. 1997;29(6 SUPPL.).

37. Stallard E. Compression of Morbidity and Mortality: New Perspectives. North Am Actuar J. 2016 Oct 1;20(4):341–54.

38. Solé-Auró A, Alcañiz M. Are we living longer but less healthy? Trends in mortality and morbidity in Catalonia (Spain), 1994-2011. Eur J Ageing [Internet]. 2015 Mar [cited 2019 Apr 3];12(1):61–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28804346

39. Riley JC. The risk of being sick: morbidity trends in four countries. Popul Dev Rev. 1990;16(3):403–32.

40. Hossin MZ, Östergren O, Fors S, Kelley-Moore J. Is the Association between Late Life Morbidity and Disability Attenuated over Time? Exploring the Dynamic Equilibrium of Morbidity Hypothesis. Kelley-Moore J, editor. Journals Gerontol - Ser B Psychol Sci Soc Sci [Internet]. 2019 Oct 4 [cited 2020 Jan 22];74(8):E97–106. Available from: https://academic.oup.com/psychsocgerontology/article/74/8/e97/3859724

41. Parker MG, Ahacic K, Thorslund M. Health changes among Swedish oldest old: Prevalence rates from 1992 and 2002 show increasing health problems. Journals Gerontol - Ser A Biol Sci Med Sci. 2005;60(10):1351–5.

42. Steensma C, Loukine L, Choi BC. Evaluating compression or expansion of morbidity in Canada: trends in life expectancy and health-adjusted life expectancy from 1994 to 2010. Heal Promot chronic Dis Prev Canada Res policy Pract [Internet]. 2017 Mar [cited 2020 Feb 6];37(3):68–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28273034

43. Seeman TE, Merkin SS, Crimmins EM, Karlamangla AS. Disability trends among older Americans: National Health and Nutrition Examination surveys, 1988-1994 and 1999-2004. Vol. 100, American Journal of Public Health. 2010. p. 100–7.

44. McCarthy M. Smoking remains leading cause of premature death in US [Internet]. Vol. 348, BMJ (Clinical research ed.). British Medical Journal Publishing Group; 2014 [cited 2020 Jan 16]. p. g396. Available from: https://www.bmj.com/content/348/bmj.g396

45. Manuel DG, Perez R, Bennett C, Rosella L T, M, Roberts M, Sanderson R, Meltem T TP, H M. Seven more years: The impact of smoking, alcohol, diet, physical activity and stress on health and life expectancy in Ontario. An ICES/PHO Report. [Internet]. Toronto: Institute for Clinical Evaluative Sciences and Public Health Ontario. 2012. Available from: http://www.ices.on.ca/~/media/Files/Atlases-Reports/2012/Seven-more-years/Full report.ashx

46. 2014 Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress [Internet]. 2018 [cited 2020 Jan 15]. Available from: https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm

47. Manuel DG, Perez R, Sanmartin C, Taljaard M, Hennessy D, Wilson K, et al. Measuring Burden of Unhealthy Behaviours Using a Multivariable Predictive Approach: Life Expectancy Lost in Canada Attributable to Smoking, Alcohol, Physical Inactivity, and Diet. PLoS Med [Internet]. 2016;13(8):e1002082. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27529741

48. Banks E, Joshy G, Weber MF, Liu B, Grenfell R, Egger S, et al. Tobacco smoking and all-cause mortality in a large Australian cohort study: findings from a mature epidemic with current low smoking prevalence. BMC Med [Internet]. 2015 Dec 24 [cited 2020 Jan 16];13(1):38. Available from: http://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-015-0281-z

49. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med [Internet]. 2013 Jan 24 [cited 2020 Jan 15];368(4):341–50. Available from: http://www.nejm.org/doi/10.1056/NEJMsa1211128

50. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Heal. 2018 Oct 1;6(10):e1077–86.

51. Ng SW, Popkin BM. Time use and physical activity: A shift away from movement across the globe. Obes Rev. 2012 Aug;13(8):659–80.

52. Mok A, Khaw KT, Luben R, Wareham N, Brage S. Physical activity trajectories and mortality: Population based cohort study. BMJ. 2019;365.

53. Reimers CD, Knapp G, Reimers AK. Does physical activity increase life expectancy? A review of the literature. Journal of Aging Research 2012.

54. Diaz KM, Howard VJ, Hutto B, Colabianchi N, Vena JE, Safford MM, et al. Patterns of sedentary behavior and mortality in U.S. middle-aged and older adults a national cohort study. Ann Intern Med. 2017 Oct 3;167(7):465–75.

55. Katzmarzyk PT, Lee IM. Sedentary behaviour and life expectancy in the USA: A cause-deleted life table analysis. BMJ Open. 2012 Jan 1;2(4):e000828.

56. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum NN, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-A systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol. 2017 Jun 1;46(3):1029–56.

57. Johnson RJ, Segal MS, Sautin Y, Nakagawa T, Feig DI, Kang DH, et al. Potential role of sugar (fructose) in the epidemic of hypertension, obesity and the metabolic syndrome, diabetes, kidney disease, and cardiovascular disease1-3. Vol. 86, American Journal of Clinical Nutrition. 2007. p.

899–906.

58. Feskens EJM, Sluik D, Van Woudenbergh GJ. Meat consumption, diabetes, and its complications. Curr Diab Rep. 2013;13(2):298–306.

59. Li Y, Pan A, Wang DD, Liu X, Dhana K, Franco OH, et al. Impact of Healthy Lifestyle Factors on Life Expectancies in the US Population. Circulation [Internet]. 2018 [cited 2020 Jan 15];138(4):345–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29712712

60. Li K, Hüsing A, Kaaks R. Lifestyle risk factors and residual life expectancy at age 40: A German cohort study. BMC Med. 2014 Apr 7;12(1).

61. O'Doherty MG, Cairns K, O'Neill V, Lamrock F, Jørgensen T, Brenner H, et al. Effect of major lifestyle risk factors, independent and jointly, on life expectancy with and without cardiovascular disease: results from the Consortium on Health and Ageing Network of Cohorts in Europe and the United States (CHANCES). Eur J Epidemiol. 2016 May 1;31(5):455–68.

62. Wang H, Naghavi M, Allen C, Barber RM, Carter A, Casey DC, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct 8;388(10053):1459–544.

63. Yusuf PS, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. Lancet. 2004 Sep 11;364(9438):937–52.

64. O'Donnell MJ, Chin SL, Rangarajan S, Xavier D, Liu L, Zhang H, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet. 2016 Aug 20;388(10046):761–75.

65. Ford ES, Capewell S. Proportion of the Decline in Cardiovascular Mortality Disease due to Prevention Versus Treatment: Public Health Versus Clinical Care. Annu Rev Public Health [Internet]. 2011 [cited 2020 Feb 2];32(1):5–22. Available from: www.annualreviews.org

66. Schulze MB, Hoffmann K. Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. Br J Nutr. 2006 May;95(5):860–9.

67. Burke GM, Genuardi M, Shappell H, D'Agostino RB, Magnani JW. Temporal Associations Between Smoking and Cardiovascular Disease, 1971 to 2006 (from the Framingham Heart Study). Am J Cardiol. 2017 Nov 15;120(10):1787–91.

68. Bakhru A, Erlinger TP. Smoking cessation and cardiovascular disease risk factors: Results from the third national health and nutrition examination survey. PLoS Med [Internet]. 2005 [cited 2020 Feb 2];2(6):0528–36. Available from: www.plosmedicine.org

69. Zimmet PZ. Diabetes and its drivers: The largest epidemic in human history? [Internet]. Clinical Diabetes and Endocrinology BioMed Central Ltd.; Dec 18, 2017 p. 1. Available from: http://clindiabetesendo.biomedcentral.com/articles/10.1186/s40842-016-0039-3

70. Guthrie JF, Lin BH, Frazao E. Role of food prepared away from home in the American diet, 1977-78 versus 1994-96: Changes and consequences. J Nutr Educ Behav. 2002;34(3):140–50.

71. Patterson R, McNamara E, Tainio M, de Sá TH, Smith AD, Sharp SJ, et al. Sedentary behaviour and

risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Vol. 33, European Journal of Epidemiology. Springer Netherlands; 2018. p. 811–29.

72. Feldman AL, Long GH, Johansson I, Weinehall L, Fhärm E, Wennberg P, et al. Change in lifestyle behaviors and diabetes risk: Evidence from a population-based cohort study with 10 year follow-up. Int J Behav Nutr Phys Act. 2017 Mar 29;14(1).

73. Ekelund U, Palla L, Brage S, Franks PW, Peters T, Balkau B, et al. Physical activity reduces the risk of incident type 2 diabetes in general and in abdominally lean and obese men and women: The EPIC-InterAct study. Diabetologia [Internet]. 2012 Jul 21 [cited 2020 Jan 30];55(7):1944–52. Available from: http://link.springer.com/10.1007/s00125-012-2532-2

74. Pereira MA, Kartashov AI, Ebbeling CB, Van Horn L, Slattery ML, Jacobs PDR, et al. Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. Lancet [Internet]. 2005 Jan 1 [cited 2020 Mar 7];365(9453):36–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15639678

75. Bray GA, Popkin BM. Dietary sugar and body weight: Have we reached a crisis in the epidemic of obesity and diabetes?: Health be damned! Pour on the sugar. Vol. 37, Diabetes Care. American Diabetes Association Inc.; 2014. p. 950–6.

76. Barnes AS. The epidemic of obesity and diabetes: Trends and treatments. Texas Hear Inst J. 2011;38(2):142–4.

77. Elwood PC, Whitmarsh A, Gallacher J, Bayer A, Adams R, Heslop L, et al. Healthy living and cancer: Evidence from UK Biobank. Ecancermedicalscience. 2018 Jan 4;12.

78. Tang R, Zheng W, Li H, Hu X, Xiang Y. [Lifestyle-related factors and cancer incidence in men: a public health assessment based on a 'healthy lifestyle score']. Zhonghua Liu Xing Bing Xue Za Zhi [Internet]. 2013 Feb [cited 2020 Feb 1];34(2):109–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23751461

79. Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: A brief review of recent epidemiological evidence. In: Lung Cancer. 2004.

80. Fong JH. Disability incidence and functional decline among older adults with major chronic diseases. BMC Geriatr [Internet]. 2019 Nov 21 [cited 2020 Feb 8];19(1):323. Available from: https://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-019-1348-z

81. Valderrama-Gama E, Damian J, Ruigomez A, Martin-Moreno JM. Chronic Disease, Functional Status, and Self-Ascribed Causes of Disabilities Among Noninstitutionalized Older People in Spain. Journals Gerontol Ser A Biol Sci Med Sci [Internet]. 2002 Nov 1 [cited 2020 Feb 9];57(11):M716–21. Available from: https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/57.11.M716

82. Gaugler JE, Duval S, Anderson KA, Kane RL. Predicting nursing home admission in the U.S: a meta-analysis. BMC Geriatr [Internet]. 2007 Dec 19 [cited 2020 Feb 10];7(1):13. Available from: https://bmcgeriatr.biomedcentral.com/articles/10.1186/1471-2318-7-13

83. Smith AK, Walter LC, Miao Y, Boscardin WJ, Covinsky KE. Disability during the last two years of life. JAMA Intern Med. 2013 Sep 9;173(16):1506–13.

84. Dunlop DD, Manheim LM, Sohn MW, Liu X, Chang RW. Incidence of functional limitation in older adults: The impact of gender, race, and chronic conditions. Arch Phys Med Rehabil. 2002 Jul 1;83(7):964–71.

85. Jindai K, Nielson CM, Vorderstrasse BA, Quiñones AR. Multimorbidity and Functional Limitations Among Adults 65 or Older, NHANES 2005–2012. Prev Chronic Dis [Internet]. 2016 Nov 3 [cited 2020 Feb 9];13(11):160174. Available from: http://www.cdc.gov/pcd/issues/2016/16_0174.htm

86. Kingston A, Comas-Herrera A, Jagger C. Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modelling study. Lancet Public Heal. 2018 Sep 1;3(9):e447–55.

87. Kraus L. A Publication of the Rehabilitation Research and Training Center on Disability Statistics and Demographics [Internet]. Institute on Disability/UCED University of New Hampshire. 2017 [cited 2020 Feb 9]. Available from: www.iod.unh.edu

88. Statistics Canada. Canadian Survey on Disability [Internet]. 2012 [cited 2020 Feb 9]. 1–69 p. Available from: http://www.statcan.gc.ca/pub/89-654-x/89-654-x2014001-eng.pdf

89. Freedman VA, Martin LG, Schoeni RF. Recent trends in disability and functioning among older adults in the United States: A systematic review. Vol. 288, Journal of the American Medical Association. 2002. p. 3137–46.

90. Liao WC, Li CR, Lin YC, Wang CC, Chen YJ, Yen CH, et al. Healthy behaviors and onset of functional disability in older adults: Results of a national longitudinal study. J Am Geriatr Soc. 2011 Feb;59(2):200–6.

91. Artaud F, Dugravot A, Sabia S, Singh-Manoux A, Tzourio C, Elbaz A. Unhealthy behaviours and disability in older adults: Three-City Dijon cohort study. BMJ. 2013 Aug 24;347(7922).

92. Kahende JW, Adhikari B, Maurice E, Rock V, Malarcher A. Disparities in health care utilization by smoking status--NHANES 1999-2004. Int J Environ Res Public Health [Internet]. 2009 Mar [cited 2014 Mar 24];6(3):1095–106. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2672402&tool=pmcentrez&rendertype=abstract

93. Azagba S, Sharaf MF, Xiao Liu C. Disparities in health care utilization by smoking status in Canada. Int J Public Health. 2013;58(6):913–25.

94. Izumi Y, Tsuji I, Ohkubo T, Kuwahara A, Nishino Y, Hisamichi S. Impact of smoking habit on medical care use and its costs: a prospective observation of National Health Insurance beneficiaries in Japan. Int J Epidemiol [Internet]. 2001 Jun 1 [cited 2020 Feb 12];30(3):616–21. Available from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/30.3.616

95. Wang Y, Sung H-Y, Lightwood J, Chaffee BW, Yao T, Max W. Health Care Utilization and Expenditures Attributable to Smokeless Tobacco Use Among US Adults. Nicotine Tob Res [Internet]. 2018 Sep 25 [cited 2020 Feb 13];20(11):1359–68. Available from: https://academic.oup.com/ntr/article/20/11/1359/4103039

96. Manuel D. The Ten-Year Impact of Improving Smoking, Alcohol, Diet and Physical Activity in Ontario. 2016.

97. Vander Weg MW, Rosenthal GE, Sarrazin MV. Smoking bans linked to lower hospitalizations for

heart attacks and lung disease among medicare beneficiaries. Health Aff. 2012 Dec 2;31(12):2699–707.

98. Mullen KA, Manuel DG, Hawken SJ, Pipe AL, Coyle D, Hobler LA, et al. Effectiveness of a hospital-initiated smoking cessation programme: 2-year health and healthcare outcomes. Tob Control. 2017 May 1;26(3):293–9.

99. Manuel DG, Perez R, Bennett C, Rosella L CB. 900,000 Days in Hospital: The Annual Impact of Smoking, Alcohol, Diet and Physical Activity on Hospital Use in Ontario. 2014.

100. Wilkins K, Shields M, Rotermann M. Smokers' use of acute care hospitals--a prospective study. Health Rep [Internet]. 2009 Dec [cited 2015 Sep 3];20(4):75–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20108608

101. Kang S wan, Xiang X. Physical activity and health services utilization and costs among U.S. adults. Prev Med (Baltim). 2017 Mar 1;96:101–5.

102. Sari N. Exercise, physical activity and healthcare utilization: A review of literature for older adults. Vol. 70, Maturitas. Elsevier; 2011. p. 285–9.

103. Langsetmo L, Kats AM, Cawthon PM, Cauley JA, Vo TN, Taylor BC, et al. The Association Between Objectively Measured Physical Activity and Subsequent Health Care Utilization in Older Men. Journals Gerontol Ser A [Internet]. 2019 May 16 [cited 2020 Feb 13];74(6):820–6. Available from: https://academic.oup.com/biomedgerontology/article/74/6/820/4399482

104. Li CL, Chu SJ, Sheu JT, Huang LYG. Impact of physical activity on hospitalization in older adults: A nationwide cohort from Taiwan. Arch Gerontol Geriatr. 2011 Sep 1;53(2):141–5.

105. Haapanen-Niemi N, Miilunpalo S, Vuori I, Pasanen M, Oja P. The impact of smoking, alcohol consumption, and physical activity on use of hospital services. Am J Public Health. 1999 Oct 7;89(5):691–8.

106. Martin MY, Powell MP, Peel C, Zhu S, Allman R. Leisure-time physical activity and health-care utilization in older adults. J Aging Phys Act. 2006 Oct 1;14(4):392–410.

107. Ackermann RT, Cheadle A, Sandhu N, Madsen L, Wagner EH, LoGerfo JP. Community exercise program use and changes in healthcare costs for older adults. Am J Prev Med [Internet]. 2003 Oct [cited 2020 Mar 5];25(3):232–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14507530

108. Munro JF, Nicholl JP, Brazier JE, Davey R, Cochrane T. Cost effectiveness of a community based exercise programme in over 65 year olds: Cluster randomised trial. J Epidemiol Community Health. 2004 Dec 1;58(12):1004–10.

109. Ford DW, Hartman TJ, Still C, Wood C, Mitchell DC, Bailey R, et al. Diet quality and body mass index are associated with health care resource use in rural older adults. J Acad Nutr Diet. 2014 Dec 1;114(12):1932–8.

110. Klijs B, Nusselder WJ, Looman CW, Mackenbach JP. Contribution of Chronic Disease to the Burden of Disability. Verbeek JH, editor. PLoS One [Internet]. 2011 Sep 22 [cited 2020 Feb 9];6(9):e25325. Available from: https://dx.plos.org/10.1371/journal.pone.0025325

111. Manuel DG, Perez R, Bennett C, Laporte A WA, Gandhi S, Yates EA HD. A $4.9 Billion Decrease Of,

in Health Care Expenditure: The Ten-Year Impact Changing Smoking, Alcohol, Diet and Physical Activity on Health Care [Internet]. 2016 [cited 2020 Feb 13]. Available from: https://www.ices.on.ca/Publications/Atlases-and-Reports/2016/A-4-9-Billion-Dollar-Decrease-in-Health-Care-Expenditure

112. Dalziel K, Segal L. Time to give nutrition interventions a higher profile: cost-effectiveness of 10 nutrition interventions. Health Promot Int [Internet]. 2007 Dec [cited 2020 Mar 7];22(4):271–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17916604

113. Barendregt JJ, Bonneux L, van der Maas PJ. The health care costs of smoking. N Engl J Med [Internet]. 1997 Oct 9 [cited 2015 Sep 3];337(15):1052–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9321534

114. Van Baal PHM, Polder JJ, De Wit GA, Hoogenveen RT, Feenstra TL, Boshuizen HC, et al. Lifetime medical costs of obesity: Prevention no cure for increasing health expenditure. PLoS Med. 2008;5(2):0242–9.

115. Cao B. Future healthy life expectancy among older adults in the US: a forecast based on cohort smoking and obesity history. Popul Health Metr [Internet]. 2016 [cited 2018 May 15];14:23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27408607

116. Nusselder WJ, Looman CW, Marang-van de Mheen PJ, van de Mheen H, Mackenbach JP. Smoking and the compression of morbidity. J Epidemiol Community Health [Internet]. 2000 Aug 1 [cited 2019 May 30];54(8):566–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10890867

117. Fine LJ, Philogene GS, Gramling R, Coups EJ, Sinha S. Prevalence of multiple chronic disease risk factors: 2001 National Health Interview Survey. Am J Prev Med. 2004;27(SUPPL.):18–24.

118. Fries JF, Bruce B, Chakravarty E. Compression of morbidity 1980-2011: A focused review of paradigms and progress. Vol. 2011, Journal of Aging Research. 2011.

119. Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. Annu Rev Sociol. 2010 Jun;36(1):349–70.

120. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health [Internet]. 2016 Dec 29 [cited 2020 Feb 12];16(1):657. Available from: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-016-3373-6

121. Lippiatt BC. Measuring medical cost and life expectancy impacts of changes in cigarette sales. Prev Med (Baltim) [Internet]. 1990 Sep [cited 2020 Mar 5];19(5):515–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2122438

122. Rasmussen SR, Prescott E, Sørensen TI a, Søgaard J. The total lifetime costs of smoking. Eur J Public Health [Internet]. 2004 Mar;14(1):95–100. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15080400

123. Lim K-L, Jacobs P, Ohinmaa A, Schopflocher D, Dewa CS. A new population-based measure of the economic burden of mental illness in Canada. Chronic Dis Can. 2008;28(3):92–8.

124. Statistics Canada. Evaluation of the Health Statistics Program (2011/2012 to 2015/2016) [Internet]. Evaluation Reports. 2017. Available from: https://www.statcan.gc.ca/eng/about/er/hspfr#fn27

125. Sanmartin C, Decady Y, Trudeau R, Dasylva A, Tjepkema M, Finès P, et al. Linking the Canadian Community Health Survey and the Canadian Mortality Database: An enhanced data source for the study of mortality. [cited 2018 Mar 8]; Available from: http://www.statcan.gc.ca/pub/82-003-x/2016012/article/14687-eng.pdf

126. Weissman J, Parker JD, Miller DM, Miller EA, Gindi RM. The relationship between linkage refusal and selected health conditions of survey respondents. Surv Pract [Internet]. 2016 Sep 1 [cited 2019 Feb 19];9(5):1–10. Available from: https://surveypractice.scholasticahq.com/article/2798-the-relationship-between-linkage-refusal-and-selected-health-conditions-of-survey-respondents

127. Wodchis WP, Bushmeneva K, Nikitovic M, McKillop I. Guidelines on person level cost using administrative databases in Ontario [Internet]. Vol. 1, Health System Performance Research Network Working Paper Series. 2013 Jan [cited 2018 Mar 13]. Available from: https://www.scienceopen.com/document?vid=19dd8115-435f-4586-81f3-eaec85280736

128. Kralj B. Measuring Rurality - Rurality Index of Ontario 2008 BASIC : Methodology and Results. 2008;1–7.

129. Canada S. Postal Code OM Conversion File (PCCF), Reference Guide. [cited 2018 Mar 18]; Available from: http://www.statcan.gc.ca/pub/92-154-g/92-154-g2017001-eng.pdf

130. Statistics Canada. Canadian Community Health Surveys [Internet]. Available from: https://www.statcan.gc.ca/eng/survey/household/3226

131. Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, 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 [Internet]. 2016 Aug 23 [cited 2018 Mar 13];54(1):337–49. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27567819

132. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying patients with physician-diagnosed asthma in health administrative databases. Can Respir J [Internet]. [cited 2018 Mar 13];16(6):183–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20011725

133. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. Chronic Dis Inj Can [Internet]. 2013 Jun [cited 2018 Mar 13];33(3):160–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23735455

134. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care [Internet]. 2002 Mar [cited 2018 Mar 13];25(3):512–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11874939

135. Tu K, Campbell NR, Chen Z-L, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. Open Med [Internet]. 2007 Apr 14 [cited 2018 Mar 13];1(1):e18-26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20101286

136. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physcian diagnosed COPD in health administrative databases. COPD [Internet]. 2009 Oct [cited 2018 Mar 13];6(5):388–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19863368

137. Report M. CIHI ' s Population Grouping Methodology Report. Vol. 1. 2017.

138. Reitsma MB, Fullman N, Ng M, Salama JS, Abajobir A, Abate KH, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet. 2017 May 13;389(10082):1885–906.

139. Alberg AJ, Shopland DR, Cummings KM. The 2014 Surgeon General's Report: Commemorating the 50th Anniversary of the 1964 Report of the Advisory Committee to the US Surgeon General and Updating the Evidence on the Health Consequences of Cigarette Smoking. Am J Epidemiol. 2014 Feb 15;179(4):403–12.

140. Reitsma MB, Kendrick PJ, Ababneh E, Abbafati C, Abbasi-Kangevari M, Abdoli A, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet [Internet]. 2021 Jun 19 [cited 2022 Jan 27];397(10292):2337–60. Available from: http://www.thelancet.com/article/S0140673621011697/fulltext

141. Chudasama Y V., Khunti KK, Zaccardi F, Rowlands A V., Yates T, Gillies CL, et al. Physical activity, multimorbidity, and life expectancy: A UK Biobank longitudinal study. BMC Med [Internet]. 2019 Jun 12 [cited 2022 Jan 27];17(1):1–13. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1339-0

142. Ho JY, Hendi AS. Recent trends in life expectancy across high income countries: retrospective observational study. BMJ [Internet]. 2018 [cited 2020 Jan 18];362:2562. Available from: http://dx.doi.org/10.1136/bmj.k2562

143. Watts P, Buck D, Netuveli G, Renton A. Clustering of lifestyle risk behaviours among residents of forty deprived neighbourhoods in London: Lessons for targeting public health interventions. J Public Heal (United Kingdom). 2016;38(2):308–15.

144. Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: A red herring? Vol. 8, Health Economics. 1999. p. 485–96.

145. NC H, C R, MC G. Determinants of health care costs in the senior elderly: age, comorbidity, impairment, or proximity to death? Eur J Health Econ [Internet]. 2018 Jul 1 [cited 2021 Sep 30];19(6):831–42. Available from: https://pubmed.ncbi.nlm.nih.gov/28856487/

146. von Wyl V. Proximity to death and health care expenditure increase revisited: A 15-year panel analysis of elderly persons. Heal Econ Rev 2019 91 [Internet]. 2019 Mar 11 [cited 2021 Sep 30];9(1):1–16. Available from: https://healtheconomicsreview.biomedcentral.com/articles/10.1186/s13561-019-0224-z

147. Nshimyumukiza L, Lieffers JRL, Ekwaru JP, Ohinmaa A, Veugelers PJ. Temporal changes in diet quality and the associated economic burden in Canada. PLoS One [Internet]. 2018 Nov 1 [cited 2022 Jan 31];13(11):e0206877. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206877

148. Wijeysundera HC, Wang X, Tomlinson G, Ko DT, Krahn MD. Techniques for estimating health care costs with censored data: an overview for the health services researcher. Clinicoecon Outcomes Res [Internet]. 2012;4:145–55. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3377439&tool=pmcentrez&rendertype=abstract

149. Tanuseputro P, Wodchis WP, Fowler R, Walker P, Bai YQ, Bronskill SE, et al. The health care cost of dying: A population-based retrospective cohort study of the last year of life in Ontario, Canada. PLoS One [Internet]. 2015;10(3):1–11. Available from: http://dx.doi.org/10.1371/journal.pone.0121759

150. Oliveira de, W KK. Phase-specific and lifetime costs of cancer care in Ontario, Canada. 2016 [cited 2022 May 17]; Available from: https://doi.org/10.1186/s12885-016-2835-7

151. Mendlowitz A, Bremner KE, Walker JD, Wong WWL, Feld JJ, Sander B, et al. Health care costs associated with hepatitis C virus infection in First Nations populations in Ontario: a retrospective matched cohort study. Can Med Assoc Open Access J [Internet]. 2021 Jul 1 [cited 2022 May 17];9(3):E897–906. Available from: https://www.cmajopen.ca/content/9/3/E897

152. Tawfik A, Wodchis WP, Pechlivanoglou P, Hoch J, Husereau D, Krahn M. Using Phase-Based Costing of Real-World Data to Inform Decision-Analytic Models for Atrial Fibrillation. Appl Health Econ Health Policy [Internet]. 2016 Jun 1 [cited 2022 May 17];14(3):313–22. Available from: https://pubmed.ncbi.nlm.nih.gov/26924098/

153. Basu A, Manning WG. Estimating lifetime or episode-of-illness costs under censoring. Health Econ. 2010;19(9):1010–28.

154. Sweegers MG, Boyle T, Vallance JK, Chinapaw MJ, Brug J, Aaronson NK, et al. Which cancer survivors are at risk for a physically inactive and sedentary lifestyle? Results from pooled accelerometer data of 1447 cancer survivors. Int J Behav Nutr Phys Act [Internet]. 2019 Aug 16 [cited 2022 Jun 1];16(1):1–15. Available from: https://ijbnpa.biomedcentral.com/articles/10.1186/s12966-019-0820-7

155. Oliveira C de, Bremner KE, Pataky R, Gunraj N, Haq M, Chan K, et al. Trends in use and cost of initial cancer treatment in Ontario: a population-based descriptive study. Can Med Assoc Open Access J [Internet]. 2013 Oct 22 [cited 2022 Jun 1];1(4):E151–8. Available from: https://www.cmajopen.ca/content/1/4/E151

156. Steensma C, Choi BCK, Loukine L, Schanzer D. Period Life Tables for Calculating Life Expectancy: Options to Assess and Minimize the Potential for Bias. Am J Public Health [Internet]. 2018 Mar 1 [cited 2022 May 25];108(3):e14. Available from: /pmc/articles/PMC5803818/

157. Chen X, Lin F. Estimating Transitional Probabilities with Cross-Sectional Data to Assess Smoking Behavior Progression: A Validation Analysis. J Biom Biostat [Internet]. 2013 [cited 2022 May 25];Suppl 1. Available from: /pmc/articles/PMC4178951/

158. Khosravi A, Mansournia MA, Mahmoodi M, Pouyan AA, Holakouie-Naieni K. Estimating the Transitional Probabilities of Smoking Stages with Cross-sectional Data and 10-Year Projection for Smoking Behavior in Iranian Adolescents. Int J Prev Med [Internet]. 2016 [cited 2022 May 25];7(1):101. Available from: /pmc/articles/PMC5007904/

159. Holliday KM, Lin D, Chakladar S, Daviglus ML, Evenson KR, Heiss G, et al. Abstract 59: Age-Specific Probabilities of Transitioning between Ideal, Intermediate, and Poor Levels of Physical Activity in the United States: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) and the National Health and Nutrition Examination Survey (NHANES). Circulation [Internet]. 2014 Mar 25 [cited 2022 May 25];129(suppl_1). Available from: https://www.ahajournals.org/doi/abs/10.1161/circ.129.suppl_1.59

160. Wakimoto P, Block G. Dietary intake, dietary patterns, and changes with age: an epidemiological perspective. J Gerontol A Biol Sci Med Sci [Internet]. 2001 [cited 2022 Jun 2];56 Spec No 2(SPEC. ISS. 2):65–80. Available from: https://pubmed.ncbi.nlm.nih.gov/11730239/

161. Johannsen DL, DeLany JP, Frisard MI, Welsch MA, Rowley CK, Fang X, et al. Physical activity in aging: Comparison among young, aged, and nonagenarian individuals. J Appl Physiol [Internet]. 2008 Aug [cited 2022 Jun 2];105(2):495. Available from: /pmc/articles/PMC2519943/

162. Thomeer MB, Hernandez E, Umberson D, Thomas PA. Influence of social connections on smoking behavior across the life course. Adv Life Course Res. 2019 Dec 1;42:100294.

163. Clarke P, O'Malley PM, Johnston LD, Schulenberg JE. Social disparities in BMI trajectories across adulthood by gender, race/ethnicity and lifetime socio-economic position: 1986-2004. Int J Epidemiol [Internet]. 2009 [cited 2022 Jun 2];38(2):499–509. Available from: https://pubmed.ncbi.nlm.nih.gov/18835869/

164. Sheehan TJ, DuBrava S, DeChello LM, Fang Z. Rates of weight change for black and white Americans over a twenty year period. Int J Obes Relat Metab Disord [Internet]. 2003 Apr 1 [cited 2022 Jun 2];27(4):498–504. Available from: https://pubmed.ncbi.nlm.nih.gov/12664083/

165. Tandon A, Sadana R, Murray C, Serdobova I. Describing Population Health in Six Domains : Comparable Results From 66 Household Surveys. WHO Glob Program Evid Heal Policy Discuss Pap. 2002;(43).

166. Corsi DJ, Neuman M, Finlay JE, Subramanian S V. Demographic and health surveys: a profile. Int J Epidemiol [Internet]. 2012 [cited 2021 May 23];41:1602–13. Available from: https://academic.oup.com/ije/article/41/6/1602/747345

167. Üstün TB, Chatterji S, Mechbal A, Murray CJL, WHS Collaborating Groups. The World Health Surveys. In: Health systems performance assessment: debates, methods and empiricism. 2003. p. 797–809.

168. Wallar LE, Rosella LC. Risk factors for avoidable hospitalizations in Canada using national linked data: A retrospective cohort study. PLoS One [Internet]. 2020 [cited 2021 Sep 17];15(3). Available from: https://doi.org/10.1371/journal.pone.0229465

169. Garner R, Tanuseputro P, Manuel DG, Sanmartin C. Transitions to long-term and residential care among older Canadians. Heal reports [Internet]. 2018 [cited 2021 May 30];29(5):13–23. Available from: https://www150.statcan.gc.ca/n1/pub/82-003-x/2018005/article/54966-eng.htm

170. Manuel DG, Tuna M, Perez R, Tanuseputro P, Hennessy D, Bennett C, et al. Predicting Stroke Risk Based on Health Behaviours: Development of the Stroke Population Risk Tool (SPoRT). Obukhov AG, editor. PLoS One [Internet]. 2015 Dec 4 [cited 2018 May 18];10(12):e0143342. Available from: http://dx.plos.org/10.1371/journal.pone.0143342

171. Lebenbaum M, Espin-Garcia O, Li Y, Rosella LC. Development and validation of a population based risk algorithm for obesity: The Obesity Population Risk Tool (OPoRT). Brusic V, editor. PLoS One [Internet]. 2018 Jan 18 [cited 2019 Feb 28];13(1):e0191169. Available from: https://dx.plos.org/10.1371/journal.pone.0191169

172. Manuel DG, Msc CB, Hennessy D, Rosella L, Sanmartin C, Tu J V., et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the cardiovascular Disease Population Risk Tool (cVDPoRT). C | JULY [Internet]. 2018 Jul 23 [cited

2018 Dec 11];23(29):871–82. Available from: http://www.cmaj.ca/content/190/29/E871

173. Ng R, Sutradhar R, Wodchis WP, Rosella LC. Chronic Disease Population Risk Tool (CDPoRT): a study protocol for a prediction model that assesses population-based chronic disease incidence. Diagnostic Progn Res [Internet]. 2018 Dec 1 [cited 2021 May 22];2(1):1–11. Available from: https://doi.org/10.1186/s41512-018-0042-5

174. LC R, K K, Z Y, DG M, C B, R F, et al. Predicting High Health Care Resource Utilization in a Single-payer Public Health Care System: Development and Validation of the High Resource User Population Risk Tool. Med Care [Internet]. 2018 Oct 1 [cited 2021 Sep 18];56(10):e61–9. Available from: https://pubmed.ncbi.nlm.nih.gov/29189576/

175. Meyer BD, C Mok WK, Sullivan JX. Household Surveys in Crisis. J Econ Perspect [Internet]. 2015 [cited 2019 Feb 26];29(1):1–6. Available from: http://dx.doi.org/10.1257/jep.29.4.1

176. Groves RM. NONRESPONSE RATES AND NONRESPONSE BIAS IN HOUSEHOLD SURVEYS. Public Opin Q [Internet]. 2006 [cited 2019 Feb 12];70(5):646–75. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.5319&rep=rep1&type=pdf

177. Massey DS, Tourangeau R. Where Do We Go from Here? Nonresponse and Social Measurement. Ann Am Acad Pol Soc Sci [Internet]. 2013 Jan [cited 2019 Feb 12];645(1):222–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25484368

178. Bladon TL. The downward trend of survey response rates: Implications and considerations for evaluators [Internet]. Vol. 24, Canadian Journal of Program Evaluation. 2009 [cited 2018 Dec 12]. p. 131–56. Available from: https://evaluationcanada.ca/secure/24-2-131.pdf

179. Muggah E, Graves E, Bennett C, Manuel DG. Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. [cited 2017 Mar 19]; Available from: http://download.springer.com/static/pdf/481/art%253A10.1186%252F1471-2458-13-16.pdf?originUrl=http%3A%2F%2Fbmcpublichealth.biomedcentral.com%2Farticle%2F10.1186%2F1471-2458-13-16&token2=exp=1489963501~acl=%2Fstatic%2Fpdf%2F481%2Fart%25253A10.1186%25252F1471-2458-13-16.pdf*~hmac=994cad660320a5d9810624762fc864282ac4fb4015ad143c8ed69953caeae6f3

180. KA F, LE G, A G, R U, R P, L F, et al. Effect of socio-demographic and health factors on the association between multimorbidity and acute care service use: population-based survey linked to health administrative data. BMC Health Serv Res [Internet]. 2021 Dec 1 [cited 2021 Sep 18];21(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33435978/

181. Griffith LE, Gruneir A, Fisher KA, Upshur R, Patterson C, Perez R, et al. The hidden complexity of measuring number of chronic conditions using administrative and self-report data: A short report: https://doi.org/101177/2235042X20931287 [Internet]. 2020 Jun 26 [cited 2021 Sep 18];10:2235042X2093128. Available from: https://journals.sagepub.com/doi/full/10.1177/2235042X20931287

182. Cheung KL, ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health [Internet]. 2017 Dec 23 [cited 2018 Jan 12];17(1):276. Available from:

http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4189-8

183. Van Loon AJM, Tijhuis M, Picavet HSJ, Surtees PG, Ormel J. Survey non-response in the Netherlands: effects on prevalence estimates and associations. Ann Epidemiol [Internet]. 2003 Feb 1 [cited 2018 Mar 4];13(2):105–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12559669

184. Zhao J, Stockwell T, Macdonald S. Non-response bias in alcohol and drug population surveys. Drug Alcohol Rev [Internet]. 2009 May 18 [cited 2018 Mar 8];28(6):648–57. Available from: http://doi.wiley.com/10.1111/j.1465-3362.2009.00077.x

185. Cohen G, Duffy JC. Are Nonrespondents to Health Surveys Less Healthy Than Respondents? J Off Stat. 2002;18(1):13–23.

186. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher D V, Scott I, et al. Data Linkage: A powerful research tool with potential problems. BMC Heal Serv Res 2010 101 [Internet]. 2010 Dec 22 [cited 2021 Sep 18];10(1):1–7. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-10-346

187. Gaugler JE, Duval S, Anderson KA, Kane RL. Predicting nursing home admission in the U.S: a meta-analysis. BMC Geriatr [Internet]. 2007 Dec 19 [cited 2018 May 18];7(1):13. Available from: http://bmcgeriatr.biomedcentral.com/articles/10.1186/1471-2318-7-13

188. Seitz D, Purandare N, Conn D. Prevalence of psychiatric disorders among older adults in long-term care homes: a systematic review. Int Psychogeriatrics C Int Psychogeriatr Assoc [Internet]. 2018 [cited 2018 May 18];22(7):1025–39. Available from: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/3424575EA5F22DBEE79FBE39132BF511/S1041610210000608a.pdf/prevalence_of_psychiatric_disorders_among_older_adults_in_longterm_care_homes_a_systematic_review.pdf

189. Statistics Canada. 2016 Census Data tables – Structural Type of Dwelling (10), Age (20) and Sex (3) for the Population in Occupied Private Dwellings of Canada, Provinces and Territories, Census Divisions and Census Subdivisions, 2016 Census - 100% Data [Internet]. 2019 [cited 2021 May 22]. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=2&Lang=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=1234492&GK=0&GRP=1&PID=109537&PRID=10&PTYPE=109445&S=0&SHOWALL=0&SUB=0&Temporal=2016&THEME=116&VID=0&VNAMEE=&VNAMEF=&D1=0

190. Poss JW, Sinn CLJ, Grinchenko G, Blums J, Peirce T, Hirdes J. Location, location, location: Characteristics and services of long-stay home care recipients in retirement homes compared to others in private homes and long-term care homes. Healthc Policy [Internet]. 2017 [cited 2021 Jun 8];12(3):80–93. Available from: http://www.rhra.ca/en/register/.

191. Information Sheet Population Grouping Methodology Overview of the methodology.

192. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. J Grad Med Educ [Internet]. 2012 Sep 1 [cited 2021 Sep 20];4(3):279–82. Available from: /pmc/articles/PMC3444174/

193. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput. 2009

Jun;38(6):1228–34.

194. Center for Health Statistics - Division of Health Interview Statistics N. 2018 National Health Interview Survey (NHIS) Public Use Data Release. Natl Heal Interview Surv. 2018;

195. 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 [Internet]. 2018 Mar [cited 2018 Mar 5];29(2):299–307. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29389712

196. The Scottish Government. Scottish Health Survey [Internet]. Survey. 2003 [cited 2021 Sep 21]. Available from: https://www.scotpho.org.uk/publications/overview-of-key-data-sources/surveys-cross-sectional/scottish-health-survey

197. Gray L, McCartney G, White IR, Vittal Katikireddi S, Rutherford L, Gorman E, et al. Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: a study protocol. BMJ Open [Internet]. 2013 [cited 2019 Feb 11];3:2647. Available from: http://bmjopen.bmj.com

198. Kouyoumdjian FG, Cheng SY, Fung K, Humphreys-Mahaffey S, Orkin AM, Kendall C, et al. Primary care utilization in people who experience imprisonment in Ontario, Canada: a retrospective cohort study. BMC Health Serv Res [Internet]. 2018 Nov 9 [cited 2021 Sep 22];18(1). Available from: /pmc/articles/PMC6234797/

199. Otero Varela L, Doktorchik C, Wiebe N, Quan H, Eastwood C. Exploring the differences in ICD and hospital morbidity data collection features across countries: an international survey. BMC Heal Serv Res 2021 211 [Internet]. 2021 Apr 7 [cited 2021 Aug 23];21(1):1–9. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-021-06302-w

200. Sadana R, Mathers CD, Lopez AD, Murray CJL, Iburg K. Comparative Analyses of More Than 50 Household Surveys on Health Status. EIP/GPE/EBD Word Heal Organ GPE Discuss Pap Ser No 15. 2001;(15):3–78.

201. Drapeau A, Boyer R, Diallo FB. Discrepancies between survey and administrative data on the use of mental health services in the general population: findings from a study conducted in Quebec. BMC Public Health [Internet]. 2011;11(1):837. Available from: http://www.biomedcentral.com/1471-2458/11/837

202. A B, T W. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev [Internet]. 2006 Apr [cited 2021 Sep 4];63(2):217–35. Available from: https://pubmed.ncbi.nlm.nih.gov/16595412/

203. PL R, AL S, H K, DS S, DA B, KR L. Self-reports of health care utilization compared to provider records. J Clin Epidemiol [Internet]. 2001 [cited 2021 Sep 4];54(2):136–41. Available from: https://pubmed.ncbi.nlm.nih.gov/11166528/

204. Quan H, Fong A, Coster C De, Wang J, Musto R, Noseworthy TW, et al. Variation in health services utilization among ethnic populations. CMAJ [Internet]. 2006 Mar 14 [cited 2021 Aug 23];174(6):787–91. Available from: https://www.cmaj.ca/content/174/6/787

205. Guo EX, Sweetman A, Guindon GE. Socioeconomic differences in prescription drug supplemental coverage in Canada: A repeated cross-sectional study. Health Policy (New York). 2020 Mar 1;124(3):252–60.

206. OECD. Health spending set to outpace GDP growth to 2030 [Internet]. 2019 [cited 2021 Aug 23]. Available from: https://www.oecd.org/health/health-spending-set-to-outpace-gdp-growth-to-2030.htm

207. Norbeck TB. Drivers of Health Care Costs: A Physicians Foundation white Paper - Second of a Three-Part Series. Mo Med [Internet]. 2013 Mar [cited 2021 Aug 23];110(2):113. Available from: /pmc/articles/PMC6179664/

208. de Boer WIJ, Dekker LH, Koning RH, Navis GJ, Mierau JO. How are lifestyle factors associated with socioeconomic differences in health care costs? Evidence from full population data in the Netherlands. Prev Med (Baltim). 2020 Jan 1;130:105929.

209. Rosella LC, Fitzpatrick T, Wodchis WP, Calzavara A, Manson H, Goel V. High-cost health care users in Ontario, Canada: demographic, socio-economic, and health status characteristics. 2014 [cited 2021 Aug 24]; Available from: http://www.biomedcentral.com/1472-6963/14/532

210. Manuel DG, Bennett C, Perez R, Wilton AS, Dass AR, Laporte A, et al. Burden of health behaviours and socioeconomic position on health care expenditure in Ontario. F1000Research [Internet]. 2019 Mar 18 [cited 2021 Aug 24];8. Available from: /pmc/articles/PMC6844135/

211. Howdon D, Rice N. Health care expenditures, age, proximity to death and morbidity: Implications for an ageing population. J Health Econ [Internet]. 2018 Jan 1 [cited 2019 Feb 18];57:60–74. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0167629617310020

212. Polder JJ, Barendregt JJ, van Oers H. Health care costs in the last year of life-The Dutch experience. Soc Sci Med. 2006 Oct;63(7):1720–31.

213. Dale MC, Burns A, Panter L, Morris J. Factors affecting survival of elderly nursing home residents. Int J Geriatr Psychiatry [Internet]. 2001 [cited 2021 Sep 4];16(1):70–6. Available from: https://pubmed.ncbi.nlm.nih.gov/11180488/

214. Hjaltadóttir I, Hallberg IR, Ekwall AK, Nyberg P. Predicting mortality of residents at admission to nursing home: A longitudinal cohort study. BMC Heal Serv Res 2011 111 [Internet]. 2011 Apr 20 [cited 2021 Sep 4];11(1):1–11. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-11-86

215. P T, M C, C B, A G, SE B, P W, et al. Hospitalization and Mortality Rates in Long-Term Care Facilities: Does For-Profit Status Matter? J Am Med Dir Assoc [Internet]. 2015 Oct 1 [cited 2021 Sep 4];16(10):874–83. Available from: https://pubmed.ncbi.nlm.nih.gov/26433862/

216. Roblin B, Deber R, Kuluski K, Silver MP. Ontario's Retirement Homes and Long-Term Care Homes: A Comparison of Care Services and Funding Regimes. Can J Aging [Internet]. 2019 Jun 1 [cited 2021 May 30];38(2):155–67. Available from: https://doi.org/10.1017/S0714980818000569

217. B R, R D, K K, MP S. Ontario's Retirement Homes and Long-Term Care Homes: A Comparison of Care Services and Funding Regimes. Can J Aging [Internet]. 2019 Jun 1 [cited 2021 Sep 4];38(2):155–67. Available from: https://pubmed.ncbi.nlm.nih.gov/30626461/

218. J P, CL S, G G, L S-W, J H. Comparing Changes and Transitions of Home Care Clients in Retirement Homes and Private Homes. Can J Aging [Internet]. 2020 Sep 1 [cited 2021 Sep 4];39(3):421–31. Available from: https://pubmed.ncbi.nlm.nih.gov/31412959/

219. Feinberg AE, Porter J, Saskin R, Rangrej J, Urbach DR. Regional variation in the use of surgery in

Ontario. Can Med Assoc Open Access J [Internet]. 2015 Jul 17 [cited 2021 Sep 4];3(3):E310–6. Available from: https://www.cmajopen.ca/content/3/3/E310

220. Rotermann M. Evaluation of the coverage of linked Canadian Community Health Survey and hospital inpatient records. Heal reports / Stat Canada, Can Cent Heal Inf = Rapp sur la sant?? / Stat Canada, Cent Can d'information sur la sant?? 2009;20(1):45–51.

221. Blewett LA, Call KT, Turner J, Hest R. Data Resources for Conducting Health Services and Policy Research. Annu Rev Public Health [Internet]. 2018 Apr 1 [cited 2021 Oct 7];39:437. Available from: /pmc/articles/PMC5880724/

222. JH K, MAK H, RA G, SF P. Life in Data Sets: Locating and Accessing Data on the Health of Americans Across the Life Span. J Public Health Manag Pract [Internet]. 2021 May 1 [cited 2021 Oct 7];27(3):E126–42. Available from: https://pubmed.ncbi.nlm.nih.gov/31688741/

223. Lin DY. Regression analysis of incomplete medical cost data. Stat Med. 2003;22(7):1181–200.

224. Luppa M, Luck T, Weyerer S, Konig H-H, Brahler E, Riedel-Heller SG. Prediction of institutionalization in the elderly. A systematic review. Age Ageing [Internet]. 2010 Jan 1 [cited 2020 Feb 10];39(1):31–8. Available from: https://academic.oup.com/ageing/article-lookup/doi/10.1093/ageing/afp202

225. Manis DR, Poss JW, Jones A, Rochon PA, Bronskill SE, Campitelli MA, et al. Rates of health services use among residents of retirement homes in Ontario: a population-based cohort study. CMAJ [Internet]. 2022 May 30 [cited 2022 Jun 6];194(21):E730–8. Available from: https://www.cmaj.ca/content/194/21/E730

226. Moore AA, Gould R, Reuben DB, Greendale GA, Carter MK, Zhou K, et al. Longitudinal Patterns and Predictors of Alcohol Consumption in the United States. https://doi.org/102105/AJPH2003019471 [Internet]. 2011 Oct 10 [cited 2022 Jun 7];95(3):458–64. Available from: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2003.019471

227. Platt A, Sloan FA, Costanzo P. Alcohol-Consumption Trajectories and Associated Characteristics Among Adults Older Than Age 50*. http://dx.doi.org/1015288/jsad201071169 [Internet]. 2015 Jan 21 [cited 2022 Jun 7];71(2):169–79. Available from: https://www.jsad.com/doi/abs/10.15288/jsad.2010.71.169