THE EVOLVING USE OF ADMINISTRATIVE HEALTH DATA FOR QUANTIFYING BURDEN OF ILLNESS

L W SVENSON

PhD 2015

THE EVOLVING USE OF ADMINISTRATIVE HEALTH DATA FOR QUANTIFYING BURDEN OF ILLNESS

LAWRENCE WALTER SVENSON

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for degree of Doctor of Philosophy by Published Work (Route 2)

Department of Health Professions
Faculty of Health, Psychology and Social Care
the Manchester Metropolitan University

2015

Table of Contents

Abstract

The aim of this thesis is to demonstrate how the use of administrative health data (AHD) has evolved over time to be a valuable resource for quantifying disease burden. AHD are defined as data routinely collected for the purposes of payment, monitoring, priority setting, and evaluation of the provision of health services. While the primary purpose of AHD is not research, and researchers typically do not have a direct role in their collection, they represent a rich source of data for secondary analyses.

The thesis presents and critiques 12 peer-reviewed publications to demonstrate how the use of AHD has evolved for better understanding the burden of disease. Each publication shows a natural evolution in thinking, and sophistication of methods that help to illustrate how secondary data can be used to augment primary data collection methods. The development of evidence using different methods, particularly when there are consistent results, works to strengthen our understanding of any given health issue.

The thesis defines AHD, as well as its strengths and limitations, and how these data can be considered 'big data'. Next, historical developments on secondary use and AHD are provided starting with the work of John Graunt, and ending with the present author.

The value of AHD is explored critically through the following themes: role of mortality and hospitalisation data; development of algorithms for improving the accuracy of AHD for determining the presence of disease (case definition algorithms); strength and value of longitudinal designs; identification of rare health events; assessing the burden of co-morbidities; assessing health outcomes; and finally how AHD can support policy development. Future directions for research are highlighted, as well as how AHD can be used to inform policy, resource allocation, and practice.

Acknowledgements

I would like to thank my academic advisor, Francis Fatoye, for his support, feedback, encouragement, and guidance throughout the process. Also, Dan O'Connor who provided guidance on navigating university requirements and the admission processes. I would like to thank my five nieces (Sydney, Tess, Gabby, Alicia, Ashley) and nephew (Michael) for motivating me to be a positive role model; my parents for being positive role models; my sister, and extended family for continuing support and encouragement; and my colleagues for teaching and inspiring me. Finally, I want to thank my wife Jill for her patience, encouragement, and confidence in me.

Chapter 1: Introduction

1.1 Aims of the Thesis

Administrative health data (AHD) are data routinely collected as part of the operation of the healthcare system. The overall aim of the thesis is to present and critique how AHD has evolved to become a valuable resource for health services research in general, and the quantification of disease burden specifically. The thesis will achieve this aim through the presentation of 12 papers, for which the present author was either the sole or co-author. The papers presented as part of this thesis were chosen from 139 papers authored or co-authored by the current author covering the period 1990 to 2014. Each paper was selected to exhibit a natural evolution in thinking with an emphasis on applied studies designed with a propensity for quantifying disease burden in order to inform and evaluate health policy, programs, practice, outcomes, and risk assessment.

Historically disease burden was quantified using mortality rates, but has evolved to use other metrics such as incidence, prevalence, disability adjusted life years, potential years of life lost, and others (Murray, 1994). It has been suggested that measures used to quantify the burden of disease should be done to: aid in priority setting for service delivery; set research priorities; identify high risk or high need populations; and provide evidence for measuring the impact of interventions (Murray, 1994).

The thesis will present and critique 12 papers in the context of when each was published, highlight unique approaches, and contributions for the use of AHD for quantifying disease burden will be demonstrated. Given that the papers span a 25-year period, earlier papers may appear somewhat simplistic in approach or content, but were unique to their time. In addition, it will describe and critique the methods used, compare and contrast results from other studies that used primary or secondary approaches. Issues identified through the critique may be specific to an individual

paper or may be more general across multiple papers. The specific issues identified for an individual paper are captured as part of the critical commentary of that chapter. However, where the issues are applicable to multiple papers, they are addressed in both the final summative methodological critique in Chapter 9, and concluding sections of the relevant chapter. Ten of the 12 papers are based exclusively on data from the province of Alberta, Canada, and the results will be placed into both the context of Canadian and international studies.

Primary approaches involve the collection of data by the researcher with a specific hypothesis that is often generated through a review of prior work. Secondary approaches draw upon already available data and information that was derived from primary research or administrative processes (Glass, 1976). While the studies included have applicability to the use of AHD broadly, they were conducted in Canada so an understanding of the Canadian context is important. International comparators will be used as appropriate.

In Canada, each province and territory has responsibility for the delivery of healthcare services, with the federal government being responsible for health services for the military, federal inmates, First Nations living on reserves, those living north of the 60th parallel, and international points of entry to the country (Shah, 1998). Within Canada, healthcare services are publicly administered and funded, and universally available to all residents (Baumgart, 1992; Shah, 1998). As part of administering the publicly funded healthcare system, each province and territory maintains a number of population-based AHD sources. These data have value for understanding, funding, and evaluating the healthcare system, as well as supporting research.

The thesis will exhibit the value of routinely collected mortality data to describe geographic and temporal differences across provinces in Canada leading to the development of hypotheses on aetiological factors. Next, the formulation of hypotheses will evolve to the development and use of AHD-based case definition algorithms for improving the accuracy of identifying disease states, use of cross-sectional and longitudinal designs, and how linkage of multiple databases can provide

accurate epidemiological measures. The thesis will also highlight how the results have direct relevance to health services and outcomes research, applied research, practice, and health policy development.

1.2 Structure of the Thesis

Chapter 1 provides the overall aim of the thesis and general information on the topics to be explored in more depth throughout the thesis. Chapter 2 will focus on administrative health data. It provides a definition, highlights strengths and weaknesses of this type of data source for research, and highlights historical developments.

Chapters 3 through 8 are organized to highlight early contributions and the evolution in the sophistication of the use of AHD for quantifying disease burden. Chapter 3 focuses on mortality and hospitalisation data, and provides a critique of strengths, limitations, and value of routinely collected data for describing the distribution of disease, and for developing aetiological hypotheses.

Chapter 4 introduces the concept of developing case definitions for attributing a disease state to a person as a means of providing more accurate estimates of incidence and prevalence of disease. The Chapter addresses issues of concern related to the ability to generate reliable and valid epidemiological measures using AHD, as well as the importance of understanding the sensitivity and specificity of any definition created.

Chapter 5 focuses on the value of longitudinal designs and how AHD can be used as an efficient means of following large cohorts of individuals. Chapter 6 explores the value of AHD for assessing the prevalence of rare conditions and critiques the use of AHD for assessing comorbidities.

Chapter 7 appraises AHD for assessing health outcomes and risk. This includes a critical commentary of the methods and interpretations. Three examples are provided: 1) parental ancestry as a risk factor for multiple sclerosis (MS); 2) MS prevalence

among First Nations people; and 3) febrile seizure risk following the introduction of a new vaccine.

The contribution of research using AHD for policy development is described in Chapter 8. The example provided looks at the impact of an immunisation program for chickenpox on the rates of shingles.

Chapter 9 provides an overall critique of the study designs used across all 12 publications. Chapter 10 presents a summary, conclusions, and recommendations for future research. It highlights the role of AHD in understanding the burden of illness, and lays out options for research to further understand the value these data can provide.

The next chapter will discuss the concept of burden of illness, definition of AHD including its strengths, limitations, and history.

Chapter 2: Administrative Health Data

2.1 Introduction

This chapter will provide a brief overview of administrative health data and its history. The narrative will not be exhaustive, but rather will focus on key developments. Also, given the papers included in this thesis are based on Canadian work, the historical advancement of the use of administrative data will end with a focus on Canadian studies as a means of providing a relevant context for this thesis and the papers included. Prior to providing an overview of AHD, the concept of burden of illness will first be explored, and placed into the context of this thesis.

2.1.1 Burden of Illness

Last (2007) defines burden of disease as:

> "The amount of ill health from a given cause (disease, injury, cause of disease, or risk factor) in a population of interest." (p. 46).

While there are numerous measures such as incidence, prevalence, mortality, disability adjusted life years, and potential years of life lost are available for measuring aspects of burden, an overarching conceptual framework is lacking (Pinheiro *et al*., 2011). However, Pinheiro *et al*. (2011) have suggested that a framework would include targeting the consequences of disease events, may include metrics related to non-health events such as social and economic impacts, and is related to populations or groups rather than on individuals. In general, assessing burden of illness can be considered a means of determining the impact of disease events on all aspects of human life (Pinheiro *et al*., 2011).

The control of mortality has been considered a paramount goal in policy and practice (Sullivan, 1966). Conceptually, this was based on the assumption that changes in mortality were indicative of changes in other aspects of health. Measures based on mortality data have been the cornerstone for assessing burden (Devleesschauwer *et al.*, 2014). However, with increasing importance of non-communicable diseases death-

related statistics are becoming increasingly inaccurate measures of the overall health status of the population (Pinheiro *et al*., 2011).

With mortality due to infectious diseases being replaced more and more by chronic, or non-communicable, conditions, other metrics of burden are required to avoid biasing our understanding of the health of the population. An understanding of the burden of disease provides the evidence needed to promote, enhance and protect the health of a population through policy and practice, and the metrics to support any actions need to be robust and relevant (Murray, 1994; Devleesschauwer *et al.*, 2014).

Actions taken, whether direct or indirect, aimed at improving the health and health needs of the population need to be based on reliable estimations on disease burden and risk (Devleesschauwer *et al*., 2014). Actions also need to be evaluated to ensure burden is reduced and not increased. Within the context of this thesis, burden of illness will not be considered as a single metric or index. Rather, it will focus on the evolution of the use of AHD to quantify aspects of burden and demonstrate the value of AHD in this endeavour.

2.2 Administrative Health Data Defined

Administrative health data (AHD) are routinely collected for the purpose of payment, monitoring, planning, priority setting, and evaluation of the provision of health services (Kreyenfeld & Wilekens, 2015; Virnig & McBean, 2001). The primary purpose of AHD is not research. However, AHD can be a rich source of data to support secondary analyses. Research using these data is considered secondary use when the researcher was not directly involved in the design or collection of the data, nor was the data collected for a specific research purpose (Potvin & Champagne, 1986; Stewart & Kamins, 1993).

The use of AHD for research can be viewed as adding to, rather than replacing, other options for data collection and research, such as randomized controlled trials (RCTs), surveys, case-control or cohort studies, meta-analyses, reviews, or observational studies. Each of these data collection methods has its strengths and

weaknesses. For example, RCTs which are regarded as a gold standard in healthcare research (Evans, 2003) tend to have stringent inclusion criteria with their findings driven by the protocol, unknown impacts of violations to the protocol, limited sample sizes, and relatively short duration of follow-up, all of which can limit generalisability of their findings (Cartwright & Munro, 2010). It is not the use of primary or secondary approaches that is important, but rather the use of an appropriate method to address a given research question. The strengths and limitations of AHD are discussed in the next section.

2.3 Strengths and Limitations of Administrative Health Data

2.3.1 Administrative Health Data Strengths

Like any source of data, AHD has its strengths and limitations, and the use of the data should be done with an understanding of when the data will work and when other strategies are required. The strengths of AHD data over primary data collection for research are: they are population-based; data collection is nonintrusive as it can be obtained without contacting individuals directly; there is no recall bias, information is readily available; collection and use is cost-effective; longer durations of follow-up are possible; multiple databases can generally be linked together; data are usually updated in a systematic and regular manner; and large sample size (Bright *et al.*, 1989; Chowdhury & Hemmelgarn, 2015). Also, the use of AHD, similar to that of a naturalistic observational study, can be used to monitor 'real-life' interactions with the healthcare system (Gallagher, 2004). This allows for a relatively inexpensive and efficient way for assessing the effectiveness of interventions, and to determine if they should continue or if further study is needed (Gallagher, 2004). While AHD have a number of strengths that make their use in research attractive, AHD also have limitations that need to be considered prior to use.

2.3.2 Administrative Health Data Limitations

Limitations of AHD include: lack of information on socio-economic status; case ascertainment is based on utilization of services which may be independent of disease incidence; prevalence, or severity; the researcher was not involved in the design of collection of the data; data privacy and protection issues need to be addressed; accuracy and precision of the data cannot always be confirmed; large sample size leading to statistically significant, but clinical or policy irrelevant findings; policies on reporting or reimbursement may lead to biased case ascertainment; and researchers are dependent on administrative authorities for access (Chowdhury & Hemmelgarn, 2015; Ray, 1997).

AHD generally do not include detailed clinical information such as blood pressure measures, or physical measurements like height, weight, and waist circumference. These may be captured within electronic medical records, but within the Canadian context these data are not generally available at the population level. The presence or absence of detailed clinical information varies by jurisdiction. These data are also generally missing information on patient behaviours such as drinking, smoking, and physical activity (Chowdhury & Hemmelgarn, 2015). Social environment factors such as family structure, stability, size, culture, ethnicity, and other attributes tend to also be missing from AHD. In essence, AHD provides a breadth of data, but is generally lacking depth.

2.4 Administrative Health Data and Big Data

In 1997, Cox and Ellsworth coined the term *Big Data*. They described the problem of increasing amounts of data taxing the ability of computers in the areas of main memory, processing, and storage. Since this time the concept has expanded to include structured and unstructured data, and is essentially inclusive of any digital storage regardless of type. Definitions of big data tend to focus on three key attributes: high volumes of data, high velocity, and high variety (Dutcher, 2014). To land on a more descriptive definition of big data, Dutcher (2014) surveyed 43 thought leaders

from multiple disciplines asking them how they would define the term *big data*. There were a variety of answers with a number of common themes. One definition appeared to particularly represent the wide array of responses:

> "While the use of the term is quite nebulous and is often co-opted for other purposes, I've understood 'big data' to be about analysis for data that's really messy or where you don't know the right questions or queries to make – analysis that can help you find patterns, anomalies, or new structures amidst otherwise chaotic or complex data points." (para. 5).

Based on the above definition, AHD can be considered a form of big data. Mackie *et al*. (2015) have cautioned that big data will likely not solve many epidemiological questions, and that consideration will need to be given to how to deal with 'messy' data. Also, that AHD may be limited in what it can contribute to big data given the need to protect the confidentiality of data. Aside from concerns with issues of confidentiality, Dixon *et al*. (2013) argued for the need to put in place strategies to ensure the integrity and quality of big data, regardless of its source. They contended that AHD may not be robust enough to support all the intended uses proposed by decision makers, and that the implementation of good data quality practices can work to improve the value of these data for decision making. However, Raghupathi and Raghupathi (2014) argued that analytic techniques used for big data could be applied to the healthcare arena offering significant benefits in the areas of clinical operations, research and development, public health, evidence-based medicine, genomic analysis, fraud analysis, device/remote monitoring, and patient profile analytics.

2.5 Administrative Health Data Historical Developments

2.5.1 John Graunt and the Bills of Mortality

The value of routinely collected health information came about in 1532 in London when the town council started to keep a count of the number of people dying as a result of the plague (Choi, 2012; Rothman, 1996). The accounting was done using

what were referred to as the *Bills of Mortality,* which were weekly mortality statistics used to monitor burials.

The *Bills* were collected off and on for more than a century, but never systematically examined until the 1600s (Choi, 2012). In 1662, the English haberdasher John Graunt, considered to be one of the first demographers, published a book titled *Natural and political observations made upon the bills of mortality* (Choi, 2012; Rothman, 1996). This marked the first comprehensive analysis and interpretation of routinely collected health information.

Graunt demonstrated the value of examining data from a perspective different from the original reason for collection. In essence, creating the concept of secondary data analysis (Rothman, 1996). Graunt also explored the data without a clear hypothesis using what is referred to as exploratory data analysis (Tukey, 1977). Graunt allowed the data to guide him on a journey of discovery. The collection of mortality data on frequency and cause evolved over the coming centuries, and remains a major source of evidence for understanding the health status of a population.

2.5.2 Understanding Burden of Disease

A number of notable individuals continued to expand upon Graunt's early work showing the value of mortality data for not only understanding population health, but also to inform and advocate for action. For example, Sir Edwin Chadwick highlighted a strong association between poverty and poor health outcomes in London. This resulted in reforms to the *Poor Law* leading to the emergence of the modern welfare state, as well as the passing of the 1848 *Public Health Act* (Choi, 2012).

One of the most noteworthy contributions to the use of mortality data came from William Farr. Farr worked at the General Register Office for 41 years (1828 to 1879) where he started the practice of collecting and analysing vital statistics on a routine basis (Choi, 2012). Not only did it include routinely collecting, analysing, and disseminating data on deaths, but also the development of standardized processes of collection making the data easier to collect and interpret.

Mortality data has been the cornerstone for estimating the health of the population. It was used for centuries as the only reliable information available. However, it is limited for developing health policy and informing practice (Devleesschauwer *et al.*, 2014). Other metrics of burden are required to avoid biasing our understanding of population health. To be useful, metrics of the burden of disease need to be robust and relevant as they provide evidence to promote, enhance, and protect population health through policy and practice (Devleesschauwer *et al.*, 2014; Murray, 1994). The quantification of disease burden carries an ethical dimension as it informs priorities for resource allocations (Murray, 1994).

2.5.3 Administrative Health Data For Health System Planning

Moving ahead, to the twentieth century, Rosenfeld *et al.* (1957) used administrative data to assess the need for general hospital care as well as reasons for a prolonged hospital stay. Alderson (1976) reviewed the computing policies of the National Health Service in the United Kingdom (UK). He argued the value of a reliable computing infrastructure as a key component of decision-making. This included the concept of a master patient index to allow for accurate identification of individual patients. Prior to this, it was only possible to count events with no ability to understand the number of people represented by those counts. While not mentioned explicitly by Alderson, it also set the stage for ensuring it was possible to link records reliably across data sources.

In 1973, Wennberg and Gittlesohn highlighted wide geographic variations in the use of high cost medical procedures. Where a person lived appeared to determine what types of services would be available. However, the geographic areas responsible for the highest health care expenditures did not necessarily have the best health outcomes. While their work had clear policy implications for resource allocations and monitoring health system performance, it is the concept of using AHD to describe the 'real-world' that represented an important contribution to the evolving use of AHD.

Roos (1989) used AHD as a way to control for case severity when trying to understand health service utilisation patterns. Roos combined hospitalisation and physician claims data, and then compared with self-reported data. AHD outperformed self-report data when the outcomes of interest were admission to a hospital or nursing home, or mortality, showing AHD were potentially valuable for the assessment of outcomes. The value of AHD was becoming increasingly evident from a health system planning perspective, but it was unknown how, or if, these data could be used to understand the health of the population.

Before considering the use of AHD for estimating measures like incidence and prevalence, it is important to understand the properties of the data, and if the data are robust enough to support this purpose. A number of issues need to be considered when using AHD for research. A single diagnostic code could be part of describing signs, interpreting a diagnostic test, a confirmed diagnosis, or an error in coding (Koran, 1975; Roos *et al*., 1982). The data may have *factual errors* that result from the methods used by clinicians to assess and diagnose their patients, meaning that the data may be an accurate representation of an error made in clinical practice. Unfortunately, it is not always possible to review all the evidence used in the clinical decision making process given the cost and logistics involved. In other words, an understanding of how a given health system and its professionals operate is important to understanding and using data derived by the system.

2.5.5 Administrative Health Data for Estimating Disease Burden in Canada

Turning to the Canadian context, the Nova Scotia-Saskatchewan Cardiovascular Disease Epidemiology Group (1989) linked data from inpatient records with vital statistics death registration data in order to estimate the incidence of acute myocardial infarction (AMI). The study was one of the first in Canada to look at the data with the explicit purpose of assessing the incidence of a health event. The authors concluded that linked AHD could be used to assess geographic distribution and temporal trends.

Young *et al*. (1991) conducted a pilot study, to estimate the burden of diabetes mellitus in Manitoba. The authors used five years of data to estimate incidence and prevalence, and then compared their results with three self-report sources – the 1978 Canadian Health Survey, 1985 Canadian General Social Survey, and the United States National Health and Nutrition Examination Survey II. Estimates derived using AHD were comparable to self-report data, leading the researchers to conclude that AHD was a potential data source for epidemiological studies.

To this point, few studies looked at AHD as a potential source for the quantification of disease burden. They had been focused on service delivery patterns, health services research, funding allocations, and health policy development. It was generally not known if AHD could be used to provide accurate estimations of disease burden.

This was the case until the mid-1990s when Svenson *et al*. (1993; 1994) used AHD to estimate the prevalence of Parkinson's disease (PD) and multiple sclerosis (MS) in the province of Alberta using a population-based cohort design. As these were the first studies in Canada to use AHD to estimate prevalence of MS and PD, there were no other similar AHD-based studies from which comparisons could be made. However, prevalence estimates were similar, but tended to be lower, than other studies at that time suggesting AHD-derived estimates were likely to undercount cases (Rajput, 1992; Svenson *et al.*, 1993; Warren & Warren, 1992, 1993). In 1993, Manfreda *et al*. examined physician records in order to estimate the prevalence of asthma, and the possibility for diagnostic exchange, in the province of Manitoba. Manfreda *et al*. (1993) found that there did appear to be evidence of a change in the use of the diagnostic codes that could influence prevalence estimation. This demonstrated that practice patterns play a role, and need to be understood when using AHD.

As the 1990s advanced, the number of Canadian papers using AHD increased substantially creating a need to better understand the validity of the data. In 1997, Robinson *et al.* reported results from a linkage of survey data with AHD in Manitoba concluding that using a single diagnostic code from AHD may increase the inclusion of

false positive cases, however, the use of a more stringent case definition increased the false negative error to a greater degree. The authors failed to provide performance metrics such as sensitivity, specificity, or predictive values, but did use Kappa statistics to assess agreement between the sources. In principle, their study was focused on congruence between data sources more so then on the validity of AHD.

The research demonstrating the value of AHD for estimating incidence and prevalence resulted in work to validate approaches to better understand the sensitivity and specificity of any approach used (Lacasse *et al.*, 2012; Marrie *et al*., 2013; Southern, *et al*., 2010; To *et al.*, 2006). Also, AHD became the backbone for the surveillance of chronic diseases in Canada (Clottey *et al*., 2001; James *et al.,* 2004; Lix *et al*., 2008). More recent work, done within the context of the Canadian healthcare system, has used increasingly sophisticated methods and a move to assessing comorbidities and health outcomes (Patten *et al*., 2007; MacDonald *et al.,* 2014; Marrie *et al*., 2013; Yurkovich *et al*., 2015).

2.6 Summary

The use of AHD has evolved from being used to reimburse health care providers and allocate funding, to a better understanding of the operation of the health care system, policy development, resource allocation, accountability, and more recently, for assessing the health of the population. Prior to 1990, AHD was rarely used to estimate disease prevalence or incidence. This represented a gap making it clear that there was an opportunity to explore how AHD could address the dearth of information on disease burden. Work of the present author, and others, began to show that the creative use of AHD could be valuable as a means of better understanding health burden.

Chapters 3 through 8 will provide a critical appraisal of the 12 papers included in this thesis. They will highlight the study design, findings, strengths, limitations, and implications as they relate to the use of AHD for the quantification of disease burden.

Chapter 3: Role of Mortality and Hospitalisation Data

3.1 Introduction

Mortality data represent one of the few data sources collected in virtually all countries and have been a mainstay for determining health priorities, and understanding population health for hundreds of years (Devleesschauwer *et al.*, 2014). Hospitalisation data has been used primarily for resource allocation, and at times disease burden, though its history is much shorter. This chapter will provide a critique of Publications 1 (Svenson, 1990) and 2 (Svenson, 1991) with a focus on the value of mortality and hospitalisation data for assessing disease burden. Both studies focused on the geographic distribution of Parkinson's disease (PD) (Svenson, 1990; Svenson, 1991). A cross-sectional study design was used for both, and the critique will highlight strengths and limitations of each study as well as their contributions to understanding disease burden. Cross-sectional designs are generally used to assess the frequency of an outcome in a given population at a particular point or period of time (Levin, 2006a).

3.2 Mortality Data

3.2.1 Study Summary and Critique – Publication 1

Mortality data had been used in a number of countries to examine the frequency and distribution of Parkinson's disease (PD), but had not been analysed in Canada until this study (Kurtzke & Goldberg, 1988). Publication 1 was the first, and so far only, Canadian study to look specifically at the geographic distribution of deaths attributed to PD using vital statistics records. The study found an uneven geographic distribution of PD in Canada, and suggested this supported aetiological hypotheses for PD centred on the role of environmental factors, primarily toxins.

Using publicly available data, from Statistics Canada, PD mortality rates were computed by year, sex, and province. The purpose of the study was to examine the geographic distribution of PD in Canada to see if there was any clustering as a means of generating potential aetiological hypotheses. Geographic clustering of disease may

imply that environmental factors play a role the aetiology of a given disease. However, it cannot not be concluded from descriptive analyses that this is, in fact, the case.

The geographic differences noted showed higher mortality rates for PD in the western provinces of Canada. These provinces also tend to have larger agricultural activity (Statistics Canada, 1999). These geographic differences were used to conclude that environmental factors likely played an aetiological role. Given the cross-sectional design, it was not possible to make inferences on the role of agricultural activity, but it was possible to develop hypotheses about the role of agricultural activity, and the possibility of chemical or other toxin exposures. Cross-sectional designs can be used to identify difference between groups, though are not well suited for understanding why the differences exist.

Until this study, researchers had to rely on a limited number of studies conducted at different points in time, different locations, and often with different methods (Warren *et al*., 2005). This study provided a systematic assessment of the geographic distribution of PD. As a result, it was possible to generate hypotheses that supported other literature on aetiological hypotheses (Rajput, 1992). This demonstrated the value of using multiple approaches to improve our understanding of the burden of disease in general, and PD specifically.

There was consistency with other studies conducted around that time pointing to higher prevalence among individuals living in more rural areas (Hertzman *et al.,* 1990; Kurtzke, 1977; Rajput, 1992). Consistency of findings across studies and locations, even if methods were different, increases the strength of the findings. In this case, consistency in finding higher disease prevalence in rural areas helped to refine aetiological hypotheses with a focus on what may be unique to the rural environment.

There were limitations related to the study. These included a lack of information on birthplace, places of residence, age of disease onset, severity of illness, and that PD may not have been recorded as the underlying cause of death. The recording of the underlying cause of death may be biased. It is unknown what information the physician used, and it is often not possible to determine if the person

died from or with a given disease. Studies have shown that among individuals with PD that have died, between 14.6% and 64% of descendants had PD listed on their death certificate (Benito-León *et al*., 2014; Paulson & Gill, 1995). Given the potential to miss a significant proportion of PD cases through the use of mortality data, it becomes difficult to make valid inference. Those with more severe presentations may be more likely to have PD listed on the death certificate.

There may be other issues with mortality data that could compromise the interpretation of any findings. These include personal preference on where to die which could impact the assessment of geographic distribution; not all deaths result in an autopsy so important causes of death may go missed; and vital statistics agencies typically do not confirm the accuracy of the information submitted meaning coding errors could go undetected (Nielsen *et al*., 1991; Schnatter *et al*., 1990). These limitations, if not considered, could lead to misinterpreting evidence on the burden of not only PD, but also any condition. This could lead to the misallocation of services, and other resources.

The limitations noted were offset by strengths such as: being population-wide coverage due to provincial laws requiring death certificates to be completed; uniformity of content and format set nationally; continuously collected over time; and the use of international standards for definitions and coding. Also, mortality data are accessible and well documented. Having systematically collected, population-based data, allowed for reliability of the results for assessing the geographic distribution. Also, mortality data represent one of the few sources of population level data collected internationally. This allows for comparisons that could not be done using other sources of information.

The study recommended future research focused on environmental factors in areas with higher mortality rates, and suggested that augmenting routinely collected data with other designs, like case control studies, would strengthen our understanding of potential risk factors for PD. It also recommended the use of hospitalisation data for estimating prevalence. The implications of these findings were that mortality data

could be used to assess geographic distribution of disease, and for hypothesis generation.

3.3 Publication 1

Svenson, L.W. (1990). Geographic distribution of deaths due to Parkinson's disease in Canada: 1979 – 1986. *Movement Disorders*, *5*, 322-324.

Link: http://onlinelibrary.wiley.com/doi/10.1002/mds.870050412/abstract

3.4 Hospitalisation Data

Publication 2 (Svenson, 1991) builds on the results from Publication 1 (Svenson, 1990). In Publication 1, Svenson (1990) suggested that mortality data, as a measure of disease burden, could be improved through the examination of hospitalisation data.

3.4.1 Study Summary and Critique – Publication 2

Publication 2 (Svenson, 1991) was a cross-sectional study that examined hospital separation (death of discharge of an inpatient) data from Statistics Canada to estimate the prevalence of Parkinson's disease by Canadian province. The study found similarities as well as a few differences from the examination of PD mortality rates (Svenson, 1990).

A male preponderance for PD was noted, and this was consistent with other hospital-based studies as well as with the findings of Publication 1 (Kessler, 1972; Kurtzke & Murphy, 1990; Svenson, 1990). The geographic distribution of hospitalisation rates differed somewhat from that noted for mortality. Such differences may be the result of a number of factors such as coding practices, underlying severity of disease, differences in survival, or differences in likelihood of being hospitalised. The prevalence of disease was only one factor of many that would influence the likelihood of being hospitalised, and subsequently identified as a prevalent case.

The concept of using differing approaches, and data sources having different value and purpose was proposed in this paper. While there were differences between Publication 1 (Svenson, 1990) and Publication 2 (Svenson, 1991) related to the geographic distribution across Canadian provinces, both studies did note higher rates in provinces that tended to have higher levels of agricultural activity, or proportion of the population living in rural areas. This added further support to the view that PD resulted from exposure to environmental factors.

While hospitalisation data are generally viewed as being accurate, there are a number of important limitations, some of which were highlighted in Publication 2 (Svenson, 1991). There were two key limitations of the paper. First, the data were

based on the number of hospital separations, the discharge or death of an inpatient, and not the number of unique patients. Second, the analysis was based on location of hospitalisation, and not the province of residence of the patient. Because the data were based on separations and not unique individuals, there was the possibility of counting the same person multiple times resulting in an over estimation of prevalence. Hospitalisation data tends to bias results towards those most ill or in need of health services, as not everyone with PD will require hospitalisation (Butt *et al*., 2014).

While not known at the time of the study, more recent work has shown that the use of hospitalisation data may significantly under estimate the burden of PD (Butt *et al*., 2014; Danila *et al*., 2014). This supports the view that it is the more severe cases that are hospitalised. This would lead to a bias in understanding the burden of PD as the cases examined would over represent severe presentations, and under estimate less severe disease. This suggests that hospitalisation data should not be used for the estimation of measures like incidence and prevalence, but could potentially be used as a measure of burden among those with PD. Policy or resource allocation decisions for PD, based solely on hospitalisation data, could have unintended consequences as the allocation could be based on biased information. The bias would be to underestimate the prevalence of disease while over estimating the severity. Limiting decisions to the hospital setting may be prudent when only hospitalisation data are available.

3.5 Publication 2

Svenson, L.W. (1991). Regional disparities in the annual prevalence rates of Parkinson's disease in Canada. *Neuroepidemiology*, *10*, 205-210.

Link: https://www.karger.com/Article/Abstract/110271

3.6 Summary

Both studies represented novel approaches, in 1990 and 1991, for trying to estimate the prevalence and burden of PD. While both studies helped to foster a better understanding of the epidemiology of PD, more recent research has shown that hospitalisation data may not be the best source for assessing prevalence (Butt *et al*., 2014; Danila *et al*., 2014). Despite this, the fact that resources were being used meant the data had value for health service planning and resource allocation. It is the number of hospital separations, more than the number of people, which drives resource consumption. As such, a good understanding of hospital use can lead to not only efficient resource allocation, but also lead to work to define appropriate use, and identify optimum models of care.

The use of either mortality or hospitalisation data must consider the primary purpose of the study. For example, mortality data provides a more accurate indication of survival or clustering of deaths temporally or geographically, while hospitalisation data will provide a better perspective on health system utilisation. Both may be used to develop hypotheses about aetiological factors, but caution is needed when used for health system planning, resource allocation, or determining health priorities. While hospitalisation data provides some indication of health system use, it does not allow one to know what proportion of individuals with PD end up hospitalised. It is possible that the more severe case are hospitalised, and this would result in a biased assessment of burden.

Both studies represented an early part of the journey towards a better understanding of AHD and how it might be used to understand the burden of disease. The next chapter will examine how issues of data validity for estimation of incidence and prevalence can be addressed by using an algorithmic approach that combines multiple data sources.

Chapter 4: Case Definition Development to Improve Data Quality

4.1 Introduction

Population-based data on the incidence and prevalence of chronic diseases is important for understanding burden, setting priorities for resource allocation, and informing practice. Two sources of population-based data are generally available for this purpose – AHD and survey data. The accuracy of diagnosis has been considered a key limitation, or concern, cited with both surveys and AHD (Lix *et al*., 2008). For example, Wilchesky *et al*. (2004) assessed the validity of diagnostic information for 18 conditions using medical claims data from Quebec, Canada and found the validity of codes varied by condition and that, in general, the specificity tended to be high, while sensitivity tended to be low.

A case definition is a set of criteria used to establish a diagnosis (Last, 2007). Because AHD does not include details on the criteria used by a given physician, case definition algorithms can be created to minimize the risk of false positive detections, and to improve the accuracy of the data. A case definition algorithm selects records from one or more AHD sources based on a combination of visits with specific diagnostic or procedure codes over a fixed period of time (Lix *et al*., 2008). The source of the data determines the number of services one would need to have to qualify as a case. The purpose of the analysis determines if one chooses to err on the side of sensitivity (gives few false negatives, but may introduce false positives) or specificity (gives few false positives, but may introduce false negatives). To highlight considerations and contributions to the concept of developing case definitions as a means of improving the quality and utility of AHD, Publications 3 (Robertson *et al*., 1998), 4 (Yiannkoulias *et al.,* 2003) and 5 (Yiannakoulias *et al*., 2007) will be critiqued.

4.2 Case Definition Development – An Example with Cerebral Palsy

Cerebral palsy (CP) is considered to be one of the most common motor disabilities of childhood. At the time of this study, there was debate in the literature

about the incidence and prevalence of this condition. The debate centred around views that the incidence of CP was changing as a result of changes in the percentage of preterm births (births prior to 37 weeks gestation), low birth weight births (weight less than 2,500 grams), and increasing survival among very low birth weight new-borns (weight less than 1,500 grams) (Pharoah *et al*., 1990; Stanley, 1994; Stanley & Blair, 1991).

4.2.1 Study Summary and Critique – Publication 3

CP can be difficult to diagnose, often requiring multiple assessments over a period of time (Jutte *et al*., 2011). To address this problem, the current study used a unique approach by using multiple birth cohorts, and seven years of follow-up. This allowed for the accumulation of service events related to CP for each child. It was possible to look at the proportion of children receiving services for CP before and after their third birthday. Next, to avoid inclusion of children as CP cases that may have been misdiagnosed, services for conditions that would be considered for a differential diagnosis were also included in the analysis. This allowed for an exclusion of potential false positive cases leading to improved specificity.

For the case definition algorithm, a child was considered to have cerebral palsy if they met the following inclusion criteria: (1) the child received a medical service where a diagnostic code was provided for cerebral palsy when the child was three years of age or older or (2) any child with a cerebral palsy diagnosis prior to age three that had at least one additional service when they were three years of age or older. Following the inclusion criteria, the child was reclassified as being a false positive case if they had medical services with codes for other progressive neurological disorders. This approach was designed to err on the side of specificity ensuring there were few false positive cases included in the estimation of incidence.

This study demonstrated that AHD could be used to develop retrospective cohorts using inclusion and exclusion criteria for case identification. A closed cohort approach was used which fixes the members of the cohort so that no new members

are added. Existing members can only exit by moving from the province or dying. This limited the potential for confounding factors to influence the findings.

Within the province of Alberta, this was the first population-based study of CP. It augmented previous work that was only focused on new-borns weighing less than 1,250 grams at birth by providing prevalence estimates for the entire range of birth weights. A better understanding of the impact of birth weight and CP can help with estimations of special needs assistance, and other services for children with CP as well as for their caregivers. It also showed that there was no difference in prevalence across three birth cohorts. Differences in prevalence across birth cohorts could indicate issues with the prevalence of underlying risk factors. The estimates derived in the study were consistent with those found by others (Murphy *et al*., 1993; Stanley, 1994; Stanley & Blair, 1991).

There were also limitations that require consideration. First, no validation was done on the diagnostic codes included in the study. While an algorithmic approach was used to improve specificity, the study would have benefitted from examining charts for a sample of the children. This would have provided additional evidence on the accuracy of the approach used. While the lack of validation does not mean the data are invalid, the assessment of validity would work to strengthen the credibility of the findings.

Case capture was based on utilisation of health services. Variations in how services were accessed, as well as where, may limit the diagnostic information available for assessing inclusion and exclusion criteria. Assuming the approach ensured all cases classified as having CP were correct, it is likely that this increased the number of false-negative cases. The application of a case definition algorithm would err on the side of specificity. This would work to underestimate the incidence and prevalence of the condition. This could lead to poor decisions on the number and type of health professionals needed to meet the demand of newly diagnosed cases. Also, it may negatively impact decisions on the placement of specialty services, particularly by geographic location, that could result in availability and accessibility issues for managing cases.

4.2.4 Publication 3

Robertson, C.M.T., **Svenson, L.W.,** & Joffres, M.R. (1998). Prevalence of cerebral palsy in Alberta. *Canadian Journal of Neurological Sciences*, *25*, 117-122.

Link:

http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=9504712&fulltextType=RA&fileId=S0317167100033710

4.3 Case Definition Development – Geographic Considerations

Cerebrovascular disease is the third leading cause of death in Canada making it an important contributor to disease burden, for the individual, their families, and the health care system. AHD have been recognised as a cost-effective, timely, and generalisable resource for the surveillance of cerebrovascular disease (Tirschwell & Longstreth, 2002). However, it has been noted that the accuracy of diagnostic codes for cerebrovascular disease varies across studies (Reker *et al*., 2001). There appeared to be uncertainty in the quality of the coding of a cerebrovascular diagnosis as well as variability in coding practices across different geographic areas (Liu *et al*., 1999). These factors make it difficult to understand if temporal or geographical variability is true or related to issues of misclassification or coding practices. The purpose of this study was to describe and evaluate geographic differences in cerebrovascular disease.

4.3.1 Study Summary and Critique – Publication 4

Given the above limitations, there was a need to understand the issue in greater detail. To do so, inpatient and outpatient data were linked from all hospitals in Alberta for patients receiving one or more cerebrovascular diagnoses during the 1999-2000 fiscal year (April to March).

The study noted important differences in how urban and rural hospitals coded individual cases. Rural facilities tended to use less specific coding for cerebrovascular disease than did urban facilities. However, for rural patients that were transferred to an urban centre, when they returned to the rural facility they retained the more specific coding of the urban facility on the administrative health record.

The approach and results demonstrated that assessing the burden of cerebrovascular disease using AHD required a large-scale person-oriented approach. The differences between urban and rural facilities was an important contribution as it demonstrated that access to both specialists and diagnostic imaging appeared to result in more specific as well as a broader range of cerebrovascular disease coding. Our findings showed the value to a person-oriented approach, particularly when trying to

assess geographic differences in the distribution of disease. By following individual patients, it was possible to attach a more specific, and potentially more accurate diagnosis, to the patient for those that travelled between rural and urban facilities. The study also provided considerations for the interpretation of geographic differences, particularly when there may be factors that may bias results, such as the availability or access to specialists, and specialized equipment.

The type of cerebrovascular event (ischaemic, haemorrhagic) may result in different approaches to managing the patient, and understanding their prognosis. Also, the geographic distribution of stroke type could help to identify potential risk factors that are unique to a geographic area. The frequency and type also provide information for service level planning, resource allocation, and staff training. The lack of specificity of coding in rural facilities suggests a need to better connect them to specialists within urban centres as a means of triaging, assessing risk, and need for medical transport. From a research perspective, our results demonstrate the need for a person-oriented approach to analysing data, and that results may be influenced by practice patterns. The study highlighted the importance of understanding not just the data source, but also the activities in the health system that might impact the precision of coding, and subsequent interpretations of the data.

While the study was well designed for its purpose, there were limitations that could impact the interpretation of findings. While there appeared to be coding practice differences that may be attributable to access to diagnostic imaging technology or specialists, the study did not look at the availability of either. The diagnostic codes provided were not validated, but assumed to be an accurate reflection of the information on the patient chart. It was also not possible to assess patient preferences in the type of facility they would visit or where the patient was at the time of the event that resulted in needing to be hospitalised. Despite these limitations, the paper demonstrated the importance of understanding the context in which the data were collected as it aids with interpretation, and use of the findings.

4.3.4 Publication 4

Yiannakoulias, N., **Svenson, L.W.,** Hill, M.D., Schopflocher, D.P., James, R.C., Wielgosz, A.T., & Noseworthy, T.W. (2003). Regional comparisons of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chronic Diseases in Canada*, *24*, 9-16.

Link: http://publications.gc.ca/collections/collection_2009/aspc-phac/H12-27-24-1E.pdf

4.4 Case Definition Development – Impact of Changes to Patient Residence

Many chronic conditions vary geographically, and this may be used to generate hypotheses about aetiological agents or to facilitate the planning of health services. Unfortunately, a number of factors can complicate the interpretation of geographic clustering. These include, latency periods from exposure(s) to disease onset, small sample size, physician practice pattern variations, availability of health services, and health service seeking behaviour of individuals with a chronic condition. Boyle (2004) has speculated that those with a chronic condition may be disproportionately more likely to move following the onset of their condition. If this were true, then geographic differences in disease prevalence will not represent variations in *in situ* risk, but rather systematic changes in residence.

The issue may negatively impact public health policy and programs developed in light of a perceived risk to a given geographic area. An inaccurate understanding of the geographic distribution could lead to less than optimal allocation of resources, whether they be funding, personnel, or physical infrastructure. Also, some areas may be viewed as better or worse for providing services to chronically ill individuals, and this has implications for resource planning, and allocation, as well as measures of equity in healthcare.

4.4.1 Study Summary and Critique – Publication 5

To better understand the relationship between the onset of a chronic illness and the probability of a change in residence, Publication 5 used administrative health data to set up a longitudinal study to quantify changes in residence among individuals newly diagnosed with either multiple sclerosis (MS) or Parkinson's disease (PD), and compared this to age- and sex-matched controls. In addition to quantifying the proportion of individuals changing residence, we also estimated the impact this change in residence would have on geographically based prevalence estimates.

To examine changes in residence between 1994 and 2004, data were extracted for the period 1983 to 2005. A case definition algorithm was used to reduce the risk of

including false positive cases. The run-in period included all data between 1983 and 1993 was a means of excluding any individuals that may have had a history of MS or PD prior to the study period. Cases were matched on sex, age ($\pm$ 5 years), socioeconomic status, and municipality of residence. Changes in residence were then identified through the Alberta Health Care Insurance Plan (AHCIP) population registry.

Individuals with MS or PD had different mobility characteristics than did the control group. Newly diagnosed cases of MS or PD were more likely to move following their diagnosis, but this was primarily within the same municipality. There was a tendency for both MS and PD groups to move to or between the two largest urban centres in the province, and both groups were significantly less likely to move out of province relative to their respective controls. While there were changes in residence following a diagnosis of MS or PD, the impact of the mobility was assessed to be minor when developing regionally based prevalence estimates.

Using a longitudinal design and linking data from multiple AHD sources, it was possible to compare changes in residence among a large number of individuals with two chronic conditions, and compare them with the general population. The design included the concept of a run-in period to increase the likelihood that those identified were newly diagnosed cases of either MS or PD. Linkage to a population-based health insurance registry that tracks changes in residence with dates on when changes were made provides a valuable tool for assessing residential changes. Having the estimated incidence date and dates of any moves allowed for changes post diagnosis to be tracked. Being able to assess address changes in this manner would appear to be superior to asking individuals for their address histories, which would introduce issues of recall bias. This would be particularly problematic for PD where cases are generally older.

Limitations of the study were primarily related to the nature of AHD in general. It was not possible to assess the true incident date for cases. We could only estimate this based on contacts with the health care system. Individuals who had either

condition, but who were not yet diagnosed, might have been more or less inclined to move, but we would not be able to assess this or determine if it had an impact on the migration patterns. This could result in an underestimation of any changes in residency.

Even with the extended run-in period, it is possible that some individuals were diagnosed with either condition but did not have any services with the specific diagnosis leading up to the study period. Thus, any diagnosis received during the study period would be viewed as incident, when they were prevalent. This also applies to any individual receiving a diagnosis from a health care provider outside of the province. This information would not be available within AHD. The study only examined the first change in residence following a diagnosis, and not all subsequent changes were captured. Finally, postal codes were used for geo-referencing, and these are not equally precise across the province. In rural areas, a single postal code may cover a large area including multiple communities. If the communities differ significantly from each other on attributes related to the likelihood of moving, then the results may be biased towards a null finding even if one or more of the communities had members more likely to move.

An understanding of migration patterns helps to improve aetiological hypotheses. If individuals with either disease have been resident in an area for their whole life, or a significant proportion of their lives, then one could look to environmental factors unique to, or in high concentrations, within that geographic area. Also, it helps to quantify the likelihood that patients and their families might move following diagnosis. Understanding migration patterns helps to inform decisions about the allocation of resources, such as health professionals and clinics. However, not knowing why the person moved, it is difficult to fully understand the implications from a health policy development perspective or to fully address patients' service needs.

4.4.4 Publication 5

Yiannakoulias, N., Schopflocher, D.P., Warren, S.A., & **Svenson, L.W.** (2007). Parkinson's disease, multiple sclerosis and changes in residence in Alberta. *Canadian Journal of Neurological Sciences*, *34*, 343-348.

Link: http://dx.doi.org/10.1017/S0317167100006806

4.5 Summary

While creating case definition algorithms using multiple services over a specified period of time appears to be a valuable approach, there are other considerations that needed to be assessed. Diagnostic coding practices change across geographic area, with urban centres more likely to use both specific codes as well as a broader array of codes. This can impact our understanding of the distribution of cerebrovascular disease type (Yiannakoulias *et al.*, 2003). While this issue may not impact estimations of the burden of cerebrovascular disease, it does need to be considered when interpreting any study that included urban and rural facilities with differing access to diagnostic technology and medical specialists. Region-to-region differences need to be understood to ensure appropriate interpretation of findings. While the example used cerebrovascular disease, the findings would generalise to any chronic condition that requires specialised tools for confirming the diagnosis.

Understanding if people move closer to specialty services can play a role in understanding current and future health service needs. This chapter demonstrated that migration does occur following a diagnosis of a chronic condition (i.e. MS, PD), but does not occur at a level that would impact the geographic estimation of prevalence, nor negatively impact the development of aetiological hypotheses, or inform resource planning and allocation (Yiannakoulias *et al.,* 2007).

The three publications demonstrated that a thorough understanding of AHD and the application of novel approaches could lead to better understanding the burden of disease. The next chapter will explore the value of longitudinal designs for improving our understanding of disease burden and risk.

Chapter 5: Using Administrative Health Data for Longitudinal Designs

5.1 Introduction

Cohort designs, a form of observational study, are used to look at associations that may be difficult, or unethical, to assess via other means (Patten, 2015). The term cohort is used to describe any group that has common features, and is typically divided into either prospective or retrospective. Cohort studies typically start with a group free from an outcome, and follow them to forward to determine which members develop the outcome of interest as a way of clarifying the temporal relationship between an exposure and outcome (Patten, 2015).

While cohort designs are well suited to determining temporality and relative risk measures, among others, they have issues. First, they tend to be expensive to conduct, and may require lengthy follow-up periods. As the follow-up period increases, issues of attrition may introduce bias. Also, prospective studies are typically inefficient for studying rare conditions given the large cohort size needed. Given the cost, logistics, and complexity associated with the development of prospective cohort studies, AHD is an attractive alternative. While AHD may not have in-depth details on each case that comes from primary data collection, they can overcome issues of cost, timeliness, and attrition. This chapter will examine and critique Publication 6 (Yiannakoulias *et al*., 2004) that used a longitudinal design to assess patient to physician and physician to patient disease transmission.

5.2 Study Summary and Critique – Publication 6

This retrospective longitudinal study was initiated, in part, to provide evidence for the development of a pandemic influenza response plan for the province of Alberta. One of the assumptions embedded in many pandemic plans was that healthcare workers would have a higher probability of being infected as a result of the number of patients they would see that had influenza (Wilson *et al*., 2005). The current pandemic response plan for the province of Alberta includes a section that speaks to the

increased demand on healthcare workers as a result of a pandemic. It recognizes that the demand for healthcare services will be higher due to the number of infected individuals, as well as likely reductions in health workforce resulting from healthcare worker illness, need to care for family members, or fear of becoming ill (Government of Alberta, 2014).

Without specific information on the direct impact of a pandemic on healthcare worker illness, non-pandemic studies can be used to provide a lower bound estimate of the impact of a pandemic on the health workforce. Prior to our study, research predicted that approximately one quarter of healthcare workers could be infected, and many would choose to continue working, even while infectious (Odelin *et al*., 1993; Wilde *et al*., 1999).

Using a longitudinal design, Publication 6 (Yiannakoulias *et al.*, 2004) examined two key questions. First, does the probability of a physician being infected vary with the number of patients seen with influenza-like illness (i.e. influenza, pneumonia, bronchitis)? Second, does a physician, who had been treated for an influenza-like illness, infect his/her patients? To answer these questions, we developed a unique approach to the use of administrative data by placing the physician in two roles – one as a healthcare provider, and second as a patient.

To ensure we could limit the data as much as possible to answering the second question about to the physician-patient interaction, a matched design was used. The study physicians were those who had received an influenza-like illness (ILI) diagnosis, and were then matched to physicians who did not have an ILI diagnosis. Matching was done on age ($\pm$ 5 years), gender, residence location, medical specialty, and the number of claims submitted in a year.

The study found that the number of patients, with influenza-like illness, seen increased the probability a physician would be diagnosed with influenza-like illness. When examining the odds of a physician infecting a patient, we found no statistically significant increase in the number of patients developing ILI following a visit with an

infected physician. The study showed the risk was primarily in the direction of patient to physician. This supported the need to ensure high vaccination rates among physicians as well as the use of other preventative measures such as good hand hygiene practices (Aiello *et al*., 2010). This is important given the recent study by Aoyagi *et al*. (2015) showing healthcare workers indicated a moderately high willingness to work during influenza pandemics, and that this was driven, in part, by their perceptions of risk.

A number of strengths can be identified with the approach used. There was no need for contact with physicians, or their patients, reducing recall bias, and the chance of low participation rates. Physicians were placed in two roles – one as a health care provider, and second as a patient. Having access to data on health system encounters in both roles allowed us to quantify the number of patients seen, and the diagnoses provided leading up to the physician being diagnosed themselves. Once the physician had been diagnosed, it was possible to determine if the physician had subsequently transmitted disease to future patients. By having the dates of all services, it was possible to look at this from a temporal perspective. The number of patients seen could be quantified as a risk factor as well as the subsequent risk for patients seeing an ill physician. The use of a control group helped to reduce the potential for confounding.

A number of limitations do need to be highlighted. First, the purpose was to provide empirical evidence on the risk of a physician treating a person with ILI becoming ill. This was done using data during a seasonal and not pandemic situation. Pandemic influenza strains tend to be more virulent, and spread more rapidly than do seasonal strains (Kasowski, *et al*., 2011). As such, the estimations of risk were likely underestimates. The diagnosis provided for both the physician and the patients could not be confirmed. It was not possible to assess how many of the cases were laboratory confirmed, or the criteria used if the diagnosis was based on a clinical assessment.

The data were based on when the person sought medical care making it difficult to fully assess the period of communicability. If a person presented late in the disease course, he or she might no longer be infectious, and this would lead to an

underestimation of risk. It is not known if physicians seek medical care in ways that are consistent with the general population. It is possible for physicians to self-diagnose, or to be seen by a colleague in the same practice with no record created. Approximations of patient to physician and physician to patient transmission would be underestimates.

5.5 Summary

The publication showed a novel application of a longitudinal design based on AHD. It contributed new knowledge on the risk of influenza-like illness transmission between patients and physicians, and from physicians to patients. While it was clear that the effects were underestimates, they still provide empirical evidence that transmission does occur. The more patients with ILI a physician sees, the more likely the physician will become ill him or herself. It does suggest that in a pandemic situation, physicians are at an increased risk of being infected, but if they choose to practice their future patients will not have a significant increase in disease transmission from their physician.

The study followed a cohort of physicians and their patients at a population level. It would have been difficult to perform this type of study using primary data collection approaches. A large prospective cohort would be required with data collection carried out across numerous family practices. The sample size that would have been required would have made such a study challenging and resource intensive. While the results have applicability for pandemic influenza planning as well as understanding physician risk in non-pandemic situations, the approach could be generalised to other situations. Longitudinal designs provide strong evidence, and can be conducted readily using AHD.

The next chapter will look to the use of AHD for estimating the prevalence of a rare health event, and also look at the value of AHD for determining the burden of comorbidities.

5.6 Publication 6

Yiannakoulias, N., Russell, M.L., **Svenson, L.W**., & Schopflocher, D.P. (2004). Doctors, patients and influenza-like illness: clinicians or patients at risk? *Public Health*, *118*, 527-531.

Link: http://dx.doi.org/10.1016/j.puhe.2004.02.005

Chapter 6: Assessing Rare Events and Comorbidities

6.1 Introduction

The strength of AHD becomes apparent when trying to quantify the prevalence or incidence of rare events, as primary data collection is often cost-prohibitive. There is no single accepted definition of rare disease available. However, the European Commission and the Canadian Organization for Rare Disorders both define a rare disease as one with a prevalence that is less than 5 in 10,000 (Canadian Organization for Rare Disorders, 2015; European Commission, n.d.). Rare conditions represent unique challenges. Population-based surveys that are large enough to ensure reliable estimates are often cost prohibitive and logistically difficult. Accessing data from disease specific societies or charities can be challenging, and result in biased samples. Given these challenges, population-based AHD may provide a viable solution to determining disease incidence and prevalence.

Multiple measures are available for assessing the burden of illness (for example incidence, prevalence, mortality, case fatality) (Patten, 2015). The complexity of an illness is increased with the number of comorbidities of the patient, and this accounts for a significant amount of the variation in health care costs across patients (Engström *et al*., 2006). Using Publications 7 (Svenson *et al*., 1999) and 8 (Patten *et al*., 2007) as examples, this chapter will highlight the value of AHD for assessing rare events and for identifying comorbidities.

6.2 Assessing Rare Events – An Example with Motor Neurone Disease

6.2.1 Study Summary and Critique – Publication 7

Motor neurone disease (MND) refers to a group of rare progressive neurological disorders that attack motor neurones that control voluntary muscle movement, with amyotrophic lateral sclerosis (ALS) being the most common (Adams *et al*., 1997). MND has a median survival of two to three years following diagnosis (Lee *et al*., 1995).

At the time of our study, there had only been two Canadian studies that assessed the epidemiology of MND. The first study was conducted by Hudson *et al.* (1986) and focused on MND in southwest Ontario between 1978 and 1982. The second study examined incidence of MND in the province of Nova Scotia (Murray *et al*., 1974). Both studies relied on a review of records from either a specialty clinic or from a single hospital thus limiting the generalisability of their findings.

We used a cross-sectional study design to estimate the prevalence of MND. Given that most studies on the epidemiology of MND were either community or clinic-based with limited ability to generalise to the population level, we felt that AHD may be able to provide a valuable contribution.

While it is not always possible to confirm the diagnostic information available through administrative health data, it was possible to examine both the location of any given service as well as the specialty of the physician. We found that 84% of the cases identified had been diagnosed by a neurologist, neurosurgeon, or internal medicine specialist, and that 66% were seen at one of the two neuromuscular clinics operating in Alberta (Svenson *et al.*, 1999). While access to specialists and specialty clinics may vary by jurisdiction, we were able to show that even in a situation where there are universal publicly funded health services available, approximately one third of patients were not seen by either a specialist or specialty clinic. This suggested that there might be issues of completeness of case ascertainment if one only uses specialty clinics as a source for estimating the incidence or prevalence of MND. The study demonstrated that administrative health data appeared to be a reliable means of studying the epidemiology of this rare condition.

Limitations to the study need to be acknowledged. First, the assessment of prevalence is dependent on individuals with MND both seeking medical services and receiving a diagnosis for MND. Given the rapidly progressive nature of MND, it was assumed that most, if not all, of the cases would have been identified. There was no validation of the diagnostic information available, or how physicians assessed cases, and determined the diagnosis. However, close to three quarters of the cases had been

seen by a medical specialist, and close to two thirds received their diagnosis from a neuromuscular disease specialty clinic. The third potential limitation was that of immigration. For any individual moving into the province that was diagnosed outside of the province then the initial diagnosis information would not be available. However, the progressive nature of the disorder would work towards limiting migration, and given this was an assessment of prevalence, it can be assumed that they would eventually receive a code for MND on one or more of their contacts with the health care system. Those diagnosed outside the province could impact measures like incidence, but once identified in Alberta, they would be included as a prevalent case. Any delays in reporting would mean under estimating prevalence.

Understanding the prevalence of a condition allows for a better understanding of the health and health services needs of a population. With rare conditions, it is important to understand where cases reside to ensure appropriate services can be provided, whether they are institution or home-based. The use of AHD for this purpose provided a less biased, and more complete understanding of the number and distribution of cases, than the use of other sources (for example specialty clinic, not-for-profit society). Changes in the prevalence of MND where the survival is short may also inform the effectiveness of interventions on survival.

The value of using of AHD for determining the prevalence of a rare condition is evident. First, identification of cases was done using a low cost and non-intrusive approach. There was no need to contact individuals, their care providers, specialty clinics, or hospitals directly. Routine AHD was generated through normal interactions with the health care system. Second, it was possible to assign cases geographically to their residence location regardless of where they sought or received care. This allowed for a more accurate representation of the geographic distribution of disease. Third, the health care system in Alberta is publicly funded and universally available so there were no financial impediments to accessing care, which works to increase the likelihood of identifying cases.

6.2.4 Publication 7

Svenson, L.W., Cwik, V.A., & Martin, W.R.W. (1999). The prevalence of motor neurone disease in the province of Alberta. *Canadian Journal of Neurological Sciences*, *26*, 119-122.

Link: http://www.ncbi.nlm.nih.gov/pubmed/10352871

6.3 Assessing Comorbidities –Motor Neurone Disease and Affective Disorders

6.3.1 Study Summary and Critique – Publication 8

A number of studies suggested that there might be an increased prevalence of affective disorders among individuals diagnosed with motor neurone disease (MND) (Hogg, *et al.*, 1994; Jau-Shin *et al.*, 2003; Kilani *et al.*, 2004; Moore *et al*., 1998). However, these studies generated conflicting results, primarily due to small sample sizes. None of these studies were population-based in nature so lacked the required statistical power to determine if MND was associated with an increased risk of an affective disorder.

Publication 8 (Patten *et al.,* 2007) was a cross-sectional study that examined the prevalence of affective disorders among individuals with MND. At the time the study was published, it was the first to use population-based administrative data to examine affective disorders among people with MND. Up to this point, studies had to rely on disease registries, disease-specific non-government organisations, specialty clinics, or hospital based ascertainment. Each of these options may be biased in how cases were ascertained leading to skewed estimations of prevalence and overall burden.

Data were extracted for six affective disorders for the entire population. Next, individuals with MND were identified allowing for estimates of affective disorder prevalence for those with and without MND. It also provided the necessary statistical power to provide stable estimates. Our analysis was able to demonstrate the existence of an association between affective disorders and MND.

Relative to other studies at the time, a key strength of this study was being population-based. This limited the potential for selection bias. It also ensured there was sufficient statistical power to assess the association between MND and affective disorders, a problem other studies had (Hogg, *et al.*, 1994; Jau-Shin *et al.*, 2003; Kilani *et al.*, 2004; Moore *et al*., 1998). The limited sample sizes in previous studies may be the reason for the inconsistent findings, resulting from an increased risk of a type II statistical error.

The population-based nature of AHD, combined with the high health care needs of MND patients worked to reduce issues other studies had related to selection bias or limited statistical power. Aside from demonstrating the value of administrative data for assessing comorbidities, particularly for rare conditions, the study also contributed meaningful results for clinicians. For example, we were able to show that the odds of being diagnosed with an affective disorder were greatest in the first year of illness, and then decreased with illness duration. We postulated that this might represent a process of adjustment whereby the psychological impact of MND decreases with time, and recommended additional studies examining the mental health care needs of individuals living with MND.

While this study advanced our understanding of the association between MND and affective disorders, there were limitations. MND is a group of neuromuscular conditions and each has a different progression. This introduced heterogeneity in the cases that may have had differences in survival probabilities, and this could have explained the weakening of the association between MND and affective disorders with increasing duration of illness. AHD is not well suited to clarify issues of this nature. The accuracy of the affective disorder diagnosis could not be confirmed through AHD. Also, it was not possible to differentiate major depressive episodes from bipolar disorders. This means the categorization of affective disorders includes a wide range of severity. Detailed clinical notes were not reviewed. A review would have helped to better understand the treatment the patient received and validate the findings.

The design allowed for a direct comparison with the entire population of the province, not something done in the other studies. It showed that MND increases the risk of being diagnosed with an affective disorder that is above what is seen in the general population. For clinicians, this provides an opportunity for earlier intervention, and to take into account the mental state of the patient when treating all aspects of MND. The presence of physician diagnosed affective disorders, even if inaccurate, represents a clinically meaningful event.

6.2.4 Publication 8

Patten, S.B., **Svenson, L.W.,** White, C.M., Khaled, S.M., & Metz, L.M. (2007). Affective disorders in motor neuron disease: a population-based study. *Neuroepidemiology*, *28*, 1-7.

Link: https://www.karger.com/Article/Abstract/97849

6.3 Summary

Using publications 7 (Svenson *et al*., 1999) and 8 (Patten *et al*., 2007) as examples, this chapter demonstrated the value of AHD for assessing the prevalence of rare events and for identifying the presence of comorbidities. When studying rare conditions, primary data collection techniques may lead to small sample sizes that limit statistical power or to biased samples that limit the ability to generalise the findings. In addition, the logistics and resources to find study subjects can be substantial. While there are limitations to the use of AHD, its population-based nature makes it a good starting point for descriptive epidemiological studies of rare conditions.

AHD allowed quantification of comorbidities associated conditions that have an impact on health system utilisation, costs, and complexity for managing the patient. Therefore, health care providers to be better prepared when consulting with patients, developing treatment and management plans, and allocating resources.

A better understanding of comorbidities helps to inform practice. The more clinicians know about all health needs of their patients, the better they can develop management plans. It is generally known that affective disorders impact on how a patient perceives their health, impacts their quality of life, and can negatively affect patient compliance with medication use and other strategies aimed at improving their health status (Cameron *et al*., 2014). The high rate of affective disorders, particularly in the year of initial diagnosis, suggests that clinicians must factor in mental health early in the management of this condition.

In addition, it provides researchers with the means of examining associations in a way that can inform resource allocations and inform front line practice. They also generate important hypotheses that could be addressed through primary data collection means, which would add the depth of understanding that is not present with AHD alone. AHD can, therefore, be used to drive research priorities and a better understanding for improving practice, and training of health professionals. The next chapter will explore the role of AHD for assessing risk and outcomes.

Chapter 7: Assessing Outcomes and Risk with Administrative Health Data

7.1 Introduction

There is general agreement that systematic reviews and meta-analyses of randomised control trials (RCTs) are trustworthy forms of evidence, followed by: RCTs, cohort studies, case-control studies, cross-sectional surveys, case reports, and finally perspectives or expert opinion (Evans, 2003). AHD represents an opportunity to evaluate the implementation of new treatments at a population level, and add information an RCT would have been underpowered to detect. This allows for a better understanding that may help explain gaps between the efficacy in an RCT and effectiveness in the community.

AHD allows for a population-level assessment of the impact of actions taken (for example policies, guidelines, and new interventions) by examining their use in the 'real-world'. For example, RCTs designed for vaccines generally include assessment of adverse events, but may be underpowered to detect rare, yet important adverse events. The linkage of multiple AHD sources, using a cohort design, can be done to quantify adverse events following the introduction of a new intervention, and thereby add information not available through an RCT. A population-based approach reduces issues with power to detect as well as the opportunity to control for a number of health conditions that may independently increase a particular adverse event (MacDonald *et al*., 2014). This will be explored in Publication 11.

Another aspect of risk that will be explored is related to the role of environmental and genetic factors in the onset of disease. Publication 9 will describe how the linkage of multiple AHD sources, using an ecological study design, was done to explore the role of parental ancestry and MS (Warren *et al*., 1996). The findings will be compared with Publication 10 that looked specifically at MS prevalence among First Nations people in Alberta, Canada, highlighting how using different approaches can lead to a better understand the epidemiology of MS (Svenson *et al.*, 2007).

7.2 An Example with Parental Ancestry and Multiple Sclerosis – Publication 9

7.2.1 Study Summary and Critique – Publication 9

The uneven geographic distribution of multiple sclerosis (MS) has been well established for decades and this knowledge has been used to create hypotheses suggesting a role for viral and/or environmental factors, along with genetic susceptibility, in the development of MS (Kurtzke, 1977). It had also been noted that MS appeared to be more common not only within certain geographic areas, but also among certain ethnic groups (Bullman & Ebers, 1992; Davenport, 1993; Page *et al*., 1993). Page *et al*. (1993) looked at the prevalence of MS across US states, and noted an association between MS and European ancestry, particularly those of Scandinavian descent. However, the authors stated they could not rule out the possibility of an ecological fallacy.

Publication 9 (Warren *et al*., 1996) was designed to examine the relationship between the geographic distribution of MS in the province of Alberta, Canada and self-reported parental ancestry, and to compare these findings with those from a case control study. This study had two embedded designs. The first was an ecological design, and the second was a case-control design.

The addition of the case control component was done to assess the potential for an ecological fallacy that could be present when using an ecological design. Prevalence rates were age-standardized, using the direct method (Fleiss, 1981) to remove any influence of differing age-distributions across the census divisions. The Canadian census data was then accessed and aggregated by sex, self-reported ancestry, and census division. This allowed for a comparison between an ecological study, and one based on primary data collection.

The ecological component of the study found results that were consistent with other research (Bulman & Ebers, 1992; Page *et al*., 1993) with Scandinavian ancestry appearing to be associated with a higher prevalence of MS for males. A negative correlation between MS prevalence and Aboriginal ancestry was noted.

The association found with the ecological component was strengthened by the inclusion of a case control study design. The combination of two designs into a single study made it possible to determine if an ecological fallacy may have been present in other studies. The use of AHD essentially added statistical power to the case-control component. It was possible that the case definition used for the ecological component may have introduced false positives, which would work to decrease the magnitude of the association. The ecological component suggested that Aboriginal ancestry was associated with lower prevalence of MS. AHD provided population-level data while the case-control study provided more depth and rich data to the process. It was possible to show that some of the variation in MS prevalence is attributed to ethnicity, and not simply geographic location.

A few key limitations need to be highlighted. The capture of MS cases was done based on a patient having at least health service for MS within a five-year period. This could have increased the number of false positive cases included in the analysis. The impact of this on the findings cannot be determined, but it is possible that association between ethnic origin and MS could be obscured. The case-control study was relatively small to provide a reliable estimate of the role of ethnic origin. As a result, countries of origin needed to be grouped. This introduced heterogeneity to the case group that would work to limit finding relationships between MS prevalence and ethnicity.

Despite the limitations, the study demonstrated that MS prevalence varies by ethnicity. This supports the role of genetics in the risk of MS, but does not allow an assessment of the nature of the genetic role. However, this adds support to justify additional aetiological research that includes a genetic component.

7.2.4 Publication 9

Warren, S., **Svenson, L.,** Woodhead, S., & Warren, K.G. (1996). Parental ancestry and risk of multiple sclerosis in Alberta, Canada. *Neuroepidemiology*, *15*, 1-9.

Link: https://www.karger.com/Article/Abstract/109883

7.3 First Nations and Multiple Sclerosis – Publication 10

7.3.1 Study Summary and Critique – Publication 10

In Publication 9, Warren *et al*. (1996) looked at parental ancestry and MS prevalence, and noted that geographic areas with a higher proportion of people of Aboriginal descent had lower prevalence rates of MS. Two Canadian studies had found MS prevalence to be lower among Aboriginal populations, but neither were population-based studies (Hader *et al*., 1985; Mirsattari *et al*., 2001). In Canada, there are three recognized Aboriginal populations – First Nations, Métis, and Inuit (Todd *et al*., 2001).

Publication 10 (Svenson *et al*., 2007) used a cross-sectional study design to assess the prevalence of MS among the First Nations population. The study found the prevalence of MS among First Nations people was approximately one third of that of the general population. However, when compared to international studies, the prevalence would still be considered high (Lauer, 1994). This supported a view that MS resulted from a complex interaction between genetically predisposed people and some factor, or factors, in the environment (Sadovnick & Ebers, 1993). It was the first population-based study in Canada to systematically look at the prevalence among an Aboriginal population. As such it provided a baseline or reference point from which other studies could compare.

A better understanding of the prevalence of MS in Aboriginal populations helps to understand where resources need to be allocated. Aboriginal populations in Alberta tend to live in more rural, or remote, areas of the province meaning access to specialists may be less than for those living in urban centres (Statistics Canada, 2012). Thus, data on prevalence can lead to better connections between family practitioners in rural areas and specialists in urban areas. From a research perspective, it may help to identify subsets of the MS population for trials on new therapies or management strategies.

It has been suggested that First Nations with MS have a more aggressive form of the disease, but AHD does not have information to allow for an assessment of disease severity or progression (Misattari *et al*., 2001). While AHD can provide evidence on disease prevalence, it cannot directly measure disease severity. Either proxies, based on health service use, or direct measurements would be needed. AHD, by itself, is not sufficient for determining if Aboriginal populations do have a more aggressive form of the disease.

Having an administrative process that captured information on First Nations status provided strength to the study. It removed the need to recruit people into the study, which may have been a challenge given the low frequency, and the high proportion of First Nations living in rural, or remote, areas of the province. The population-based nature of the study allowed for a direct comparison between First Nations and the general population.

While the validity of the diagnosis is a common limitation of studies using AHD, this study did use a case definition algorithm and had findings that were consistent with other studies (Warren & Warren, 1992; Warren & Warren, 1993; Klein *et al*., 1994). It is not known if First Nations with MS utilize health services differently than other MS cases. First Nations are more likely to live in rural and remote locations, and may not have the same access to health services, particularly specialty services. This could limit, or delay, the discovery of MS among First Nations leading to lower prevalence estimates. However, given the nature of MS, it is unlikely health service access would fully explain the difference in prevalence.

7.3.4 Publication 10

Svenson, L.W., Warren, S., Warren, K.G., Metz, L.M., Patten, S.B., & Schopflocher, D.P. (2007). Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, *34*, 175-180.

Link: http://dx.doi.org/10.1017/S0317167100006004

7.4 Assessing Vaccine-Associated Febrile Seizure Risk – Publication 11

7.4.1 Study Summary and Critique – Publication 11

Publication 11 (MacDonald *et al*., 2014) examined the risk of a child developing a febrile seizure following administration of measles-mumps-rubella-varicella (MMRV) vaccine relative to same-day administration of separate measles-mumps-rubella and varicella vaccines (MMR+V). The design was a retrospective population-based cohort covering the period 2006 to 2012. It estimated the relative risk for children receiving the MMRV from 2010 to 2012, and for children receiving MMR+V from 2006 onward via linkage of multiple databases.

Pre-licensure clinical trials had found a safety profile, for MMRV, similar to that found for MMR+V, except for an increased incidence of fever (Czajka *et al.*, 2009; GlaxoSmithKline, 2014). In 2010, Klein *et al*. reported an increased risk of a febrile seizure associated with the use of MMRV vaccine in the United States, when compared to the administration of MMR and varicella as separate vaccines. The U.S. study found there was one additional seizure event for every 2,300 doses of MMRV vaccine administered. The MMRV product used was in the U.S. (ProQuad®) was different than what was being used in Canada (Priorix-Tetra®) and some European countries. The lack of evidence on febrile seizure risk for Priorix-Tetra® was a catalyst for the current study.

Health insurance registry data was used for demographics, and to censor the data for anyone moving out of the province or dying. Immunisation data were extracted from a population-based immunisation repository which was used to determine which vaccines were administered and when. Data from hospitalisations, outpatient clinics, and physician visits were then used to define not only febrile seizures, but also comorbid conditions that may influence a seizure risk. The results were consistent with that of Klein *et al*. (2010), finding both an increased risk of a febrile seizure (one additional seizure for every 2,841 doses administered) and a similar risk window (7 – 10 days). At the time the paper was in press, Schink *et al*. (2014)

published a paper showing a febrile seizure risk with MMRV vaccine, used in Germany, of one additional seizure for every 2,747 doses administered.

This study was able to demonstrate that the linkage of multiple administrative health databases, combined with a longitudinal design could be used to augment passive surveillance programs for adverse events following immunisation. Also, it demonstrated the value of AHD as a novel way to examine risks that may not have been included in clinical trials. The clinical trials assessing the efficacy and safety of the vaccine were underpowered to detect the risk found in this study. The results contribute to the evidence base on vaccine safety, and provided front-line public health practitioners with reliable evidence that could be used in discussions with parents about the risks and benefits of vaccination. It also lays the groundwork for providing assurances that vaccine safety is well monitored.

It is possible that children who had a febrile seizure may not have been taken for medical care. Given the study was focused on children under two years of age, it is believed that a seizure would be sufficient to generate parental concern to prompt a visit. It was not possible in this study to examine medical charts as part of the analysis, but given seizure risk was evaluated pre- and post-vaccination, the relative risk measures should remain unbiased. In support of this, Klein *et al*., (2010) conducted a chart review as part of their assessment of febrile seizure risk and found a modest change in their relative risk measure. Their chart review showed their estimates were conservative which would under estimate the actual risk.

7.4.4 Publication 11

MacDonald, S.E., Dover, D.C., Simmonds, K.A., & **Svenson, L.W**. (2014). Risk of febrile seizures after the first dose of measles-mumps-rubella-varicella vaccine: a population-based cohort study. *Canadian Medical Association Journal*, *186*, 824-829.

Link: http://www.cmaj.ca/content/186/11/824.abstract

7.5 Summary

This chapter highlighted three publications that contributed to a better understanding of risk and outcomes. The first two examined the role of ethnicity as a risk factor for MS, while the third evaluated the risk associated with a new vaccine. While each had their limitations, they generated results that were consistent with other research.

The comparison of an ecological study with that of a case control study suggested the relationships found in other studies on ethnicity might have been attributed to an ecological fallacy. Despite this, it also demonstrated the value of ecological studies in better understanding the distribution of a health event, and possible explanations for the geographic distribution observed. In addition, the data can also be used to guide the planning of services for individuals living with MS. It allows clinicians to offer more patient-centred care, may contribute to conversations about familial risk, and may influence patient satisfaction, perceived benefits, and compliance with disease management services (Buchanan *et al*., 2010; *Kim et al*., 2014).

Publication 11 (MacDonald *et al*., 2014) was the first Canadian study to look at the risk of a febrile seizure associated with MMRV vaccine (Law *et al.*, 2014). RCTs designed for vaccines generally include assessment of adverse events, but may be underpowered to detect rare, but important adverse events. In this case the initial RCTs that demonstrated the efficacy of the vaccine were underpowered to detect this adverse event (GlaxoSmithKline Inc., 2014). This allowed for a relatively inexpensive and efficient way to assess the effectiveness of interventions, and determine if they should continue, or if further study is warranted (Gallagher, 2004).

A better understanding of risk and outcomes of new interventions can help to inform health policy development, research agendas, adverse event surveillance, and provide front line health professionals with information when advising patients and their family members. In the next chapter, the role AHD plays in the creation of health policy will be explored.

Chapter 8: Supporting Health Policy Development

8.1 Introduction

The application of research findings to 'real world' situations is the goal of many researchers. A general theme with the evolving use of AHD is to ensure its use supports the development of evidence-based policies and actions. Cooke *et al.* (2013) argued that AHD could have a significant impact on practice. They do caution that researchers must be careful in the data selected and approach to avoid developing biased results, but that a well-designed study can lead to valuable improvements in practice.

Once a policy decision has been made and implemented, it becomes important to determine if the policy had the intended outcome. Also, it becomes important to understand if there were any untoward outcomes. In this chapter, Publication 12 (Russell *et al*., 2014) will be used to as an example for assessing the impact of the introduction of a new vaccine for preventing chickenpox on the incidence of shingles, a condition that manifests from the same varicella virus that produces chickenpox.

8.2 Informing Health Policy – An Example with Varicella Vaccination

8.2.1 Study Summary and Critique – Publication 12

Varicella zoster virus is responsible for chickenpox in children and herpes zoster, also known as shingles, in adults. Varicella vaccines were licensed for use in Canada in 1998 with the province of Alberta introducing varicella vaccine, for the prevention of chickenpox, to the publicly funded immunisation program in 2001 (Russell *et al.*, 2005). Brisson *et al.* (2003) have suggested that the introduction of varicella vaccine programs for children may result in an increased incidence of shingles among older individuals due to a lack of immunological boosting that comes from exposure to varicella virus.

With the introduction of any public health intervention, it is important to understand any potential impact, whether positive or negative, on population health. It was unclear what the potential impact of childhood varicella immunisation would be on the incidence of shingles. Two studies, with results from Canada and the United

Kingdom, had observed increases in the incidence of shingles prior to the introduction of vaccine (Russell *et al.*, 2007; Brisson *et al.*, 2001).

To assess the impact of varicella vaccine on the incidence of shingles, data over a 10-year period were examined using a quasi-experimental design. While the approach was observational in nature, there was an intervention (vaccine introduction), and the intervention was not controlled, nor initiated by the researchers. Changes in the incidence of shingles were studied using three time periods – the period prior to the introduction of vaccine (1994-1998), the period when it was available privately (1999-2001), and the period when it was publicly funded (2002-2010).

Using Poisson regression models, the study found the incidence of shingles had increased over the entire study period, and the presence of one or more comorbidities was associated with an increased incidence of shingles. There was also a statistically significant decrease in the incidence rate among children less than 10 years of age.

An increasing rate of shingles among older individuals was a trend that began prior to the implementation of the immunisation program. As this appeared to be a continuation of the trend that pre-dated the introduction of the vaccine, it was not possible to assess the contribution of the immunisation program to this increasing trend. While it was not possible to fully assess the impact of the immunisation program on overall rates of shingles, we were able to show there were decreases in the incidence of shingles among children demonstrating the value of AHD for assessing outcomes related to healthcare interventions.

Alberta has a publicly funded, universally available health care system that is less prone to selection bias that may be seen from data that are specific to private insurance plans. The publicly funded aspect of the system reduces financial barriers from accessing services. The study also included comorbidities that may impact on the probability of developing shingles as a means of explaining variance seen in the statistical models. We also separated incidence shingles events from recurrent by using a 180-day buffer. Any shingles diagnosis received more than 180 days from the initial shingles diagnosis was considered a recurrent event and any less than 180 days was

considered part of the initial episode. This ensured a more accurate classification of incident and recurrent cases.

Limitations of this study primarily centred on the classification of cases and the potential for misclassification. Only a single record with shingles recorded was sufficient to be deemed a case. This could result in a higher number of false positive cases. Such a bias could skew results, and make it difficult to know if the trends observed were real, or an artefact of how cases were identified. However, Yawn *et al*. (2011) compared administrative health data and medical charts for shingles, and found that a single shingles code was correct approximately 85% of the time, and that AHD tended to underestimate the incidence of shingles. This means that any misclassification bias would be to err away from an increasing trend in shingles incidence.

With the introduction of any new vaccine awareness among health professionals usually increases, and this may impact on how physicians diagnose and report on cases of disease. This may also introduce a bias in reporting resulting in more cases being reported in the period after vaccine introduction than the period prior to introduction. It would be expected that this would bias towards increased reporting, but it is not possible to determine what impact, if any, this would have on the identification of cases. Any cases of shingles that did not seek medical care would be missed. This would bias results towards seeing no increase in shingles rates both before and after public funding.

8.3 Summary

This paper represented a novel approach to trying to assess the risk of a health condition increasing as the result of the introduction of an immunisation program designed to prevent another condition. This type of evidence is valuable to policy makers in making decisions to fund, or defund, a program as well as their ability to respond to elected officials and the general public. Front line practitioners, when speaking to risks associated with the vaccine, can also use the information. The program appears to have contributed to a decrease in shingles incidence among younger age groups, but it was not possible to fully assess the overall impact of the varicella immunisation program on the incidence of shingles. Despite not providing definitive results, the study demonstrated a novel way of monitoring policy decisions. It also implies that more than AHD is needed to address some questions. Simply put, AHD provides valuable information on what is happening, but may need to be augmented with other approaches to provide a more fulsome understanding.

8.4 Publication 12

Russell, M.L., Dover, D.C., Simmonds, K.A., & **Svenson, L.W**. (2014). Shingles in Alberta: Before and after publicly funded varicella vaccination. *Vaccine*, *32*, 6319-6324.

Link: http://www.sciencedirect.com/science/article/pii/S0264410X13012498

Chapter 9: Methodological Critique

9.1 Introduction

Chapters 3 through 8 provided a summary of each paper included in this thesis. This chapter will focus on the methodological aspects of each publication by grouping them together by study design. Strengths and limitations of the designs will be discussed. The emphasis will be on highlighting both strengths and weaknesses that were common across multiple papers.

9.2 Cross-Sectional Designs

Cross-sectional studies are a form of observational study that include either the entire population, or a representative sample of the population, with the goal of describing the population at a point in time (Rothman & Greenland, 1998). The purpose of a cross-sectional design is descriptive and usually there is no hypothesis being tested (Levin, 2006a). Cross-sectional studies typically collect exposure and outcome data at the same time, so the temporal relationship cannot be determined. However, they may provide clues for the development of hypotheses that could be tested with an experimental design (Rothman & Greenland, 1998; Levin, 2006a).

Publications 1 (Svenson, 1990), 2 (Svenson, 1991), 4 (Yiannakoulias *et al*., 2003), 7 (Svenson *et al*., 1999), 8 (Patten *et al*., 2007), and 10 (Svenson *et al*., 2007) were all population-based observational studies using a cross-sectional study design. While the approach for each study was appropriate to the given research questions, cross-sectional studies come with strengths and limitations that need to be considered when interpreting findings.

9.2.1 Cross Sectional Designs – Strengths

There are general advantages to cross-sectional designs that apply to each study presented. These include data access was inexpensive; data were population-

based allowing for estimations of prevalence; data could be used to support priority setting; data could be used to support health system planning; and development of aetiological hypotheses (Levin, 2006a). In addition to these advantages, the population-based nature of AHD reduces the risk of both selection and recall bias. Also, there was no need to contact patients or their family members to obtain the information.

Each study included strengths that were unique to the study. Publications 1 (Svenson, 1990) and 2 (Svenson, 1991) benefitted from the completeness of each data source, consistency in coding, accessibility of the data, and the data were continuously collected over time. For publications 4 (Yiannakoulias *et al*., 2003) and 8 (Patten *et al*., 2007), an additional strength was the ability to link unique individuals across each source of AHD. This allowed for a more thorough understanding of patient interactions with the health system and to attach attributes of interest to the individual. Publication 7 (Svenson *et al.*, 1999) demonstrated the value of AHD for capturing data on rare health events and Publication 10 (Svenson *et al*., 2007) showed the advantage of having additional demographic information allowing for the assessment of a health event among a particular subgroup in the population (First Nations). While not a specific strength of a cross-sectional design, these could be considered as strengths in that they were non-intrusive ways of assembling the data.

In addition to the above strengths unique to each study, there were general strengths that applied to each. The data were population based, reducing issues of selection bias. There is no need to contact individuals, which represented an efficient and cost-effective approach for collecting data. This approach was also free from issues of recall bias. The population-based level of the analysis resulted in large sample sizes that allowed for more precise point estimates, and greater confidence that the studies represented meaningful results.

9.2.2 Cross Sectional Designs – Weaknesses

There were general limitations that applied to Publications 1 (Svenson, 1990), 2 (Svenson, 1991), 4 (Yiannakoulias *et al*., 2003), 7 (Svenson *et al*., 1999), 8 (Patten *et al*., 2007), and 10 (Svenson *et al*., 2007). It was not possible to make any links with causal factors. Cross-sectional data are a description of the time period under investigation, and the situation may change if another time period is examined. However, the data can be used to generate hypotheses about causal relationships. For Publications 2, 4, 7, and 8 it was not possible to separate incident and prevalent cases of disease. For Publication 10, this limitation did not apply as a case definition algorithm was applied. This allowed incident and prevalent cases to be distinguished.

Other limitations were specific to each study more so than a limitation to the study design. For Publications 1 (Svenson, 1990) and 2 (Svenson, 1990) there was a lack of information on birthplace, places of residence, age of disease onset, severity of illness, and bias in vital event recording of underlying cause of death. Underreporting for the underlying cause of death may lead to under estimations of disease prevalence, and be biased towards the more severe presentations of disease. There may be other issues with mortality data that could compromise the interpretation of any findings. These include personal preference on where to die, which could impact the assessment of geographic distribution; not all deaths result in an autopsy, so important causes of death may go missed; and vital statistics agencies typically do not confirm the accuracy of the information submitted, meaning coding errors could go undetected (Nielsen *et al*., 1991; Schnatter *et al*., 1990). These limitations, if not considered, could lead to misinterpreting evidence on disease burden, which may in turn negatively impacts priority-setting and resource allocation.

It is also possible to over-estimate disease frequency should criteria for case identification be too lenient or inclusive. For studies of outcomes, the more stringent criteria that underestimates cases will provide meaningful results as each case included will have a higher probability of being a true case. If the study is focused on the overall

use of the health care system, then over or under estimation of disease frequency may not result in significant impacts, as they will represent meaningful information from a planning or resource allocation perspective. In any retrospective analysis of health system use, the accuracy of the diagnosis can be less important as it is the use of services that is the unit of analysis. While these represent important considerations when interpreting results, the strengths when combined with the consistency of results with other studies, demonstrate the advantages of using AHD outweigh the weaknesses.

9.3 Cohort Study Designs

Levin (2006b) defines a cohort, or longitudinal, study as one where a group of individuals is studied over a period of time. Longitudinal studies may be prospective or retrospective in nature. Cohort studies can be open with people moving in and out of the study or closed where a fixed population is followed with only exit from the cohort permitted. An advantage of this approach over cross-sectional designs includes being able to examine the temporal relationship between an event and an outcome. Also, cohort studies tend to have less potential for bias than other designs, but still may be vulnerable to issues of confounding. When primary data collection is done, cohort studies can be cost-prohibitive, as they generally need large sample sizes and long follow-up periods. This also means they are also not well suited for the study of rare events. Publications 3 (Robertson *et al*., 1998), 5 (Yiannakoulias *et al*., 2007), 6 (Yiannakoulias *et al*., 2004), and 11 (MacDonald *et al*., 2014) were all population-based retrospective cohort studies. The strengths and limitations of the cohort approach for these publications will be discussed in the following sections.

9.3.1 Cohort Design Strengths

Publications 3 (Robertson *et al*., 1998), 5 (Yiannakoulias *et al*., 2007), 6 (Yiannakoulias *et al*., 2004), and 11 (MacDonald *et al*., 2014) each used multiple AHD sources to assess outcomes. They also shared the following general strengths:

avoidance of recall bias, no need to contact the families of cases, large population-based sample size ensuring adequate statistical power, low cost compared with primary collection methods, and generalisability of the results due to the population-based nature of the data. The longitudinal design made it possible to look at temporal associations between the exposure under study and the outcome of interest. The use of AHD provided an efficient means of keeping track of demographic information such as address allowing for an accurate assessment of geography.

9.3.2 Cohort Design Limitations

There were limitations associated with this design and AHD that need to be considered when evaluating Publications 3 (Robertson *et al*., 1998), 5 (Yiannakoulias *et al*., 2007), 6 (Yiannakoulias *et al*., 2004), and 11 (MacDonald *et al*., 2014). First, the diagnostic information used was not validated. While this does not mean the data are invalid, it does mean that there may have been misclassification of cases. Also, it was not possible to know how complete the data were, and if there was a bias on health seeking behaviours that may have resulted in a diagnosis being missed or delayed. The criteria used by physicians for diagnosis is not captured within AHD making it difficult to discern those with a diagnosis and those being investigated for a given diagnosis.

9.4 Ecological Study Design

An ecological study is a form of observational study where the unit of observation is the group, and not the individual (Rothman & Greenland, 1998; Levin, 2006c). Levin (2006c) states that an ecological design is used when: there is a need to monitor population health; large-scale comparisons are to be made; the purpose is to assess population level risk factors and disease incidence; and when individual level measurements are not available. Publication 9 (Warren *et al*., 1996) used a geographical ecological design. It also compared results with a case-control study as a means of reducing potential confounding such as the risk of an ecological fallacy.

9.5 Quasi-Experimental Design

A quasi-experimental design is one that sits between an observational study and an experimental study. It can be considered a non-random intervention study with pre- and post-test components used to determine the impact of an intervention or program (Harris *et al.,* 2006; Polit & Hungler, 1987).

Publication 12 (Russell *et al*., 2014) was a form of quasi-experimental design as it looked at whether the implementation of a new vaccine (i.e., varicella vaccine) would have an impact on disease (i.e., herpes zoster) incidence. The strength of the approach was that it was population-based, and used multiple databases for case ascertainment. The study also included comorbidities that may impact on the probability of developing shingles as a means of explaining variance seen in the statistical models used. Quasi-experimental designs tend to offer good external validity as they are assessing interventions that have happened in the 'real world'.

With quasi-experimental designs, it is not possible to control as many factors as in an experimental design. Also, there was no control group as the intervention was population based. This limits the ability to make assertions about the causal nature of the data. However, the observational nature of the study does allow for an understanding of any potential impacts, whether positive or negative. The implementation of policies, or programs, provide a natural experiment whereby AHD can be used to identify whether the goals of the program were met and if any risk was created as an unintended consequence. This could lead to the development of hypotheses that may be tested using experimental designs.

9.6 Summary

Each design highlighted in the thesis has its own inherent strengths and weaknesses. This is not unique to AHD, but to any study. While limitations exist, the thesis has demonstrated that the strengths of AHD outweigh the limitations. Despite the limitations noted, there was consistency in findings with other studies that works

to add confidence to the value of AHD. AHD represent a valuable data source for understanding disease burden.

Each paper included in the thesis had consistency in their strengths. These included: the population-based nature of the data; no recall bias; cost-effective access to the data; data collection was non-intrusive; and the data were readily available. Other strengths were unique to specific studies included. AHD allows for long duration of follow-up without being resource prohibitive (see Publications 3, 5, 6, and 11) (MacDonald *et al*., 2014; Robertson *et al*., 1998; Yiannakoulias *et al*., 2004; Yiannakoulias *et al*., 2007). Another strength, linking multiple databases, was specific to all publications except Publications 1 and 2. These strengths of AHD combined with the use of appropriate research designs have demonstrated the value of AHD for quantifying burden of disease.

There were also limitations specific to the use of AHD highlighted in Chapter 2. While these general limitations also applied, it was possible to overcome some of the limitations. For example, Publication 9 used multiple designs and compared AHD with case-control data to address the potential for an ecological fallacy (Warren *et al*., 1996). Publication 11 linked multiple databases together to ensure the analysis could be adjusted for health events that might influence the outcome of interest (MacDonald *et al*., 2014). While limitations need to be considered, creative use of the data combined with good study design may help to mitigate perceived and actual risks to the validity of findings. Overall, the 12 papers have established that AHD can and do play a valuable role for understanding the burden of illness.

The next chapter will provide a summary, recommendations for future research, and concluding remarks.

Chapter 10: Summary, Recommendations, and Conclusions

10.1 Summary

The overall aim of the thesis was to demonstrate how AHD has evolved to become a valuable source of data for quantifying disease burden. Through a critical review of 12 publications, the evolution of the use of AHD has been shown over 25 years. Each paper was selected to demonstrate a concept as well as show a unique contribution to better understanding the use and value of AHD. The use has evolved from utilisation measures to descriptive epidemiological studies to more sophisticated study designs that allowed for the assessment of risk and outcomes. Initially, there were views that AHD was of limited value for understanding disease burden due to concerns over the integrity of the data. However, descriptive studies on the incidence and prevalence of chronic conditions have generated estimates and patterns consistent with studies using primary data collection strategies. They have also supported the development of aetiological hypotheses. This was demonstrated by the examination of mortality and hospitalisations attributed to Parkinson's disease (Svenson, 1990; Svenson, 1991). It was further advanced by looking at parental ancestry of multiple sclerosis by including AHD and case-control data to tease out the relationship of ancestry and MS risk while also addressing the possibility of an ecological fallacy (Warren *et al*., 1996). The consistency of findings between the two embedded designs worked to strengthen both approaches.

The application of case definition algorithms, with a person-oriented approach, has allowed for the creation of accurate estimates of disease burden. The estimates tend to err on the side of specificity when early concerns were that the error would lead to an increase in false positive cases. By adding a longitudinal component, it becomes possible to accumulate evidence at the person level that strengthens the likelihood that a case of a disease is truly a case. It also allows for a better understanding of outcomes. The concept that a case definition, with inclusion and exclusion criteria, resulted in reliable measures of burden was shown through the

examination of the birth cohort prevalence of cerebral palsy in Alberta (Robertson *et al.*, 1998). This was a novel approach at the time of the study.

In addition to the use of a longitudinal approach to assigning cases status, both geographical considerations and the concept of run-in periods were added as important considerations. Differences in the coding of stroke events between urban and rural health care facilities highlighted the value of geography and the availability of diagnostic technology (Yiannakoulias *et al*., 2003). This contributed to a better understanding of how the specificity of a diagnosis is influenced by where the patient is diagnosed. Patient movement following a diagnosis was shown to not have an impact on geography-based incidence and prevalence estimations (Yiannakoulias *et al.*, 2007). The concept of a run-in period was also a novel contribution as it aided in differentiating between incident and prevalent cases of disease (Yiannakoulias *et al*., 2007).

It has also been shown that AHD are valuable for the identification of people with rare conditions where primary data collection mechanisms are either too costly or are drawn from biased samples. AHD represents a cost-effective means of determining the burden of rare conditions. Assessing the prevalence of motor neurone disease, the first population-based study in Canada, was used to highlight the value of AHD for studies related to rare health conditions (Svenson *et al*., 1999).

AHD acts as an efficient source of evidence for understanding the presence of comorbidities. Given comorbidities represent complexities for the treatment and management of chronic conditions, understanding their frequency provides needed evidence for health policy development, resource allocation, and practice or service development. The value of AHD for assessing comorbidities was demonstrated by looking at the presence of affective disorders among individuals with MND (Patten *et al.*, 2007). This represented a unique approach in Canada at the time of the study.

AHD has been shown to add value for understanding the impact of policy decisions through an examination of 'real world' situations. The example provided looked at the potential negative impacts following the introduction of a new vaccine to

an immunisation program (MacDonald *et al*., 2014). AHD has been shown to be valuable to assess outcomes in ways that RCTs cannot. At the population level these outcomes can become significant. RCTs play an important and valuable role in understanding the efficacy of a new intervention relative to existing practice, while AHD plays an important and valuable role in understanding the population-level impacts, and a better understanding of the effectiveness of the new intervention. The thesis has shown that the use of AHD can augment the evidence from primary data collection studies.

The development of evidence in the health sector is created to inform and drive action. The analysis of AHD provides the necessary support to either directly or indirectly influence practice. For example, assessing the risk of a vaccine has both an indirect and direct impact on practice. At the indirect level, it informs health policies and at the practice level it provides health professionals with the information they need to decide on the best options for patients as well as providing an opportunity to inform and educate (MacDonald *et al*., 2014; Russell *et al*., 2014).

In addition to the above examples, the papers included in this thesis have also shown that AHD can be used in novel methodological ways to understand both disease burden and risk. For example, Yiannakoulias *et al*. (2004) placed physicians in both the role of a health care provider and as a patient. The feasibility of doing this study through primary data collection would have been limited and costly. Thus the population-based nature of the data as well as not having to interview study subjects allowed for an efficient way of quantifying and understanding both burden and risk.

While the thesis has highlighted how AHD can be a source of valuable evidence, it has also highlighted some important limitations associated with the use of AHD. These limitations can be used to develop recommendations for future research as a means of improving not only data quality, but also the value of any findings. With continued and increasing use, it is important to ensure the data can continue to support research in a meaningful way.

10.2 Recommendations for Research

There are a number of avenues for future research involving the use of AHD. Recommendations for future research will focus on AHD and not a specific disease or group of diseases. Regardless of the topic under study, researchers should start by understanding what AHD they have access to, how it was collected, the scope of the population included, and whether it is the right option for the study question.

The risk of having biased results, when using AHD, can be minimized by providing thorough descriptions of the data used, reporting on the accuracy of diagnostic of procedural codes used, distinguishing between statistical and clinical significance, understanding the time dependent nature of the data, and exploring the influence of clustering within the data on the study conclusions (van Walraven & Austin, 2012).

10.2.1 Validation Studies

An important aspect to understanding the value of any data source is to understand its validity and reliability (McPheeters *et al*, 2013; van Walraven *et al*., 2011). The introduction of errors may lead to one of two negative issues. First, it may lead to the adoption of evidence that is wrong as a result of not validating the data. Second, for studies where the findings are valid, there may be difficulty with adoption should people have concerns with validity.

Validation studies need to examine evidence drawn from multiple AHD sources and then determine the accuracy of the final assignment of disease status. Researchers need to carefully evaluate any algorithm in light of the data sources available to them. As part of the validation, there is a need to understand the impact of the deployment of new technologies as these may influence the diagnostic decision making process. Access to these technologies also needs to be assessed. Are there urban and rural differences in access or the criteria used for a diagnosis? Are there factors that create an inequity to accessing health services that may influence when and how a diagnosis

is provided? Are there issues with the completeness of case ascertainment that are related to administrative practices or data entry error?

10.2.2 Systematic Reviews

There is a need for systematic reviews of the evidence on the validity and reliability of AHD. This review should include the assessment of individual databases, performance of case definition algorithms, impact of population migration, immigration and emigration, availability and accessibility of health care professionals and facilities, completeness of coverage of AHD within different jurisdictions, population attributes, and administrative practices that guide the collection of the data. There are numerous factors, beyond the incidence, prevalence, or severity of disease, that influence health service use. This in turns has an influence on the collection, completeness, and reliability of findings based on AHD.

10.2.3 Evaluating New Technology and Data Sources

The healthcare system is complex and ever changing. New technologies can assist with providing a more accurate and timely diagnosis. The timing of the introduction of new technologies will have an influence on AHD-based studies. It will change the accuracy of the data over time, and algorithms used for case identification may not work across all time periods under study. As such, it is important to understand how to adapt approaches to ensure the integrity of the conclusions from any study.

Not all data collected within any given health system will be immediately available to researchers. When new data sources become available and can be linked to other data, this drives a need to examine the impact of the new data. In essence, it means continually refining approaches as new data becomes available. New data sources offer opportunities to improve the value of AHD and need to be systematically assessed.

There is a need to explore the methods for linking AHD with data collected from other study designs. Linking of data can help to fill gaps that one data source cannot address on its own. For example, linkage of RCT data to AHD would help to better understand gaps that occur between the efficacies found in the RCT, and the effectiveness in the community. This approach would provide a transition from the RCT to a natural setting. AHD offers population-based 'real world' perspectives that RCTs are unable to provide.

10.2.4 Leveraging New Methods and Emerging Trends

The emergence of the concept of 'big data' has led to new approaches to managing large structured and unstructured data. There is a need to determine how to apply the general approaches used for 'big data' and data visualisation to AHD-based research. Initially this needs to focus on the how to best link AHD to non-traditional sources of information, and to evaluate the value of such linkages.

Big data analytics in healthcare can be considered a growing and nascent field. There is need to develop methods for combining and mining AHD. While 'big data' may help to provide answers to clinical or policy questions more quickly, the data and methods need to be evaluated to ensure the rapid output is reliable and valid, and does not negatively impact decision-making. Research focused on new methods for data visualization will provide opportunities for hypothesis generation. There is a need to ensure the use of new technologies to go beyond descriptive or predictive analytics to the development of explanatory models that could lead to value insights that drive policy and practice.

10.2.5 Ensuring Generalisability

The context in which research is conducted plays an important role in the interpretation of findings. Any research on the validity or reliability of AHD needs to consider if the results can be generalised beyond the location of the study. Do validation studies in one country, or jurisdiction within a country, apply to others?

Under what conditions do the results generalise? What are the factors that can improve the generalisability of findings from one study to the next? These represent questions that help to ensure the reliable use of AHD. What works well in one context may not work in another. Understanding this will help reduce the risk of contributing poor evidence. Systematic reviews may help to answer some of these questions and should be explored. Also, validation studies should be conducted across countries or across administrative areas (provinces or states) within a country, and correlate results from various databases and approaches. The greater the consistency in findings across locations, the more confidence researchers can have in the results. The papers critiqued have shown that while the data for ten were from a single Canadian province, when compared to studies from other jurisdictions there were consistent findings. For example, estimates of birth cohort prevalence of cerebral palsy were consistent with results from international studies (Murphy *et al*., 1993; Robertson *et al.*, 1998; Stanley & Blair, 1991). The risk of febrile seizures following the introduction of MMRV vaccine were consistent, both in terms of magnitude and timing, as studies from the Unite States of America and Germany (Klein *et al*., 2010; MacDonald *et al*., 2014; Schink *et al*., 2014). This suggests that the findings of studies based in one location may generalise well to others.

10.3 Implications for Practice, Policy Development and Resource Allocation

The papers presented have implications for practice, policy development, and resource allocation. They have shown that AHD can be used to quantify the burden of disease, which contributes to setting priorities, and understanding health system burden. Also, an understanding of the distribution of disease helps to generate hypothesis about aetiological factors, which, in turn, contributes to our understanding of how to reduce the risk or burden of disease.

AHD is a valuable tool for identifying rare conditions, as information is often lacking to support the needs of policy makers and practitioners. Understanding the location of people with chronic conditions helps to ensure there are adequate

resources allocated to address population needs. This would include informing how many health professionals are needed, the type of training required, whether facilities are needed, and other supports.

The thesis has also shown that AHD can be used to augment other data sources and to evaluate interventions. Two examples were provided related to immunisation programs with both having policy and practice implications. By monitoring the impact of new programs and risks associated with new interventions, it becomes possible to develop good public health policy, and provide the public with assurance that systems are in place to protect health. From a practice standpoint the examples showed that AHD could quantify risk in a way that allows health professionals to speak honestly and with confidence about the safety of the intervention, what to watch for following the intervention, and time period to which the risk is greatest. This type of evidence allows health professionals to be well informed, and to have meaningful conversations with patients and colleagues.

10.4 Conclusions

By understanding both the strengths and limitations of AHD, we improve our ability to use these data effectively. AHD represent a rich source of data with great potential. The development of evidence using different methods, particularly when there are consistent results, works to strengthen our understanding of any given health issue. AHD have the advantage of covering nearly the entire population thus providing an opportunity to support stronger health system planning, policy development, and evaluation. Also, they lack many of the biases seen with other data sources making AHD a solid starting point for the development of hypotheses.

The thesis has highlighted the evolution of thinking and use of AHD as well as the application of novel methods. The 12 papers included in the thesis have demonstrated an evolution in the use of AHD and have clearly shown that AHD is valuable for quantifying the burden of illness. Each paper had results consistent with other studies that used either secondary or primary data collection. Also, there was an

increasing sophistication to the methods applied to AHD as well as the development of novel methods. These were used to increase the accuracy and, therefore, relevance of the findings.

While there may be risks that decisions could be made in error because the data used was of insufficient quality to appropriately address the question, this is not unique to AHD. Where AHD are of sufficient quality and timeliness, they represent a cost-effective and efficient means for answering important questions about the health of the population, the development of health policies, evaluation of interventions and outcomes, and the allocation of resources. The use of AHD will continue to grow, but to make the most of any research study researchers will need to understand and work within both the limits and strengths of the data. Researchers need to evaluate the data sources available to them, and critically evaluate the applicability to the study question.

The aim of this thesis was to highlight the evolution of thinking around the use of AHD and has shown that it can be used to accurately quantify the burden of illness. Strengths and weaknesses have been highlighted with the strengths outweighing the weaknesses. The 12 papers critiqued have shown that AHD can be used to provide reliable and accurate estimates of disease burden as well demonstrating value to understanding 'real world' applications of policy and practice. The data work to augment or compliment designs based on primary data collection. The evolution of the use of AHD has shown that robust measures of burden of illness could be developed and used. If this evolution had not occurred, it is possible that mortality data would have remained as the measure of choice for disease burden. As shown, mortality data are valuable, but tend to result in more biased estimates of disease burden.

It has been demonstrated that AHD plays an important role in providing evidence to support the process of developing policies, monitoring impact of decisions, setting priorities, allocating resources, and ultimately informing practice. It represents an important and valuable resource for understanding population health and the use of health services. It has been demonstrated that AHD is of sufficient quality and value

to play a central role in the quantification of the burden of illness. Its continued evolution will work to improve data quality and the utility of these data for understanding the burden of illness and to better inform actions taken to improve health and quality of life.

References

Adams, R.D., Victor, M., & Ropper, A.H. (1997). *Principles of Neurology* (6th ed.). New York: McGraw-Hill.

Aiello, A.E., Murray, G.F., Perez, P., Coulborn, R.M., Davis, B.M., Uddin, M., Shay, D.K., Waterman, S.H., & Monto, A.S. (2010). Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial. *The Journal of Infectious Diseases*, *201*, 491-498.

Alderson, M. (1976). A review of the National Health Service's computing policy in the 1970s. *British Journal of Preventive and Social Medicine*, *30*, 11-16.

Aoyagi, Y., Beck, C.R., Dingwall, R., & Nguyen-Van-Tam, J.S. (2015). Healthcare workers' willingness to work during an influenza pandemic: A systematic review and meta-analysis. *Influenza and Other Respiratory Viruses*, *9,* 120-130.

Baumgart, A. (1992). Evolution of the Canadian health care system. In A. Baumgart & J. Larsen (Eds.), *Canadian nursing faces the future* (pp. 22-43). Toronto: Mosby Year Book.

Benito-León, J., Louis, E.E., Villarejo-Galende, A., Romero, J.P., & Bermego-Pareja, F. (2014). Under reporting of Parkinson's disease on death certificates: A population-based study (NEDICES). *Journal of Neurological Sciences*, *347*, 188-192.

Bright, R.A., Avorn, J., & Everitt, D.E. (1989). Medicaid data as a resource for epidemiologic studies: Strengths and Limitations. *Journal of Clinical Epidemiology*, *42*, 937-945.

Brisson, M., Edmunds, W.J., & Gay, N.J. (2003). Varicella vaccination: impact of vaccine efficacy on the epidemiology of VZV. *Journal of Medical Virology*, *70 (Suppl. 1)*, S31-S37.

Brisson, M., Edmunds, W.J., Law, B., Gay, N.J., Walld, R., Brownell, M., Roos, L.L., & De Serres, G. (2001). Epidemiology of varicella zoster virus infection in Canada and the United Kingdom. *Epidemiology and Infection*, *127*, 3015-314.

Buchanan, R.J., Zuniga, M.A., Carrillo-Zuniga, G., Chakravorty, B.J., Tyry, T., Moreau, R.L., Huang, C., & Vollmer, T. (2010). Comparisons of Latinos, African Americans, and Caucasians with multiple sclerosis. *Ethnicity and Disease*, *20*, 451-457.

Bullman, D.E., & Ebers, G.C. (1992). The geography of MS reflects genetic susceptibility. *Journal of Tropical Geography and Neurology*, *2*, 66-72.

Butt, D.A., Tu, K., Young, J., Green, D., Wang, M., Ives, N., Jaakkimainen, L., Lam, R., & Guttman, M. (2014). A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. *Neuroepidemiology, 43*, 28-37.

Cameron, C., Habert, J., Anand, L., & Furtado, M. (2014). Optimizing the management of depression: primary care experience. *Psychiatric Research*, *220 (Suppl. 1)*, S45-S57.

Canadian Organization for Rare Disorders. (2015). *Now is the time: A strategy for rare diseases is a strategy for all Canadians*. Toronto: Author.

Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, *16*, 260-266.

Choi, B.C.K. (2012). The past, present, and future of public health surveillance. *Scientifica*, *2012*, 875253.

Chowdhury, T.T., & Hemmelgarn, B. (2015). Evidence-based decision-making 6: Utilization of administrative databases for health services research. In P.S. Parfrey & B.J. Brendan (Eds.), *Clinical Epidemiology: Practice and Methods, Methods in Molecular Biology* (pp. 469-484). New York, NY: Springer.

Clottey, C., Mo, F., LeBrun, B., Mickelson, P., Niles, J., & Robbins, G. (2001). The development of the national diabetes surveillance system (NDSS) in Canada. *Chronic Diseases in Canada*, *22*, 67-69.

Cooke, C.R., & Iwashyna, T.J. (2013). Using existing data to address important clinical questions in critical care. *Critical Care Medicine*, *41*, 886-896.

Cox, M., & Ellsworth, D. (1997, October). *Application-controlled demand paging for out-of-core visualization*. Paper presented at VIS97 IEEE Visualization ’97 conference, Phoenix, AZ.

Czajka, H., Schuster, V., Zepp, F., Espostio, S., Douha, M., & Willems, P. (2009.) A combined measles, mumps, rubella and varicella vaccine (Priorix-Tetra): immunogenicity and safety profile. *Vaccine*, *27*, 6501-65011.

Danila, O., Hirdes, J.P., Maxwell, C.J., Marrie, R.A., Patten, S., Pringsheim, T., & Jetté, N. (2014). Prevalence of neurological conditions across the continuum of care based on interRAI assessments. *BMC Health Services Research*, *14*, 29.

Devleesschauwer, B., Maertens de Noordhout, C., Smit, G.S.A., Duchateau, L., Dorny, P., Stein, C., Van Oyen, H., & Speybroeck, N. (2014). Quantifying burden of disease to support public health policy in Belgium: Opportunities and constraints. *BMC Public Health*, *14*, 1196.

Dixon, B.E., Rosenman, M., Xia, Y., & Grannis, S.J. (2013). A vision for the systematic monitoring and improvement of the quality of electronic health data. *Studies in Health Technology and Informatics*, *192*, 884-888.

Dutcher, J. (2014). *What is big data?* Retrieved from http://datascience.berkeley.edu/what-is-big-data/

Engström, S.G., Carlsson, L., Östgren, C.J., Nilsson, G.H., & Borgquist, L.A. (2006). The importance of comorbidity in analysing patient costs in Swedish primary care. *BMC Public Health*, *6*, 36.

European Commission (n.d.). *Useful information on rare diseases from an EU perspective*. Retrieved from http://ec.europa.eu/health/ph_information/documents/ev20040705_rd05_en.pdf

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*, 77-84.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley & Sons.

Gallagher, R.M. (2004). Naturalistic study designs in samples with refractory pain: advantages and limitations. *Pain Medicine*, 5, 143-145.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *10*, 3-8.

GlaxoSmithKline Inc. (2014). *Product monograph: Priorix-Tetra®: Combined measles, mumps, rubella and varicella vaccine, live, attenuated*. Mississauga, ON: Author.

Government of Alberta. (2014). *Alberta's pandemic influenza plan*. Edmonton: Author.

Hader, W.J., Feasby, T.E., Rice, G.P.A., & Ebers, G.C. (1985). Multiple sclerosis in Canadian native people. *Neurology*, *35 (Suppl. 1)*, 300.

Harris, A.D., McGregor, J.C., Perencevich, E.N., Furuno, J.P., Zhu, J., Peterson, D.E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in health informatics. *Journal of the American Medical Informatics Association*, *13*, 16-23.

Hertzman, C., Weins, M., Bowering, D., Snow, B., & Calne, D. (1990). Parkinson's disease: A case-control study of occupational and environmental risk factors. *American Journal of Industrial Medicine*, *17*, 349-355.

Hogg, K.E., Goldstein, L.H., & Leigh, P.N. (1994). The psychological impact of motor neurone disease. *Psychological Medicine*, *24*, 625-632.

Hudson, A.J., Davenport, A., and Hader, W. (1986). The incidence of amyotrophic lateral sclerosis in southwestern Ontario, Canada. *Neurology*, *36*, 1525-1528.

James, R.C., Blanchard, J.F., Campbell, D., Clottey, C., Osei, W., Svenson, L.W., & Noseworthy, T.W. (2004). A model for non-communicable disease surveillance in Canada: The prairie pilot diabetes surveillance system. *Chronic Diseases in Canada*, *25*, 7-12.

Jau-Shin, L., Reeves, A., Benice, T., & Sexton, G. (2003). Fatigue and depression are associated with poor quality of life in ALS. *Neurology*, *60*, 122-123.

Jutte, D.P., Roos, L.L., & Brownell, M.D. (2011). Administrative record linkage as a tool for public health research. *Annual Review of Public Health*, *32*, 91-108.

Kasowski, E.J., Garten, R.J., & Bridges, C.B. (2011). Influenza pandemic epidemiologic and virologic diversity: Reminding ourselves of the possibilities. *Clinical Infectious Diseases*, *52 (Suppl. 1), S44-S49.*

Kessler, I.I. (1972). Epidemiologic studies of Parkinson's disase II. A hospital-based survey. *American Journal of Epidemiology*, *95*, 308-318.

Kilani, M., Micallef, J., Soubrouillard, C., Rey-Lardiller, D., Demattei, C., Dib, M., Phillippot, P., Ceccaldi, M., Pouget, J., & Blin, O. (2004). A longitudinal study of the evolution of cognitive function and affective state in patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, *5*, 46-54.

Kim, G., Parton, J.M., Ford, K.L., Bryant, A.N., Shim, R.S., & Parmelee, P. (2014). Geographic and racial-ethnic differences in satisfaction with and perceived benefits of mental health services. *Psychiatric Services*, *65*, 1774-1482.

Klein, G., Rose, M.S., & Seland, T.P. (1994). The prevalence of multiple sclerosis in the Crowsnest Pass region of southern Alberta. *Canadian Journal of Neurological Sciences*, *21*, 262-265.

Klein, N.P., Fireman, B., Yih, W.K., Lewis, E., Kulldorff, M., Ray, P., Baxter, R., Hambridge, S., Nordin, J., Naleway, A., Belongia, E.A., Lieu, T., Baggs, J., Weintraub, E., & Vaccine Safety Datalink. (2010). Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics*, *126*, e1-e8.

Koran, L.M. (1975). The reliability of clinical methods, data and judgements. *New England Journal of Medicine*, *239*, 642-646.

Kreyenfeld, M., & Wilekens, F. (2015). Data bases and statistical systems: Demography. *International Encyclopedia of the Social and Behavioral Sciences*, *5*, 735-741.

Kurtzke, J.F. (1977). Geography in multiple sclerosis. *Journal of Neurology*, *215*, 1-26.

Kurtzke, J.F., & Goldberg, I.D. (1988). Parkinsonism death rates by race, sex, and geography. *Neurology*, *38*, 1558-1561.

Kurtzke, J.F. ,& Murphy, F.M. (1990). The changing patterns of death rates in Parkinsonism. *Neurology*, *40*, 42-49.

Lacasse, Y., Daigle, J.M., Martin, S., & Maltais, F. (2012). Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. *Canadian Respiratory Journal*, *19*, 113-120.

Lauer, K. (1994). The risk of multiple sclerosis in the U.S.A. in relation to sociogeographic features: A factor analytic study. *Journal of Clinical Epidemiology*, *47*, 43-48.

Last, J.M. (2007). *A dictionary of public health*. New York: Oxford University Press, Inc.

Law, B.J., Laflèche, J., Ahmadipour, N., & Anyoti, H. (2014). Canadian adverse events following immunization surveillance system (CAEFISS): Annual report for vaccines administered in 2012. *Canada Communicable Diseases Report*, *40*, 7-23.

Lee, J.R., Annegers, J.F., & Appel, S.H. (1995). Prognosis of amyotrophic lateral sclerosis and the effect of referral selection. *Journal of the Neurological Sciences*, *132*, 207-215.

Levin, K.A. (2006a). Study design III: Cross-sectional studies. *Evidence Based Dentistry*, 7*,* 24-25.

Levin, K.A. (2006b). Study design IV: Cohort studies. *Evidence Based Dentistry*, *7,* 51-52.

Levin, K.A. (2006c). Study design VI: Ecological studies. *Evidence Based Dentistry*, 7*,* 108.

Liu, L., Reeder, B., Shuaib, A., & Mazagri, R. (1999). Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: Implications for stroke surveillance. *Cerebrovascular Diseases*, *9*, 224-230.

Lix, L.M., Yogendran, M.S., Leslie, W.D., Shaw, S.Y., Baumgartner, R., Bowan, C., Metge, C., Gumel, A., Hux, J., & James, R.C. (2008). Using multiple data features improve

the validity of osteoporosis case ascertainment from administrative databases. *Journal of Clinical Epidemiology, 61*, 1250-1260.

Lix, L.M., Yogendran, M.S., Shaw, S.Y., Burchill, C., Metge, C., & Bond, R. (2008). Population-based data sources for chronic disease surveillance. *Chronic Diseases in Canada*, *29*, 31-38.

MacDonald, S.E., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Risk of febrile seizures after the first dose of measles-mumps-rubella-varicella vaccine: a population-based cohort study. *Canadian Medical Association Journal*, *186*, 824-829.

Mackie, P., Sim, F., & Johnman, C. (2015). Big data! Big deal? *Public Health*, *129*, 189-190.

Manfreda, J., Becker, A.B., Wang, P.Z., Roos, L.L., & Anthonisen, N.R. (1993). Trends in physician-diagnosed asthma prevalence in Manitoba between 1980 and 1990. *Chest*, *103*, 151-157.

Marrie, R.A., Yu, B.N., Leung, S., Elliot, L., Caetano, P., Warren, S., Patten, S.B., Svenson, L.W., Tremlett, H., Fisk, J., Blanchard, J.F., & CIHR Team in the Epidemiology and Impact of Comorbidity on Multiple Sclerosis. (2013). The utility of administrative data for surveillance of comorbidity in multiple sclerosis: A validation study. *Neuroepidemiology*, *40*, 85-92.

McPheeters, M.L., Sathe, N.A., Jerome, R.N., & Carnahan, R.M. (2013). Methods for systematic reviews of administrative database studies capturing health outcomes of interest. *Vaccine*, *315*, K2-K6.

Misattari, S.M., Johnston, J.B., McKenna, R., Del Bigio, M.R., Orr, P., Ross, R.T., & Power, C. (2001). Aboriginals with multiple sclerosis: HLA types and predominance of neuromyelitis opitca. *Neurology*, *56*, 317-323.

Moore, M.J., Moore, P.B., & Shaw, P.J. (1998). Mood disturbances in motor neurone disease. *Journal of Neurological Sciences*, *160 (Suppl. 1)*, S53-S56.

Murphy, C.C., Yeargin-Allsopp, M., Decoufle, P., & Drews, C.D. (1993). Prevalence of cerebral plays among ten-year-old children in metropolitan Atlanta, 1985 through 1987. *Journal of Pediatrics*, *123*, 230-237.

Murray, C.J.L. (1994). Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the World Health Organization*, *72*, 429-445.

Murray, T.J., Pride, S., & Haley, G. (1974). Motor neuron disease in Nova Scotia. *Canadian Medical Association Journal*, *110*, 814-817.

Nielsen, G.P., Björnsson, J., & Jonasson, J.G. (1991). The accuracy of death certificates: Implications for health statistics. *Virchows Archiv A Pathological Anatomy and Histopathology*, *419*, 143-146.

Nova Scotia-Saskatchewan Cardiovascular Disease Epidemiology Group (1989). Estimation of the incidence of acute myocardial infarction using record linkage: A feasibility study in Nova Scotia and Saskatchewan. *Canadian Journal of Public Health*, *80*, 412-417.

Odelin, M.F., Pozzetto, B., Aymard, M., Defayolle, M., & Jolly-Million, J. (1993). Role of influenza vaccination in the elderly during an epidemic of A/H1N1 virus in 1988-1989: Clinical and serological data. *Gerontology*, *39*, 109-116.

Page, W.F., Kurtzke, J.F., Murphy, F.M., & Norman, J.E., Jr. (1993). Epidemiology of multiple sclerosis in US veterans. V. Ancestry and the risk of multiple sclerosis. *Annals of Neurology*, *33*, 632-639.

Patten, S.B. (2015). *Epidemiology for Canadian students: Principles, methods and critical appraisal*. Edmonton: Brush Education Inc.

Patten, S.B., Svenson, L.W., White, C.M., Khaled, S.M., & Metz, L.M. (2007). Affective disorders in motor neuron disease: a population-based study. *Neuroepidemiology*, *28*, 1-7.

Paulson, G.W., & Gill, W.M. (1995). Are death certificates reliable to estimate the incidence of Parkinson's disease? *Movement Disorders*, *10*, 678.

Pharoah, P.O.D., Cooke, R.W.I., & Rosenbloom, L. (1990). Birthweight specific trends in cerebral palsy. *Archives of Diseases in Childhood*, *65*, 602-606.

Pinheiro, P., Plaß, D., & Krämer, A. (2011). The burden of disease approach for measuring population health. In A. Krämer, M. Khan, & F. Kraas (Eds.), *Health in megacities and urban areas*. Berlin: Springer-Verlag.

Polit, D.F., & Hungler, B.P. (1987). *Nursing research: Principles and methods, 3rd ed.* Philadelphia: J.B. Lippincott Company.

Potvin, L., & Champagne, F. (1986). Utilization of administrative files in health research. *Social Indicators Research*, *18*, 409-423.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*, 3.

Rajput, A.H. (1992). Frequency and cause of Parkinson's disease. *Canadian Journal of Neurological Sciences*, *19 (Suppl. 1)*, 103-107.

Ray, W.A. (1997). Policy and program analysis using administrative databases. *Annals of Internal Medicine*, *127*, 712-718.

Reker, D.M., Hamilton, B.B., Duncan, P.W., Shu-Chuan, J.Y., & Rosen, A. (2001). Stroke: Who's counting what? *Journal of Rehabilitation Research and Development*, *38*, 281-289.

Robertson, C., Sauve, R., & Christianson, H. (1994). Province-based study of neurologic disability among survivors weighing 500 through 1249 grams at birth. *Pediatrics*, *93*, 636-640.

Robertson, C.M.T., Svenson, L.W., & Joffres, M.R. (1998). Prevalence of cerebral palsy in Alberta. *Canadian Journal of Neurological Sciences*, *25*, 117-122.

Roos, L.L., Roos, N.P., Cageorge, S.M., & Nicol, J.P. (1982). How good are the data? Reliability of one health care data bank. *Medical Care*, *20*, 266-276.

Roos, N.P. (1989). Using administrative data from Manitoba, Canada to study treatment outcomes: Developing control groups and adjusting for case severity. *Social Science and Medicine*, *28*, 109-113.

Rosenfeld, L.S., Goldman, F., & Kaprio, L.A. (1957). Reasons for prolonged hospital stay: A study of need for hospital care. *Journal of Chronic Disease*, *6*, 141-152.

Rothman, K.J. (1996). Lessons from John Graunt. *Lancet*, *347*, 37-39.

Rothman, K.J., & Greenland, S. (1998). Types of epidemiologic studies. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd Ed.) (pp. 67-78). Philadelphia: Lippincott-Raven Publishers.

Russell, M.L., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Shingles in Alberta: Before and after publicly funded varicella vaccination. *Vaccine*, *32*, 6319-6324.

Russell, M.L., Schopflocher, D.P., Svenson, L.W., Yiannakoulias, N., Grimsrud, K., & Virani, S.N. (2005). The changing epidemiology of chickenpox in Alberta. *Vaccine*, *23*, 5398-53403.

Russell, M.L., Schopflocher, D.P., Svenson, L., Virani, S.N. (2007). Secular trends in the epidemiology of shingles in Alberta. *Epidemiology and Infection*, *135*, 908-913.

Sadovnick, A.D., & Ebers, G.C. (1993). Epidemiology of multiple sclerosis: A critical overview. *Canadian Journal of Neurological Sciences*, *20*, 17-29.

Schnatter, A.R., Acquavella, J.F., Thompson, F.S., Donaleski, D., & Thériault, G. (1990). An analysis of death ascertainment and follow-up through Statistics Canada's mortality data base system. *Canadian Journal of Public Health*, *81*, 60-65.

Schink, T., Holstiege, J., Kowalzik, F., Zepp, F., & Garbe, E. (2014). Risk of febrile convulsions after MMRV vaccination in comparison to MMR or MMR+V vaccination. *Vaccine*, *32*, 645-650.

Shah, C.P. (1998). Evolution of national health insurance. In C.P. Shah (Ed.), *Public health and preventive medicine in Canada* (4th ed.) (pp. 283-313). Toronto: University of Toronto Press.

Shwartz, M., Iezzoni, L.I., Moskowitz, M.A., Ash, A.A., & Sawitz, E. (1996). The importance of comorbidities in explaining patient costs. *Medical Care*, *34*, 767-782.

Southern, D.A., Roberts, B., Edwards, A., Dean, S., Norton, P., Svenson, L.W., Larsen, E., Sargious, P., Lau, D.C., & Ghali, W.A. (2010). Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. *Canadian Journal of Public Health*, *101*, 61-64.

Stanley, F.J. (1994). Cerebral palsy trends – implications for perinatal care. *Acta Obstetricia et Gynecologica Scandinavica*, *73, 5-9.*

Stanley, F.J., & Blair, E. (1991). Why have we failed to reduce the frequency of cerebral palsy? *Medical Journal of Australia*, *154*, 623-626.

Statistics Canada (1999). *National and provincial highlights from the 1996 census of agriculture-population linkage base*. Ottawa: Author.

Statistics Canada. (2012). Aboriginal peoples. In Statistics Canada (Ed.), *Canada year book, 2012* (pp. 1-13). Ottawa: Author.

Stewart, D.W. & Kamins, M.A. (1993). *Secondary research: Information sources and methods* (2nd ed.). Newbury Park, U.S.A.: Sage Publications, Inc.

Sullivan, D.F. (1966). Conceptual problems in developing an index of health. *Vital and Health Statistics 2*, *17*, 1-18.

Svenson, L.W. (1990). Geographic distribution of deaths due to Parkinson's disease in Canada: 1979 – 1986. *Movement Disorders*, *5*, 322-324.

Svenson, L.W. (1991). Regional disparities in the annual prevalence rates of Parkinson's disease in Canada. *Neuroepidemiology*, *10*, 205-210.

Svenson, L.W., Cwik, V.A., & Martin, W.R.W. (1999). The prevalence of motor neurone disease in the province of Alberta. *Canadian Journal of Neurological Sciences*, *26*, 119-122.

Svenson, L.W., Platt, G.H., & Woodhead, S.E. (1993). Geographic variations in the prevalence rates of Parkinson's disease in Alberta. *Canadian Journal of Neurological Sciences*, *20*, 307-311.

Svenson, L.W., Warren, S., Warren, K.G., Metz, L.M., Patten, S.B., & Schopflocher, D.P. (2007). Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, *34*, 175-180.

Svenson, L.W., Woodhead, S.E., & Platt, G.H. (1994). Regional variations in the prevalence rates of multiple sclerosis in the province of Alberta, Canada. *Neuroepidemiology*, *13*, 8-13.

Tirschwell, D.L., & Longstreth, W.T. (2002). Validating administrative data for stroke research. *Stroke*, *33*, 2465-2470.

To, T., Dell, S., Dick, P.T., Cicutto, L., Harris, J.K., MacLusky, I.B., & Tassoudji, M. (2006). Case verification of children with asthma in Ontario. *Pediatric Allergy and Immunology*, *17*, 69-76.

Todd, R., Thornton, M., & Collins, D.N. (2001). *Aboriginal people and other Canadians: shaping new relationships*. Ottawa: University of Ottawa Press.

Tuky, J.W. (1977). *Exploratory data analysis*. Boston: Addison-Wesley.

van Walraven, C., & Autsin, P. (2012). Administrative database research has unique characteristics that can risk biased results. *Journal of Clinical Epidemiology*, *65*, 126-131.

van Walraven, C., Bennett, C., & Forster, A.J. (2011). Administrative database research infrequently used validated diagnostic or procedural codes. *Journal of Clinical Epidemiology*, *64*, 1054-1059.

Virnig, B.A., & McBean, M. (2001). Administrative health data for public health surveillance and planning. *Annual Review of Public Health*, *22*, 213-230.

Warren, S., Svenson, L., Woodhead, S., & Warren, K.G. (1996). Parental ancestry and risk of multiple sclerosis in Alberta, Canada. *Neuroepidemiology*, *15*, 1-9.

Warren, S. & Warren, K.G. (1992). Prevalence of multiple sclerosis in Barrhead County, Alberta, Canada. *Canadian Journal of Neurological Sciences*, *19*, 72-75.

Warren, S. & Warren, K.G. (1993). Prevalence, incidence, and characteristics of multiple sclerosis in Westlock County, Alberta, Canada. *Neurology*, *43*, 1760-1763.

Warren, S., Warren, K.G., Svenson, L.W., Schopflocher, D.P., and Jones, A. (2005). Mortality rates for multiple sclerosis (MS) worldwide: Geographic and temporal trends (pp. 1 – 26). In F. Columbus (ed.), *Progress in Multiple Sclerosis Research*. Hauppauge, New York: Nova Science Publishers, Inc.

Wennberg, J., & Gittelshohn, A. (1973). Small area variations in health care delivery a population-based health information system can guide planning and regulatory decision-making. *Science*, *182*, 1102-1108.

Wilchesky, M., Tamblyn, R.M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. *Journal of Clinical Epidemiology*, *57*, 131-141.

Wilde, J.A., McMillan, J.A., & Serwint, J. (1999). Effectiveness of influenza vaccine in health care professionals – a randomized trial. *Journal of the American Medical Association*, *281*, 908-913.

Wilson, N., Baker, M., Crampton, P., & Mansoor, O. (2005). The potential impact of the next influenza pandemic on a national primary care medial workforce. *Human Resources for Health*, *3*, 7.

Yawn, B.P., Saddler, P., Wollman, P.C., St. Sauver, J.L., Kurland, M.J., & Sy, L.S. (2007). A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. *Mayo Clinic Proceedings*, *82*, 1341-1349.

Yiannakoulias, N., Russell, M.L., Svenson, L.W., & Schopflocher, D.P. (2004). Doctors, patients and influenza-like illness: clinicians or patients at risk? *Public Health*, *118*, 527-531.

Yiannakoulias, N., Schopflocher, D.P., Warren, S.A., & **Svenson, L.W.** (2007). Parkinson's disease, multiple sclerosis and changes in residence in Alberta. *Canadian Journal of Neurological Sciences*, *34*, 343-348.

Yiannakoulias, N., Svenson, L.W., Hill, M.D., Schopflocher, D.P., James, R.C., Wielgosz, A.T., & Noseworthy, T.W. (2003). Regional comparisons of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chronic Diseases in Canada*, *24*, 9-16.

Young, T.K, Roos, N.P., & Hammerstrand, K.M. (1991). Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: A pilot study. *Canadian Medical Association Journal*, *144*, 318-324.

Yurkovich, M., Avina-Zubieta, J.A., Thomas, J., Gorenchtein, M., & Lacaille, D. (2015). A systematic review identifies valid comorbidity indices derived from administrative health data. *Journal of Clinical Epidemiology*, *68*, 3-14.

Appendices

Appendix A: Peer Reviewed Publication List

Papers included in this thesis have been marked with an asterisk. At the time of admission to the program, there were 139 publications spanning the period 1990 through to 2014. As of July 1, 2015, the total number of peer reviewed professional publications stands at 144 spanning the health, education, social sciences, and geography literature.

A.1 Peer Reviewed Publications

Svenson, L.W. and Varnhagen, C.K. (1990). Knowledge, attitudes and behaviours related to AIDS among first year university students. *Canadian Journal of Public Health, 81*, 139-140.

*__Svenson, L.W.__ (1990). The geographic distribution of deaths due to Parkinson's disease in Canada: 1979-1986. *Movement Disorders*, *5*, 322-324.

Svenson, L.W. (1990). Mental health services in Edmonton: An assessment of service availability. *Canadian Journal of Public Health, 81*, 394-395.

Svenson, L.W. (1990). A model of co-operation: Community schools and public health officials. *Community Education Journal*, *18,* 16-17.

Svenson, L.W., Varnhagen, C.K., Godin, A.M., and Salmon, T. (1991). Installing condom machines in high schools. *The Canadian School Executive*, *19(8)*, 13-16.

Varnhagen, C.K., **Svenson, L.W.**, Godin, A.M., Johnson, L., and Salmon, T. (1991). Sexually transmitted diseases and condoms: High school students' knowledge, attitudes and behaviours. *Canadian Journal of Public Health*, *82,* 129-132.

Archer, W., van Roodselaar, C., and **Svenson, L.** (1991). The role of further education councils in distance education. *Journal of the Alberta Association for Continuing Education, 19*, 67-80.

*__Svenson, L.W.__ (1991). Regional disparities in the annual prevalence rates of Parkinson's disease in Canada. *Neuroepidemiology*, *10*, 205-210.

Campbell, R.L. and **Svenson, L.W.** (1992). Drug use among university undergraduate students. *Psychological Reports, 70*, 1039-1042.

Svenson, L.W. and Campbell, R.L. (1992). Perceived health status and desired health information needs of university students. *Canadian Journal of Public Health, 83*, 167-168.

Svenson, L.W., Woodhead, S.E., and Platt, G.H. (1992). Preliminary results of a review of hysterectomy in Alberta, 1984 to 1989. *Chronic Diseases in Canada, 13*, 37-41.

Campbell, R.L., **Svenson, L.W.**, Jarvis, G.K. (1992). Perceived level of stress among university undergraduate students in Edmonton, Canada. *Perceptual and Motor Skills, 75*, 552-554.

Svenson, L.W., Varnhagen, C.K., Godin, A.M., and Salmon, T.L. (1992). Rural high school students' knowledge, attitudes and behaviours concerning sexually transmitted diseases. *Canadian Journal of Public Health*, *83*, 260-263.

Svenson, L.W. (1992). Breast self-examination behaviours of female university students. *Canadian Journal of Public Health*, *83*, 388-389.

Campbell, R.L., **Svenson, L.W.**, and Jarvis, G.K. (1993). Age, gender and location as factors in permission to smoke among university students. *Psychological Reports, 72*, 1231-1234.

Svenson, L.W., Woodhead, S.E., and Platt, G.H. (1993). Estimating the prevalence of asthma in Alberta: A study using provincial health care records. *Chronic Diseases in Canada, 14,* 28-33.

Platt, G.H., **Svenson, L.W.**, and Woodhead, S.E. (1993). Coronary artery bypass grafting in Alberta from 1984 to 1989. *Canadian Journal of Cardiology, 9*, 621-624.

Svenson, L.W., Platt, G.H., and Woodhead, S.E. (1993). Geographic variations in the prevalence rates of Parkinson's disease in Alberta. *Canadian Journal of Neurological Sciences, 20*, 307-311.

Platt, G.H., **Svenson, L.W.**, and Woodhead, S.E. (1994). Geographic distribution in the performance of hysterectomy in the province of Alberta from 1984 to 1989. *Journal of the Society of Obstetricians and Gynaecologists of Canada, 16*, 1497-1501.

Svenson, L.W., Woodhead, S.E., Platt, G.H. (1994). Regional variations in the

prevalence rates of multiple sclerosis in the province of Alberta, Canada. *Neuroepidemiology, 13*, 8-13.

Svenson, L.W., Jarvis, G.K., and Campbell, R.L. (1994). Age and gender differences in the drinking behaviours of university students. *Psychological Reports, 75,* 395-402.

Huang, J., **Svenson, L.W.**, and Lindsay, J. (1994). Monograph series on aging-related diseases: V. Alzheimer's disease (senile dementia of the Alzheimer type). *Chronic Diseases in Canada*, *15*, 59-76.

Holden, R.W., **Svenson, L.W.**, Jarvis, G.K., Campbell, R.L., Lagace, D.R., and Backs, B.J. (1995). Drinking behaviors of Canadian correctional officers. *Psychological Reports*, *76*, 651-655.

Svenson, L.W., Jarvis, G.K., Campbell, R.L., Holden, R.W., Backs, B.J, and Lagace, D.R. (1995). Past and current drug use among Canadian correctional officers. *Psychological Reports, 76,* 977-978.

Svenson, L.W., Forster, D.I., Woodhead, S.E., and Platt, G.H. (1995). Health care utilization by individuals with a chemically dependent family member. *Canadian Family Physician*, *41*, 1488-1493.

*Warren, S., **Svenson, L.**, Woodhead, S., and Warren, K.G. (1996). Parental ancestry and the prevalence of multiple sclerosis (MS) in Alberta, Canada. *Neuroepidemiology, 15*, 1-10.

Svenson, L.W., Carmel, S., and Varnhagen, C.K. (1997). A review of the knowledge, attitudes and behaviours of university students concerning HIV/AIDS. *Health Promotion International*, *12*, 61-68.

*Robertson, C.M.T., **Svenson, L.W.,** and Joffres, M.R. (1998). Prevalence of cerebral palsy in Alberta. *Canadian Journal of Neurological Sciences*, *25*, 117-122.

Svenson, L.W., Schopflocher, D.P., Sauve, R.S., and Robertson, C.M.T. (1998). Alberta's infant mortality rate: the effect of the registration of live newborns weighing less than 500 grams. *Canadian Journal of Public Health*, *89*, 188-189.

***Svenson, L.W.,** Cwik, V.A., and Martin, W.R.W. (1999).The prevalence of motor

neurone disease in the province of Alberta. *Canadian Journal of Neurological Sciences, 26,*119-122.

Tough, S.C., Greene, C.A., **Svenson, L.W.,** and Belik, J. (2000). Effects of in vitro fertilization on low birth weight, preterm delivery, and multiple birth. *Journal of Pediatrics*, *136*, 618-622.

Tough, S.C., Greene, C.A., **Svenson, L.W.,** and Belik, J. (2000). Does maternal age predict multiple birth, preterm delivery or low birth weight in successful in vitro fertilization pregnancies? *Journal of the Society of Obstetricians and Gynaecologists of Canada, 22,* 938-941.

Sin, D.D., **Svenson, L.W.,** and Man, S.F.P. (2001). Do area-based markers of poverty accurately measure personal poverty? *Canadian Journal of Public Health*, *92*, 184-187.

Tough, S.C., **Svenson, L.W.,** Johnston, D.W, and Schopflocher, D.P. (2001). Characteristics of preterm delivery and low birth weight among 113,994 infants in Alberta: 1994-1996. *Canadian Journal of Public Health*, *92*, 276-280.

Spady, D.W., Schopflocher, D.P., **Svenson, L.W.,** and Thompson, A.H. (2001). Prevalence of mental disorder in children living in Alberta, Canada, as determined from physician billing data. *Archives of Pediatrics and Adolescent Medicine, 155*, 1153-1159.

Sin, D.D., Bell, N.R., **Svenson, L.W.**, and Man, S.F.P. (2002). The impact of follow-up visits to physicians' offices on repeat visits to emergency departments for patients with asthma and COPD: A population-based study. *American Journal of Medicine*, *112*, 120-125.

Tough, S.C., Newburn-Cook, C., Johnston, D.W., **Svenson, L.W.,** Rose, S., and Belik, J. (2002). Delayed childbearing and its impact on population rate changes in lower birth weight, multiple birth, and pre-term delivery. *Pediatrics*, *109*, 399-403.

Robertson, C.M.T., **Svenson, L.W.,** and Kyle, J.M. (2002). Birth weight by gestational age for Albertan liveborn infants, 1985 through 1998. *Journal of Obstetrics and Gynaecology Canada*, *24*, 138-148.

Sin, D.D., Wells, H., **Svenson, L.W.,** and Man, S.F.P. (2002). Asthma and COPD among Aboriginals in Alberta, Canada. *Chest*, *121*, 1841-1846.

Newburn-Cook, C.V., White, D., **Svenson, L.W.,** Demainczuk, N.N., Bott, N., and Edwards, J. (2002). Where and to what extent is prevention of low birth weight possible? *Western Journal of Nursing Research*, *24*, 887-904.

Warren, S., Warren, K.G., **Svenson, L.W.**, Schopflocher, D.P., and Jones, A. (2003). Geographic and temporal distribution of mortality rates for multiple sclerosis (MS) in Canada, 1965-94. *Neuroepidemiology*, *22*, 75-81.

*Yiannakoulias, N., **Svenson, L.W.**, Hill, M.D., Schopflocher, D.P., James, R.C., Wielgosz, A.T., and Noseworthy, T.W. (2003). Regional comparison of inpatient and outpatient patterns of cerebrovascular diagnosis in the province of Alberta. *Chronic Diseases in Canada, 24*, 9-16.

Tough, S.C., Faber, A.J., **Svenson, L.W.,** and Johnston, D.W. (2003). Is paternal age associated with an increased risk of low birthweight, preterm delivery, and multiple birth? *Canadian Journal of Public Health*, *94*, 88-92.

Dryden, D.M., Saunders, L.D., Rowe, B.H., May, L.A., Yiannakoulias, N., **Svenson, L.W.,** Schopflocher, D.P., and Voaklander, D.C. (2003). The epidemiology of traumatic spinal cord injury in Alberta, Canada, *Canadian Journal of Neurological Sciences, 30*, 113-121.

Sin, D.D., **Svenson, L.W.,** Cowie, R.L., and Man, S.F.P. (2003). Can universal access to health care eliminate health inequalities between children of poor and non-poor families? A case study of childhood asthma in Alberta. *Chest*, *124*, 51-56.

Tough, S.C., Newburn-Cook, C.V., White, D.E., Fraser-Lee, N.J., Faber, A.J., Frick, C., Sauve, R., and **Svenson, L.W.** (2003). Do maternal characteristics and past pregnancy experiences predict delivery among women aged 20-34? *Journal of Obstetrics and Gynaecology Canada*, *25*, 656-666.

Kelly, K.D., Pickett, W., Yiannakoulias, N., Rowe, B.H., Schopflocher, D.P., **Svenson, L.,** and Voaklander, D.C. (2003). Medication use and falls in community-dwelling older persons. *Age and Ageing*, *32*, 503-509.

Yiannakoulias, N., Rowe, B.H., **Svenson, L.W.,** Schopflocher, D.P., and Voaklander, D.C. (2003). Zones of prevention: the geography of fall injuries in the elderly, *Social Sciences and Medicine*, *57*, 2065-2073.

Yiannakoulias, N., **Svenson, L.W.,** Hill, M.D., Schopflocher, D.P., Rowe, B.H., James, R.C., Wielgosz, A.T., and Noseworthy, T.W. (2004). Incident cerebrovascular disease in rural and urban Alberta. *Cerebrovascular Diseases*, *17*, 72-78.

Schopflocher, D.P., Russell, M.L., **Svenson, L.W.,** Nguyen, T., and Mazurenko, I. (2004). Pandemic influenza planning: using the U.S. Centers for Disease Control FluAid software for small area estimation in the Canadian Context. *Annals of Epidemiology, 14*, 73-76.

Sin, D.D., Spier, S., **Svenson, L.W.,** Schopflocher, D.P., Senthilselvan, A., Cowie, R.L., and Man, S.F.P. (2004). The relationship between birth weight and childhood asthma: A population-based cohort study. *Archives of Pediatrics and Adolescent Medicine*, *158*, 60-64.

Spady, D.W., Saunders, L.D., Schopflocher, D.P., and **Svenson, L.W.** (2004). Patterns of injury in children: A population-based approach. *Pediatrics*, *113*, 522-529.

James, R.C., Blanchard, J.F., Campbell, D., Chapman, S., Clottey, C., Osei, W., **Svenson, L.W.,** and Noseworthy, T.W. (2004). A model for non-communicable disease surveillance in Canada: The prairie pilot diabetes surveillance system. *Chronic Diseases in Canada*, *25*, 7-12.

Hill, M.D., Yiannakoulias, N., Jeerakathil, T., Tu, J.V., **Svenson, L.W.,** and Schopflocher, D.P. (2004). The high risk of stroke immediately after transient ischemic attack: A population-based study. *Neurology*, *62*, 2015-2020.

Joffres, M.R., Kamath, T.V., Williams, G.R., Casey, J., and **Svenson, L.W.** (2004). Impact of guidelines on health care use for the management of dyslipidemia in two Canadian provinces, Alberta and Nova Scotia, 1990-2001. *Canadian Journal of Cardiology*, *20*, 767-772.

Humphries, K.H., Jackevicius, C.A., Gong, Y., **Svenson, L.,** Cox, J., Tu, J.V., and Laupacis,

A. (2004). Population rates of hospitalization for atrial fibrillation/flutter in Canada. *Canadian Journal of Cardiology*, *20*, 869-876.

*Yiannakoulias, N., Russell, M.L., **Svenson, L.W.,** and Schopflocher, D.P. (2004). Doctors, patients and influenza like illness: clinicians or patients at risk? *Public Health*, *118*, 527-531.

Dryden, D.M., Saunders, L.D., Rowe, B.H., May, L.A., Yiannakoulias, N., **Svenson, L.W.**, Schopflocher, D.P., and Voaklander, D.C. (2004). The utilization of health services following traumatic spinal cord injury: a 6-year follow-up study. *Spinal Cord*, *42*, 513-525.

Sin, D.D., Man, S.F.P., Cowie, R.L., Sharp, H.M., Andrews, E.M., Bell, N.R., Nilsson, C.I., Spier, S., **Svenson, L.W.,** Trachsel, L.C., and Tsuyuki, R.T. (2004). Recruitment for a provincial asthma study: Participation of network and non-network primary care physicians. *Canadian Family Physician*, *50*, 1251-1254.

Colman, I., Yiannakoulias, N., Schopflocher, D., **Svenson, L.W.,** Rosychuk, R.J., and Rowe, B.H. (2004). A population-based study of medically treated self-inflicted injuries, *Canadian Journal of Emergency Medicine*, *6*, 313-320.

Laupland, K.B., Zygun, D.A., Doig, C.J., Bagshaw, S.M., **Svenson, L.W.,** and Fick, G.H. (2005). One-year mortality of bloodstream infection-associated sepsis and septic shock among patients presenting to a regional critical care system. *Intensive Care Medicine*, *31*, 213-219.

Spady, D.W., Schopflocher, D.P., **Svenson, L.W.,** and Thompson, A.H. (2005). Medical and psychiatric comorbidity and health care use among children 6 to 17 years old. *Archives of Pediatrics and Adolescent Medicine*, *159*, 231-237.

Cox, J.L, Ramer, S.A., Lee, D.S., Humphries, K., Pilote, L., **Svenson, L.,** and Tu, J.V. (2005). Pharmacological treatment of congestive heart failure in Canada: a description of care in five provinces. *Canadian Journal of Cardiology*, *21*, 337-343.

Patten, S.B., **Svenson, L.W.,** and Metz, L. (2005). Descriptive epidemiology of affective disorders in multiple sclerosis. *CNS Spectrums*, *10*, 365-371.

Dryden, D.M., Saunders, L.D., Rowe, B.H., May, L.A., Yiannakoulias, N., **Svenson, L.W.,**

Schopflocher, D.P., and Voaklander, D.C. (2005). Depression following traumatic spinal cord injury. *Neuroepidemiology*, *25*, 55-61.

Beck, C.A., Metz, L.M., **Svenson, L.W.,** and Patten, S.B. (2005). Regional variation of multiple sclerosis prevalence in Canada. *Multiple Sclerosis*, *11*, 516-519.

Yiannakoulias, N., **Svenson, L.W.,** and Schopflocher, D.P. (2005). Commentary: Diagnostic uncertainty and medical geography: what are we mapping? *The Canadian Geographer*, *49*, 291-300.

Patten, S.B., **Svenson, L.W.,** and Metz, L.M. (2005). Psychotic disorders in MS: Population-based evidence of an association, *Neurology*, *65*, 1123-1125.

Bagshaw, S.M., Laupland, K.B., Doig, C.J., Mortis, G., Fick, G.H., Mucenski, M., Godinez-Luna, T., **Svenson, L.W.,** and Rosenal, T. (2005). Population-based epidemiology of severe acute renal failure and prognosis for renal recovery in critically ill patients. *Critical Care*, *9*, R700 – R709.

Russell, M.L., **Svenson, L.W.,** Yiannakoulias, N., Schopflocher, D.P., Virani, S.N., and Grimsrud, K. (2005). The changing epidemiology of chickenpox in Alberta, *Vaccine*, *23*, 5398 – 5403.

Dryden, D.M., Saunders, L.D., Jacobs, P., Schopflocher, D.P., Rowe, B.H., May, L.A., Yiannakoulias, N., **Svenson, L.W.,** and Voaklander, D.C. (2005). Direct health care costs following traumatic spinal cord injury. *Journal of Trauma Injury, Infection, and Critical Care*, *59*, 443 – 449.

Jayaraman, G.C., Singh, A.E., Yiannakoulias, N., Gratrix, J., Anderson, B., Johnson, M., **Svenson, L.,** Kress, B., Tyrell, G.J., and Wong, T. (2006). Sub-regional variations in the epidemiology of *Neisseria gonorrhea* in a large urban region in Alberta, Canada: Results from spatial analyses using routinely collected surveillance data. *Canada Communicable Disease Report*, *32*, 29-38.

Voaklander, D.C., Kelly, K.D., Rowe, B.H., Schopflocher, D.P., **Svenson, L.,** Yiannakoulias,

N., and Pickett, W. (2006). Pain, medication and injury in older farmers. *American Journal of Industrial Medicine*, *49*, 374-382.

Johnson, D.W., Craig, W., Brant, R., Mitton, C., **Svenson, L.,** and Klassen, T.P. (2006). A cluster randomized controlled trial comparing three methods of disseminating practice guidelines for children with croup [ISRCTN73394937]. *Implementation Science*, *1*, 10.

Pearl, D.L., Louie, M., Chui, L., Doré, K., Grimsrud, K.M., Leedell, D., Martin, S.W., Michel, P., **Svenson, L.W.,** and McEwen, S.A. (2006). The use of outbreak information in the interpretation of clustering of reported cases of *Escherichia coli* O157 in space and time in Alberta, Canada, 2000-2002. *Epidemiology and Infection*, *134*, 699-711.

Bernstein, C.N., Wajda, A., **Svenson, L.W.**, McKenzie, A., Koehoorn, M., Jackson, M., Fedorak, R., Israel, D., and Blanchard, J.F. (2006). The epidemiology of inflammatory bowel disease in Canada: A population-based study. *American Journal of Gastroenterology*, *101*, 1559-1568.

Plotnikoff, R.C., Taylor, L.M., Wilson, P.M., Courneya, K., Sigal, R.J., Birkett, R., Raine, K., and **Svenson, L.W.** (2006). Factors associated with physical activity in Canadian adults with diabetes. *Medicine and Science in Sports and Exercise*, *38*, 1526- 1534.

Rabi, D.M., Edwards, A.L., Southern, D.A., **Svenson, L.W.,** Sargious, P.M., Norton, P., Larsen, E.T., and Ghali, W.A. (2006). Association of socio-economic status with diabetes prevalence and utilization of diabetes care services. *BMC Health Services Research, 6,* 124.

Yiannakoulias, N.W., Schopflocher, D.P., and **Svenson, L.W.** (2006). Modelling geographic variations in West Nile virus. *Canadian Journal of Public Health*, *97*, 374-378.

Gilbride, S.J., Wild, T.C., Spady, D.W., **Svenson, L.W.,** and Wilson, D. (2006). Socio-economic Status and Types of Childhood Injury in Alberta: A Population Study. *BMC Pediatrics*, *6*, 30.

Pearl, D.L., Louie, M., Chui, L., Doré, K., Grimsrud, K.M., Martin, S.W., Michel, P., **Svenson, L.W.,** and McEwen, S.A. (2007). The use of randomization tests to assess the degree of similarity in PFGE patterns of *E. coli* O157 isolates from know outbreaks and statistical space-time clusters. *Epidemiology and Infection*, *135*, 100-109.

*Patten, S.B., **Svenson, L.W.,** White, C.M., Khalid, S.M., and Metz, L.M. (2007). Affective disorders in motor neuron disease: a population-based study. *Neuroepidemiology*, *28*, 1-7.

Warren, S., **Svenson, L.W.,** Warren, K.G., Metz, L.M., Patten, S.B., and Schopflocher, D.P. (2007). Incidence of multiple sclerosis among First Nations people in Alberta, Canada. *Neuroepidemiology*, *28*, 21-27.

Sun, J., Gooch, K., **Svenson, L.W.,** Bell, N.R., and Frank, C. (2007). Estimating osteoarthritis incidence from population-based administrative health care databases. *Annals of Epidemiology*, *17*, 51-56.

Rabi, D.M., Edwards, A.L., **Svenson, L.W.,** Sargious, P.M., Norton, P., Larsen, E.T., and Ghali, W.A. (2007). Clinical and medication profiles stratified by household income in patients referred to diabetes care. *Cardiovascular Diabetology*, *6*, 11.

Choi, B.C.K., Frank, J., Mindell, J.S., Orlova, A., Lin, L., Vaillancourt, A.D.M.G., Puska, P., Pang, T., Skinner, H., Marsh, M., Mokdad, A.H., Yu, S.-Z., Lindner, M.C., Sherman, G., Barreto, S.M., Green, L.W., **Svenson, L.W.,** Sainsbury, P., Yan, Y., Zhang, Z.-F., Zevallos, J.C., Ho, S., and de Salazar, L.M. (2007). Vision for a global registry of anticipated public health studies. *American Journal of Public Health*, *97 (Suppl. 1)*,

S82-S87.

Yiannakoulias, N.W., and **Svenson, L.W.** (2007). West Nile virus. Strategies for predicting municipal-level infection. *Annals of the New York Academy of Sciences*, *1102(1)*, 135-148.

***Svenson, L.W.,** Warren, S., Warren, K.G., Metz, L.M., Patten, S.B., and Schopflocher, D.P. (2007). Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, *34*, 175-180.

Yiannakoulias, N., Karosas, A., Schopflocher, D.P., **Svenson, L.W.,** and Hodgson, M.J. (2007). Using quad trees to generate grid points for applications in geographic disease surveillance. *Advances in Disease Surveillance*, *3*, 2.

Russell, M.L., Schopflocher, D.P., **Svenson, L.W.,** and Virani, S.N. (2007). Secular trends in the epidemiology of shingles in Alberta. *Epidemiology and Infection*, *135*, 908-913.

*Yiannakoulias, N., Schopflocher, D.P., Warren, S.A., and **Svenson, L.W.** (2007). Parkinson's disease, multiple sclerosis and changes of residence in Alberta. *Canadian Journal of Neurological Sciences*, *34*, 343-348.

Gao, S., Manns, B.J., Culleton, B.F., Tonelli, M., Quan, H., Crowshoe, L., Ghali, W.A., **Svenson, L.W.,** and Hemmelgarn, B.R. (2007). Prevalence of chronic kidney disease and survival among Aboriginal people. *Journal of the American Society of Nephrology*, *18*, 2953-2959.

Russell, M.L., Schopflocher, D.P., and **Svenson, L.W.** (2008). Health disparities in chickenpox or shingles? *Canadian Journal of Public Health*, *99*, 41-45.

Hauck, L.J., White, C., Feasby, T.E., **Svenson, L.W.,** Hill, M.D. (2008). Guillain-Barré syndrome in Alberta, Canada: an administrative data study. *Journal of Neurology,*

Neurosurgery, and Psychiatry, *79*, 318-320.

Pearl, D.L., Louie, M., Chui, L., Doré, K., Grimsrud, K.M., Martin, S.W., Michel, P., **Svenson, L.W.,** McEwen, S.A. (2008). Epidemiological characteristics of reported sporadic and outbreak cases of e. coli O157 in people in Alberta, Canada (2000-2002): Methodological challenges of comparing clustered to unclustered data. *Epidemiology and Infection*, *136*, 483-491.

Plotnikoff, R.C., Kotovych, M., Johnson, J.A., Karunamuni, N., and **Svenson, L.W.** (2008). Health-related behaviours in adults with diabetes: Associations with health care utilization and costs. *Canadian Journal of Public Health*, *99*, 227-231.

Warren, S., **Svenson, L.W.,** and Warren, K.G. (2008). Contribution of incidence to increasing prevalence of multiple sclerosis (MS) in Alberta, Canada. *Multiple Sclerosis*, *14*, 872-879.

Kwong, J.C., Stukel, T.A., Lim, J., McGeer, A.J., Upshur, R.E.G., Johansen, H., Sambell, C., Thompson, W.W., Thiruchelvam, Marra, F., D., **Svenson, L.W.,** and Manuel, D.G. (2008). The effect of universal influenza immunization on mortality and health care use. *PLoS Medicine*, *5*, e211.

Gao, S., Manns, B.J., Culleton, B.F., Tonelli, M., Quan, H., Crowshoe, L., Ghali, W.A., **Svenson, L.W.,** Ahmed, S., and Hemmelgarn, B.R. (2008). Access to health care among status Aboriginal people with chronic kidney disease. *Canadian Medical Association Journal*, *179*, 1007-1012.

Yiannakoulias, N., and **Svenson, L.W.** (2009). Differences between notifiable and administrative health information in the spatial-temporal surveillance of enteric infections. *International Journal of Medical Informatics*, *79*, 417-424.

Yiannakoulias, N.W., Hill, M.D., and **Svenson, L.W.** (2009). Geographic hierarchies of diagnostic practice style in cerebrovascular disease. *Social Science and Medicine*,

68, 1985-1992.

Pearl, D.L., Louie, M., Chui, L., Doré, K., Grimsrud, K.M., Martin, S.W., Michel, P., **Svenson, L.W.,** and McEwen, S.A. (2009). A multi-level approach for investigating socio-economic and agricultural risk factors associated with rates of reported cases of Escherichia coli O157 in humans in Alberta, Canada, 2000-2002. *Zoonoses and Public Health*, *56*, 455-464.

Yiannakoulias, N, **Svenson, L.W.**, and Schopflocher, D.P. (2009). An integrated framework for the geographic surveillance of chronic disease. *International Journal of Health Geographics*, *8*, 69.

Lee, D.S., Chiu, M., Manuel, D.G., Tu, K., Wang, X., Austin, P.C., Mattern, M.Y., Mitiku, T.F., **Svenson, L.W.,** Putnam, W., Flanagan, W.M., Tu, J.V. (2009) Trends in risk factors for cardiovascular disease in Canada: temporal, socio-demographic and geographic factors. *Canadian Medical Association Journal*, *181*, E55-E66.

Johnson, J.A., Balko, S.U, Hugel, G., Low, C., and **Svenson, L.W.** (2009) Increasing incidence and prevalence with limited survival gains among rural Albertans with diabetes, 1995 – 2006. *Diabetic Medicine*, *26*, 989-995.

Yiannakoulias, N., Schopflocher, D.P., and **Svenson, L.W.** (2009). Using administrative data to understand the geography of case ascertainment. *Chronic Diseases in Canada, 30,* 20-28.

Laupland, K.B., **Svenson, L.W.,** Grant, V., Ball, C.C., Mercado, M., Kmet, L., Kirkpatrick, A.W. (2010). Long-term mortality outcome of victims of major trauma. *Injury, 41,* 69-72.

Rabi, D.M, Edwards, A.L., **Svenson, L.W.,** Graham, M., Knutdson, M., and Ghali, W.A. (2010). Association of median household income with burden of coronary artery disease among individuals with diabetes. *Circulation: Cardiovascular Quality and*

Outcomes, *3*, 48-53.

Southern, D.A., Roberts, B., Edwards, A., Dean, S., Norton, P., **Svenson, L.W.**, Larsen, E., Sargious, P., Lau, D.C.W., and Ghali, W.A. (2010). Validity of Administrative Data Claim-Based Methods for Identifying Individuals with Diabetes at a Population Level. *Canadian Journal of Public Health*, *101*, 61-64.

Plotnikoff, R.C., Lightfoot, P., McFall, S., Spinola, C., Johnson, S.T., Prodaniuk, T., Predy, G., Tremblay, M.S., and **Svenson, L.** (2010). Child Health Ecological Surveillance System (CHESS) for childhood obesity: a feasibility study. *Chronic Diseases in Canada*, *30*, 95-106.

Sadowski, D.C., Ackah, F., Jiang, B., and **Svenson, L.W.** (2010). Achalasia: incidence, prevalence and survival. A population-based study. *Neurogastroenterology and Motility, 22,* e256-e261.

Laupland, K.B., **Svenson, L.W.,** Gregson, D.B., and Church, D.L. (2011). Long-term mortality associated with community-onset bloodstream infection. *Infection*, *39*, 405-410.

Oster, R.T., Johnson, J.A., Hemmelgarn, B.R., King, M., Balko, S.U., **Svenson, L.W.,** Crowshoe, L., and Toth, E.L. (2011). Recent epidemiologic trends of diabetes mellitus among status Aboriginal adults. *Canadian Medical Association Journal*, *183*, E803-E808.

Kim, J., Ferrato, C., Golding, G.R., Mulvey, M.R., Simmonds, K.A., **Svenson, L.W.,** Keays, G., Chui, L., Lovgren, M., and Louie, M. (2011). Changing epidemiology of methicillin-resistant *Staphylococcus aureus* in Alberta, Canada: population-based surveillance 2005-2008. *Epidemiology and Infection*, *139*, 1009-1018.

Barnabe, C., Joseph, L., Belisle, P., Labrecque, J., Edworthy, S., Barr, S.G., Fritzler, M., **Svenson, L.W.**, Hemmelgarn, B., and Bernatsky, S. (2012). Prevalence of systemic

lupus erythematosus and systemic sclerosis in the First Nations population of Alberta, Canada. *Arthritis Care & Research*, *64*, 138-143.

Oster, R.T., Johnson, J.A., Balko, S.U., **Svenson, L.W.,** and Toth, E.L. (2012). Increasing rates of diabetes amongst status Aboriginal youth in Alberta, Canada. *International Journal of Circumpolar Health*, 71, 1-7.

Crago, B., Ferrato, C., Drews, S.J., **Svenson, L.W.,** Tyrrell, G., and Louie, M. (2012). Prevalence of *staphylococcus aureus* and methicillin-resistant *S. aureus* (MRSA) in food associated with foodborne illness in Alberta, Canada, 2007 to 2010. *Food Microbiology*, *32*, 202-205.

Marrie, R.A., Yu, B.N, Leung, S., Elliot, L., Caetano, P., Warren, S. Wolfson, C., Patten, S.B., **Svenson, L.W.,** Tremlett, H., Fisk, J., and Blanchard, J.F. (2012). Rising prevalence of vascular comorbidities in MS: validation of administrative definitions for diabetes, hypertension, hyperlipidemia. *Multiple Sclerosis Journal*, *18*, 1310-1319.

Barnabe, C., Joseph, L., Belisle, P., Labrecque, J., Barr, S.G. Fritzler, M., **Svenson, L.,** Peschken, C., Hemmelgarn, B., and Bernatsky, S. (2012). Prevalence of autoimmunie inflammatory myopathy in Alberta's First Nations population. *Arthritis Care & Research*, *64,* 1715-1719.

Marrie, R.A., Yu, B.N. Leung, S., Elliot, L., Caetano, P., Warren, S., Wolfson, C., Patten, S.B., **Svenson, L.W.,** Tremlett, H., Fisk, J., Blanchard, J.F. (2013). The utility of administrative data for surveillance of comorbidity in multiple sclerosis: A validation study. *Neuroepidemiology*, *40*, 85-92.

Marrie, R.A., Fisk, J.D., Yu, B.N., Elliot, L., Caetano, P., Warren, S., Evans, C., Wolfson, C., **Svenson, L.W.,** Tremlett, H., Blanchard, J.F., and Patten, S.B. (2013). Mental comorbidity and multiple sclerosis: Validating administrative data to support

population-based surveillance. *BMC Neurology*, 13: 16.

Quan, H., Chen, C., Tu, K., Bartlett, G., Butt, D.A., Campbell, N.R.C., Hemmelgarn, B.R., Hill, M.D., Johansen, J., Khan, N., Lix, L.M., Smith, M., **Svenson, L.,** Walker, R.L., Wielgosz, A., McAlister, F.A., for Hypertension Outcome and Surveillance Team. (2013). Outcomes among 3.5 million newly diagnosed hypertensive Canadians. *Canadian Journal of Cardiology*, *29*, 592-597.

Yiannakoulias, N., Bland, W., and **Svenson, L.W.** (2013). Estimating the effect of turn penalties and traffic congestion on measuring spatial accessibility to primary health care. *Applied Geography*, *39*, 172-182.

So, H.C., Pearl, D.L., von Königslow, T., Louie, M., Chui, L., and **Svenson, L.W.** (2013). Spatio-temporal scan statistics for the detection of outbreaks involving common molecular subtypes: Using human cases of *Escherichia coli* O157:H7 provincial PFGE pattern 8 (national-designation ECXAI.0001) in Alberta as an example. *Zoonoses and Public Health*, *60*, 341-348.

Marrie, R.A., Yu, B.N., Leung, S., Elliot, L. Caetano, P., Warren, S., Wolfson, C., Patten, S.B., **Svenson, L.W.,** Tremlett, H., Fisk, J., and Blanchard, J.F. (2013). Prevalence and incidence of ischemic heart disease in multiple sclerosis: A population-based validation study. *Multiple Sclerosis and Related Disorders*, *2*, 355-361.

Robitaille, C., Bancej, C., Dai, S., Tu, K., Rasali, D., Blais, C., Plante, C., Smith, M., **Svenson, L.W.,** Reimer, K., Casey, J., Puchtinger, R., Johansen, H., Gurevich, Y., Waters, C., Lix, L.M., and Quan, H. (2013). Surveillance of ischemic heart disease should include physician billing claims: Population-based evidence from administrative health data across seven Canadian provinces. *BMC Cardiovascular Disorders*, *13*, 88

Fiest, K.M., Pringsheim, T., Patten, S.B., **Svenson, L.W.,** and Jetté, N. (2014). The role of

systematic reviews and meta-analyses of incidence and prevalence studies in neuroepidemiology. *Neuroepidemiology*, *42*, 16-24.

Choi, B.C.K., Decou, M.L., Rasali, D., Martens, P.J., Mancuso, M., Plotnikoff, R.C., Neufdorf, C., Thanos, J., **Svenson, L.W.,** Denny, K., Orpana, H., Stewart, P., King, M., Griffith, J., Erickson, T., van Dorp, R., White, D., and Ali, A. (2014). Enhancing capacity for risk factor surveillance at the regional/local level: a follow-up review of the findings of the Canadian Think Tank Forum after 4 years. *Archives of Public Health*, *72*, 2.

Blais, C., Dai, S., Waters, C., Robitaille, C., Smith, M., **Svenson, L.W.,** Reimer, K., Casey, J., Puchtinger, R., Johansen, H., Gurevich, Y., Lix, L.M., Quan, H., and Tu, K. (2014). Assessing the burden of hospitalized and community-care heart failure in five Canadian provinces. *Canadian Journal of Cardiology*, *30*, 352-358.

Broten, L., Aviña-Zubieta, J.A., Lacaille, D., Joesph, L., Hanly, J.G., Lix, L., O'Donnell, S., Barnabe, C., Fortin, P.R., Hudson, M., Jean, S., Peschken, C., Edworthy, S.M., **Svenson, L.,** Pineau, C.A., Clarke, A.E., Smith, M., Bélisle, P., Badley, E.M., Bergeron, L., and Bernatsky, S. (2014). Systemic autoimmune rheumatic disease prevalence in Canada: Updated analyses across seven provinces. *The Journal of Rheumatology*, *41*, 673-679.

*MacDonald, S.E., Dover, D.C., Simmonds, K.A., and **Svenson, L.W.** (2014). Risk of febrile seizures after first dose of measles-mumps-rubella-varicella vaccine: a population-based cohort. *Canadian Medical Association Journal*, 186, 824-829.

Reich, K.M., Chang, H.J., Rezaie, A., Wang, H., Goodman, K.J., Kaplan, G.G., **Svenson, L.W.,** Lees, G., Fedorak, R.N., and Kroeker, K.I. (2014). The incidence rate of colectomy for medically-refractory ulcerative colitis has declined in parallel with increasing anti-TNF use: A time-trend study. *Alimentary Pharmacology and Therapeutics*, *40*, 629-638.

Cunningham, C.T., Cai, P., Topps, D., **Svenson, L.W.,** Jetté, N., and Quan, H. (2014). Mining rich health data from Canadian physician claims: Features and face validity. *BMC Research Notes*, *7*, 682.

*Russell, M.L., Dover, D.C., Simmonds, K.A., and **Svenson, L.W.** (2014). Shingles in Alberta: Before and after publicly funded varicella vaccination. *Vaccine*, *47*, 6319-6324.

Liu, X.C., Simmonds, K.A., Russell, M.L., and **Svenson, L.W.** (2014). Herpes zoster vaccine (HZV): Utilization and coverage 2009 – 2013, Alberta, Canada. *BMC Public Health*, *14*, 1098.

Bifolchi, N., Michel, P., Talbot, J., **Svenson, L.,** Simmonds, K., Checkly, S., Chui, L., Dick, P., and Wilson, J.B. (2014). Weather and livestock factors for *Escherichia coli* O157 human infection in Alberta, Canada. *Epidemiology and Infection*, *142*, 2302-2313.

Cherry, N., Beach, J., Burstyn, I., Parboosingh, J., Schouchen, J., Senthilselvan, A., **Svenson, L.,** Tamminga, J., and Yiannkoulias, N. (2015). Genetic susceptibility to beryllium: a case-referent study of men and women of working age with sacroidosis or other lung disease. *Occupational Environmental Medicine, 72*, 21-27.

Shiff, N.J., Lix, L.M., Oen, K., Joseph, L., Duffy, C., Stringer, E., Tucker, L.B., **Svenson, L.W.,** Belisle, P., and Bernatsky, S. (2015). Chronic inflammatory arthritis prevalence estimates for children and adolescents in three Canadian provinces. *Rheumatology International*, *35*, 345-350

Shiff, N.J., Lix, L.M., Joseph, L., Duffy, C., Tucker, L.B., **Svenson, L.W.,** Belisle, P., and Bernatsky, S. (2015). The prevalence of systemic autoimmune rheumatic diseases in Canadian pediatric populations: Administrative database estimates.

Rheumatology International, *35*, 569-573.

Peng, M., Chen, G., Lix, L.M., McAlister, F.A., Campbell, N., Hemmelgarn, B.R., **Svenson, L.W.,** and Quan, H. (2015). Refining hypertension surveillance to account for potentially misclassified cases. *PLoS One*, *10*, e0119186.

Bernatsky, S., Smargiassi, A., Johnson, M., Kaplan, G.G., Barnabe, C., **Svenson, L.,** Brand, A., Bertazzon, S., Hudson, S., Clarke, A.E., Fortin, P., Edworthy, S., Bélisle, P., and Joseph, L. (2015). Fine particulate air pollution, nitrogen dioxide, and systemic autoimmune rheumatic disease in Calgary, Alberta. *Environmental Research*, *140*, 474-478.

Bennett, D.A., Brayne, C., Feigin, V.L., Brainin, M., Davis, D., Gallo, V., Jetté, N., Karch, A., Kurtzke, J.F., Lavados, P.M., Logroscino, G., Nagel, G., Preux, P.M., Rothwell, P.M., and **Svenson, L.W.** (2015). Development of the standards of reporting neurological disorders (STROND) checklist: A guideline for the reporting of incidence and prevalence studies in neuroepidmiology. *European Journal of Epidemiology, 30*, 569-576.

Bennett, D.A., Brayne, C., Feigin, V.L., Brainin, M., Davis, D., Gallo, V., Jetté, N., Karch, A., Kurtzke, J.F., Lavados, P.M., Logroscino, G., Nagel, G., Preux, P.M., Rothwell, P.M., and **Svenson, L.W.** (in press). Development of the standards of reporting neurological disorders (STROND) checklist: A guideline for the reporting of incidence and prevalence studies in neuroepidmiology. *Neurology, 85*, 821-828.

Kingwell, E., Zhu, Z., Marrie, R.A., Fisk, J.D., Wolfson, C., Warren, S., Profetto-McGrath, J., **Svenson, L.W.,** Jetté, N., Bhan, V., Yu, N., Elliot, L., Tremlett, H., for the CIHR Team in the Epidemiology and Impact of Comorbidity in MS. (2015). High incidence and increasing prevalence of multiple sclerosis in British Columbia, Canada. *Journal of Neurology*, *262*, 2352-2363.

Appendix B: Declarations of Contribution

Twelve publications, of 139 total publications at the time of program admission, have been included in this thesis. Of the twelve, two were single authored while the remaining ones were co-authored. I believe in collaboration as a means of improving the quality and usefulness of the research as well as a way of fostering professional development among all collaborators, myself included. The publications included cover the period 1990 through to 2014. For each paper with multiple authors, one co-author completed the RDPUB form. The table below summaries my percentage contribution to each paper included in the thesis. The publications are listed in the order in which they are cited in the thesis.

Table B.1: List of publications, and percentage contribution, presented in this thesis

Number	Publication	Percentage Contribution
1	Svenson, L.W. (1990). Geographic distribution of deaths due to Parkinson's disease in Canada: 1979 – 1986. *Movement Disorders*, *5*, 322-324.	100%
2	Svenson, L.W. (1991). Regional disparities in the annual prevalence rates of Parkinson's disease in Canada. *Neuroepidemiology*, *10*, 205-210.	100%
3	Robertson, C.M.T., Svenson, L.W., & Joffres, M.R. (1998). Prevalence of cerebral palsy in Alberta. *Canadian Journal of Neurological Sciences*, *25*, 117-122.	60%
4	Yiannakoulias, N., Svenson, L.W., Hill, M.D., Schopflocher, D.P., James, R.C., Wielgosz, A.T., & Noseworthy, T.W. (2003). Regional comparisons of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chronic Diseases in Canada*, *24*, 9-16.	50%
5	Yiannakoulias, N., Schopflocher, D.P., Warren, S.A., & **Svenson, L.W.** (2007). Parkinson's disease, multiple sclerosis and changes in residence in Alberta. *Canadian Journal of Neurological Sciences*, *34*, 343-348.	35%

Number	Publication	Percentage Contribution
6	Yiannakoulias, N., Russell, M.L., Svenson, L.W., & Schopflocher, D.P. (2004). Doctors, patients and influenza-like illness: clinicians or patients at risk? *Public Health*, *118*, 527-531.	30%
7	Svenson, L.W., Cwik, V.A., & Martin, W.R.W. (1999). The prevalence of motor neurone disease in the province of Alberta. *Canadian Journal of Neurological Sciences*, *26*, 119-122.	80%
8	Patten, S.B., Svenson, L.W., White, C.M., Khaled, S.M., & Metz, L.M. (2007). Affective disorders in motor neuron disease: a population-based study. *Neuroepidemiology*, *28*, 1-7.	40%
9	Warren, S., Svenson, L., Woodhead, S., & Warren, K.G. (1996). Parental ancestry and risk of multiple sclerosis in Alberta, Canada. *Neuroepidemiology*, *15*, 1-9.	40%
10	Svenson, L.W., Warren, S., Warren, K.G., Metz, L.M., Patten, S.B., & Schopflocher, D.P. (2007). Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, *34*, 175-180.	40%
11	MacDonald, S.E., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Risk of febrile seizures after the first dose of measles-mumps-rubella-varicella vaccine: a population-based cohort study. *Canadian Medical Association Journal*, *186*, 824-829.	25%
12	Russell, M.L., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Shingles in Alberta: Before and after publicly funded varicella vaccination. *Vaccine*, *32*, 6319-6324.	25%

Appendix C: Publication Citation Reports

The table below provides a listing of the number of times each article included in the thesis has been cited according to two sources – Web of Science® and Google Scholar™. For both sources, the citation counts are as of July 21, 2015. The counts only include citations in peer reviewed professional publications and may be undercounts due to lags in the time from which a one of the papers was cited to when that information is available to the two query tools used. The publications are listed in the order in which they are cited in the thesis.

Table C.1: Citation counts from each publication included in the thesis

Number	Publication	Web of Science	Google Scholar
1	Svenson, L.W. (1990). Geographic distribution of deaths due to Parkinson's disease in Canada: 1979 – 1986. *Movement Disorders*, *5*, 322-324.	3	4
2	Svenson, L.W. (1991). Regional disparities in the annual prevalence rates of Parkinson's disease in Canada. *Neuroepidemiology*, *10*, 205-210.	10	26
3	Robertson, C.M.T., Svenson, L.W., & Joffres, M.R. (1998). Prevalence of cerebral palsy in Alberta. *Canadian Journal of Neurological Sciences*, *25*, 117-122.	35	57
4	Yiannakoulias, N., Svenson, L.W., Hill, M.D., Schopflocher, D.P., James, R.C., Wielgosz, A.T., & Noseworthy, T.W. (2003). Regional comparisons of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chronic Diseases in Canada*, *24*, 9-16.	7	17
5	Yiannakoulias, N., Schopflocher, D.P., Warren, S.A., & **Svenson, L.W.** (2007). Parkinson's disease, multiple sclerosis and changes in residence in Alberta. *Canadian Journal of Neurological Sciences*, *34*, 343-348.	2	2
6	Yiannakoulias, N., Russell, M.L., Svenson, L.W., &	2	6

Number	Publication	Web of Science	Google Scholar
	Schopflocher, D.P. (2004). Doctors, patients and influenza-like illness: clinicians or patients at risk? *Public Health*, *118*, 527-531.		
7	Svenson, L.W., Cwik, V.A., & Martin, W.R.W. (1999). The prevalence of motor neurone disease in the province of Alberta. *Canadian Journal of Neurological Sciences*, *26*, 119-122.	8	11
8	Patten, S.B., Svenson, L.W., White, C.M., Khaled, S.M., & Metz, L.M. (2007). Affective disorders in motor neuron disease: a population-based study. *Neuroepidemiology*, *28*, 1-7.	5	7
9	Warren, S., Svenson, L., Woodhead, S., & Warren, K.G. (1996). Parental ancestry and risk of multiple sclerosis in Alberta, Canada. *Neuroepidemiology*, *15*, 1-9.	5	5
10	Svenson, L.W., Warren, S., Warren, K.G., Metz, L.M., Patten, S.B., & Schopflocher, D.P. (2007). Prevalence of multiple sclerosis in First Nations people of Alberta. *Canadian Journal of Neurological Sciences*, *34*, 175-180.	13	18
11	MacDonald, S.E., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Risk of febrile seizures after the first dose of measles-mumps-rubella-varicella vaccine: a population-based cohort study. *Canadian Medical Association Journal*, *186*, 824-829.	5	12
12	Russell, M.L., Dover, D.C., Simmonds, K.A., & Svenson, L.W. (2014). Shingles in Alberta: Before and after publicly funded varicella vaccination. *Vaccine*, *32*, 6319-6324.	2	12