Examining Diabetes Inequalities Using Individual and Area-level Income in Urban and Rural Saskatchewan

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Community Health and Epidemiology

University of Saskatchewan

Saskatoon

By

Lisa Thurairasu

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Community Health & Epidemiology
Box 7, Health Science Building, 107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5
Canada

OR

Dean
College of Graduate Studies and Research
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan S7N 5A2
Canada

ABSTRACT

Background and Rationale:

Diabetes is a growing health problem in Saskatchewan, disproportionately affecting people in low socioeconomic groups compared to people more well off (1,2). The absence of individual-level socioeconomic status (SES) data in administrative databases necessitates researchers to use area-level SES as a proxy for measuring health inequalities (3). This study compares individual and area-level income for measuring diabetes inequalities in a CCHS study sample and estimated population to determine whether an area-based measure can be used as a proxy for individual-level data in urban and rural Saskatchewan.

Methods:

Health administrative data containing diabetes cases was linked to the 2007-2008 CCHS combined cycle containing individual-level income, which was merged with the 2006 Canadian Census containing area-level income at the dissemination area (DA)-level. Individual and area-level incomes were compared for the 'unweighted' and 'weighted' CCHS population. The 'unweighted' population was the CCHS study sample of Saskatchewan respondents in which no sampling weights or bootstrap weights were used. Contrarily, the 'weighted' population was the estimated Saskatchewan population derived from applying sampling weights and bootstrap weights to the study sample.

The statistical methods used in this study included descriptive analyses, bivariate analyses, and multivariable analyses. Odds ratios of the final multivariable models for the 'unweighted' and 'weighted' population were compared to determine whether area-based income underestimates diabetes inequalities. The software, SAS Enterprise Guide 6.1, was used to analyze the data.

Results:

There was relatively low agreement between individual and area-level income. Individual and area-level income had varying patterns of influence on diabetes in the study sample and in the estimated population. Overall, income gradients were larger in the 'unweighted' population compared to the 'weighted' population. The over-representation of older individuals (who have a higher proportion of diabetes than younger people) and the under-representation of younger individuals in the study sample compared to the estimated population and the 'actual' Census population as seen in Table 4.1, could have led to the stronger association in the ('unweighted') study sample (4,5). However, as individuals generally earn higher income with age (6) and as seniors often earn low income (2), these factors can only partially explain the observed patterns.

Of individuals with diabetes, gradients were observed between area-level income and proportion of individuals with diabetes, while a pattern was present for individual-level income, in the study sample and estimated population. Within each income category, the reverse pattern occurred for individual-level income and area-level income in the study sample and estimated population. However, age-adjusted rates revealed clear downward gradients. Based on the final logistic model, the odds of the 'unweighted' individual-level income model produced a downward gradient, while statistically insignificant U-shaped curves were present for the 'unweighted' area-level income model, 'weighted' individual-level income model, and the 'weighted' area-level income model. The odds of diabetes was also smaller in rural areas compared to urban areas in the final model, contradictory to literature (7,8), however the results were statistically insignificant.

Study Implications:

This study provides decision support for the use or disuse of area-level income in measuring diabetes inequalities accurately in Saskatchewan. Future research, especially qualitative research, can examine the mechanisms of individual and area-level income on health.

ACKNOWLEDGEMENTS

I would like to express my wholehearted gratitude to my supervisor, Dr. Cory Neudorf, for his patience, intelligent advice, guidance, and feedback during my studies. Dr. Neudorf gave me the opportunity to learn invaluable skills, which I am very grateful. He also supported me through challenges experienced pertinent to the study and provided research funding, which I am very thankful for.

I would also like to thank my thesis committee members: Dr. Nazeem Muhajarine, Dr. Gary Teare, and Dr. Anne Leis, who provided intelligent feedback, guidance, and motivation for carrying out my thesis. Each member brought a different strength and perspective to the research study and I am very appreciative to have such strong team members. I am also utmost grateful to Dr. Punam Pahwa, who was not on my committee, but provided invaluable guidance for my research study.

I would also like to thank the SPHERU health inequalities research team supervised by Dr. Neudorf, who contributed to the larger provincial project. I would specifically like to thank Dr. Dan Fuller for his time and effort in guiding the research study and Stuart Lockhart for obtaining the 2006 Saskatchewan Census profile from Statistics Canada.

I am also very thankful towards the HQC staff and Dr. Teare for allowing me to conduct my analysis at the Health Quality Council of Saskatchewan. I would like to specifically thank Nedeene Hudema who transferred countless files for me, Jacqueline Quail for her support and kindness, Eric Wang, Nicole Bilodeau, and other HQC staff for their support and kindness.

I am also appreciative of the staff in the Department of Community Health and Epidemiology. I would like to specifically thank my coursework professors, Dr. Punam Pahwa, Dr. Bonnie Janzen, Dr. Anne Leis, Dr. Rachel Engler-Stringer, Dr. June Lim, and Dr. Cindy Feng (School of Public Health) for their knowledge and expertise in the field, which lay the foundation for my thesis work. I would also like to express my gratitude to Cheryl Bolster, Cindy Elchuk, and Kathy Evans for their kindness and support. I am very appreciative of their friendship during the time I have known them.

I would like to thank my friends and colleagues who were kind and supportive during my time in Saskatoon. I will forever remember all the wonderful memories you have given me.

Last but not least, I would like to thank my family for all their love and support during my studies. They made many sacrifices for me, which I am very grateful for.

Disclaimer

This study was based, in part, on de-identified data made available by the Saskatchewan Ministry of Health to the Health Quality Council. The interpretations and conclusions herein do not necessarily represent those of the Government of Saskatchewan or the Saskatchewan Ministry of Health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABSM –Area-based socioeconomic measure

BMI –Body mass index

CAI –Computer Assisted Interviewing

CAPI –Computer Assisted Person Interviewing

CATI –Computer Assisted Telephone Interviewing

CCDSS –Canadian Chronic Disease Surveillance System

CCHS –Canadian Community Health Survey

CV –Coefficient of Variation

DA –Dissemination area

DAD –Discharge Abstract Database

ENWB –Elimination of Non-Working Banks

HQC –Health Quality Council

MAUP –Modifiable Area Unit Problem

MIZ –Metropolitan influenced zone

MSB –Medical Services Branch

PPS –Probability proportional to size

PRS –Person Registry System

RDD –Random Digit Dialing

SEP –Socioeconomic position

SES –Socioeconomic status

SPHERU –Saskatchewan Population Health and Evaluation Research Unit

LIST OF APPENDICES

CHAPTER 1:
INTRODUCTION

Chapter 1 provides the background context of the research, the study rationale, the study objectives, and the research questions with corresponding hypotheses. The purpose of Chapter 1 is to familiarize the reader with the research topic and provide the basis for undertaking the study.

1.1 Background

Health, which refers to physical and psychosocial well-being, is very important to Canadians, yet health differences exist among more or less socially disadvantaged groups based on race/ethnicity, gender, socioeconomic status, or other discriminating features used to establish a social hierarchy in a population (9,10). These health differences are known as social inequalities in health, and terms like *health inequalities* or *health disparities* are used in this paper to refer to the same concept (9-11). According to Braveman et al. (2011), health inequalities are “systematic, plausibly avoidable health differences,” which put socially disadvantaged groups at further disadvantage (10). Hence, these specific social inequalities constitute inequities because there is an unfair and unequal distribution of social determinants of health (i.e., access and availability to healthcare, education, etc.) across social groups (11,12).

Considerable evidence has shown that health differences resulting from varying levels of social advantage/disadvantage across socioeconomic status (SES), measured by income, education, and/or occupational level, produces a socioeconomic gradient (10,12,13). The socioeconomic gradient in health, also known as a “dose-response,” is a graded relationship between SES and health occurring at every level of the social hierarchy in which people from a low SES often have poorer health compared to people from a higher SES for several health outcomes (8,12). However, there are instances when the burden of disease is greater among higher socioeconomic groups compared to lower socioeconomic groups, in the case of breast cancer or melanoma, or the unequal distribution of leukemia among socioeconomic groups (8). Nonetheless, prominent studies like the *Black Report*, *Whitehall*, and *Acheson Report* first published findings on health inequalities and the socioeconomic gradient, encouraging further research into the causal pathways and determinants linking SES and health (14,15).

Socioeconomic indicators like income, education, and employment are used in deprivation indices like the Pampalon index to measure health inequalities in Canada (3). Deprivation, which was originally termed by Peter Townsend in the 1980s, was defined as a "state of observable and demonstrable disadvantage relative to the local community or the wider society or nation to which the individual, family or group belongs (16)." He classified deprivation into two categories: material and social. Material deprivation refers to the lack of goods and services associated with modern living such as access to motor vehicle, adequate housing, etc., whereas social deprivation is the disintegration of social relationships among family members, community members, and colleagues (16). Since these terms can be interpreted in different ways, operational definitions are needed to measure social inequalities in health (11).

Despite the versatility and flexibility of the Pampalon index, socioeconomic indicators can be used independently to measure health inequalities in a population. Previous studies have found income to be a widely-used measure of economic resources and a strong predictor in determining health outcomes like diabetes (17-21). Diabetes prevalence was 4.1 times higher for the lowest income earners than the highest income earners in Canada (18). In addition, low SES Canadians have higher rates of primary care visits and hospital admissions, yet greater difficulty accessing speciality clinics than higher SES Canadians (19,20). Individual and area-level incomes may have independent effects on diabetes development. Metcalf et al. (2008) found low household income and low area-based income to increase cardiovascular disease and diabetes risk factor rates (17). The importance of individual and area-based income measures should be studied in greater detail.

1.2 Study Rationale

It is important to conduct this study to determine whether the level of spatial scale is an important aspect for measuring health inequalities accurately. Due to the unavailability of individual-level socioeconomic data in administrative databases, researchers have compensated for the shortcoming by using area-level SES as a "proxy" for individual-level SES (3). Using area-level SES as a proxy for measuring health inequalities is quite debatable as discussed in section 2.2.3.a *Utilizing ABSMs with Caution* (22-24). In this study, individual and area-level income in measuring health inequalities are compared, which could determine whether an area-based socioeconomic measure (ABSM) underestimates health inequalities. As previous studies have found ABSMs to attenuate health disparities, it could be that area-based measures based on administrative boundaries may not adequately measure neighbourhood-specific place effects affecting a neighbourhood (25).

1.3 Objectives

The primary objective of this study is to determine whether area-level income at the dissemination area[1] (DA) geography can be used as a proxy for individual-level income for measuring diabetes inequalities in urban and rural Saskatchewan. Individual and area-level income inequalities will be compared in the study sample ('unweighted'[2]) and in the estimated population ('weighted'[3]) of the Canadian Community Health Survey (CCHS) to determine whether differences in inequalities exist from the sampling and weighting methods employed in the CCHS. The study sample refers to the respondents who were surveyed in the CCHS sample, while the estimated population is derived using sampling weights on the study sample. The first objective will be accomplished through descriptive analyses and separate logistic models in which area-based estimates are compared to individual-level estimates to determine whether "no to little" changes exist. The phrase *no to little* is subjective and depends on the researcher's viewpoint, but an overall examination of all findings from this study should allow the researcher to determine whether area-level income can be used as a proxy.

---

[1] A dissemination area (DA) is a spatial unit of geography consisting of approximately 125 to 144 households or about 400 to 700 people. It is the smallest available unit in Canada by which census data is aggregated (26).
[2] 'Unweighted' refers to the CCHS study sample of respondents originally surveyed in which sampling weights and bootstrap weights have not been applied.
[3] 'Weighted' refers to the Saskatchewan estimated population in which sampling weights and bootstrap weights have been applied to the study sample. More information on sampling and bootstrap weights can be found in section 3.1.5.d *Weighting.*

The secondary objective is to determine whether area-level income has an effect on diabetes development in urban and rural Saskatchewan of the study sample and the estimated population. The secondary objective will be completed through descriptive analyses and multivariate analysis to determine whether area-level income has a relationship with diabetes and geography type (another area-level variable).

1.4 Research Questions and Hypotheses

1.4.1 Research Questions

1. What is the type of relationship between diabetes and income (at the individual and area-level) in the study sample and in the estimated population?
2. How much agreement exists between individual-level income and area-level income in the study sample and in the estimated population?
    a. How does the relationship between individual and area-level income differ in urban and rural areas?
3. Are diabetes inequalities underestimated when measured using area-level income compared to individual-level income in urban and rural Saskatchewan?
    a. What are the predicted probabilities of diabetes when individual and area-level income are used in urban and rural Saskatchewan?
    b. How do the odds ratios of diabetes differ between individual and area-level income categories and between urban and rural areas?

1.4.2 Hypotheses

1.4.2.a Research Question#1 Hypotheses

It is well-known that socioeconomic gradients exist for several health outcomes including diabetes (8,12). Therefore, it is predicted that an inverse relationship between diabetes and income (at the individual and area-level) will exist in the study sample and in the estimated population. It is also predicted that gradients will be different in the study sample compared to the estimated population due to selection bias occurring in the study sample and the application of sampling weights used in the estimated population to adjust for selection differences (4).

Although Statistics Canada over-samples for youth aged 12 to 19 years of age, the representation of different individuals (i.e., age groups) in the study sample and estimated population is unknown prior to conducting the study. Since age is a confounder in the income-diabetes relationship (1), it is difficult to determine the direction and magnitude of any potential gradients.

It is hypothesized that individual-level income (represented as household income) will show a steeper gradient than area-level income since past studies have found area-based SES measures to underestimate socioeconomic gradients (27-30). Literature has shown that even the smallest spatial units (like the dissemination area (DA) or block group[4]) at which socioeconomic data is available lead to diminished gradients (27-30). These smaller gradients can be a result of any added and unmeasured contextual effects of a surrounding area or the misclassification of individuals by assigning them to an area-level income (27,30-32).

1.4.2.b Research Question#2 Hypotheses

Although there is evidence for both the use and disuse of area-based SES measures as a proxy for individual-level data (27,33-37), it is difficult to hypothesize how much agreement exists between individual and area-level income for the CCHS study population. While some studies have shown that individual and area-based SES measures can be used synonymously, other studies have proven otherwise (27,33-37). This study will measure percentage agreement between individual and area-level income through a weighted kappa statistic.

In urban and rural areas, it is predicted that the income gradient will be steeper for individual-level income compared to area-level income based on prior studies that have found smaller socioeconomic gradients for area-based measures (27-30). Although rural residents on average have poorer health than urban residents (7,8), previous studies have found smaller SES gradients in rural areas than in urban areas (3,33,38). In other words, the inequality between the residents is smaller in rural areas than between residents in urban areas (3,33,38), but the overall health of rural residents is poorer than their urban counterparts (7,8). Therefore, steeper gradients

[4] A block group is a spatial unit of geography in the U.S. consisting of approximately 1,000 individuals. It is the smallest available unit in the U.S. by which census data is aggregated (31).

are expected in urban areas than in rural areas for this study. As previously stated, it is also predicted that the gradients in the study sample will be different than the gradients in the estimated population due to selection bias in the study sample and the utilization of sampling weights in the estimated population to correct for selection differences (4). It is difficult to determine the magnitude and direction of the gradients since it is unknown beforehand how the study sample differs from the estimated population in terms of representation of certain individuals (i.e., older versus younger individuals).

1.4.2.c Research Question#3 Hypotheses

Previous studies have found socioeconomic gradients to be smaller when using area-based SES compared to individual-level SES (27-30). Therefore, it is expected that diabetes inequalities (measured via predicted probabilities and odds ratios) will be underestimated with area-level income compared to individual-level income in urban and rural areas. In this study, predicted probabilities are used to examine the probability of diabetes at each income level, while the odds ratios in the final logistic model reveal the probability of diabetes as one income level compares to another.

Based on prior studies that have found rural residents to have poorer health than urban residents (7,8), it is predicted that predicted probabilities and odds ratios for diabetes will be greater in rural areas than in urban areas. Secondly, gradients are expected to be smaller in rural areas based on previous research stating that health inequalities are smaller in rural areas than in urban areas (3,33,38).

Comparing the study sample to the estimated population, it is expected that predicted probabilities and odds ratios will be different in the study sample and in the estimated population due to selection bias in the sample and the use of sampling weights in the estimated population to adjust for selection differences (4). However, it is difficult to determine the magnitude and direction of the estimates since it is unknown prior to conducting the study the degree of representation of particular individuals (i.e., older versus younger individuals).

CHAPTER 2:
LITERATURE REVIEW

Chapter 2 provides the theoretical perspective and the background literature on the research study. First, a conceptual framework pertaining to the relationship between socioeconomic status (SES) and diabetes is presented to set the foundation of the thesis paper. Second, literature on the topic is discussed to provide the context and familiarize the reader with the field of study.

2.1 Conceptual Framework

The conceptual framework used to guide this research study is by Brown et al. (2004) and it focusses on the relationship between socioeconomic position (SEP) and diabetes. Socioeconomic position, which can be measured over time, refers to the socioeconomic status of individuals and surrounding areas in addition to the social relationships between individuals and the connections they have with their communities (39). The mechanisms linking SEP to diabetes are illustrated in the framework as seen in Figure 2.1 below. A more detailed version of the distal[5] factors at the various societal levels is seen in Figure 2.2. The underlying assumption behind the framework is that people of lower SEP are disproportionately affected by poorer health outcomes compared to people of higher SEP (39). Although the framework emphasizes the effect of SES on health, the authors also acknowledge reverse causality in which diabetes has an effect on income earnings. However, due to strong evidence from previous research, they focus on the influences of SEP on diabetes (39).

Socioeconomic position can influence health directly and indirectly through characteristics at the individual, community, health care provider, and health care system level as illustrated in pathways 1,2, and 3 of Figure 2.1 (39). Proximal factors, which directly affect health, include health behaviours, health care access, and health professional care, while distal factors act through proximal factors to influence health (39). For example, proximal factors like blood glucose monitoring and exercise are important for controlling diabetes (39). However, the

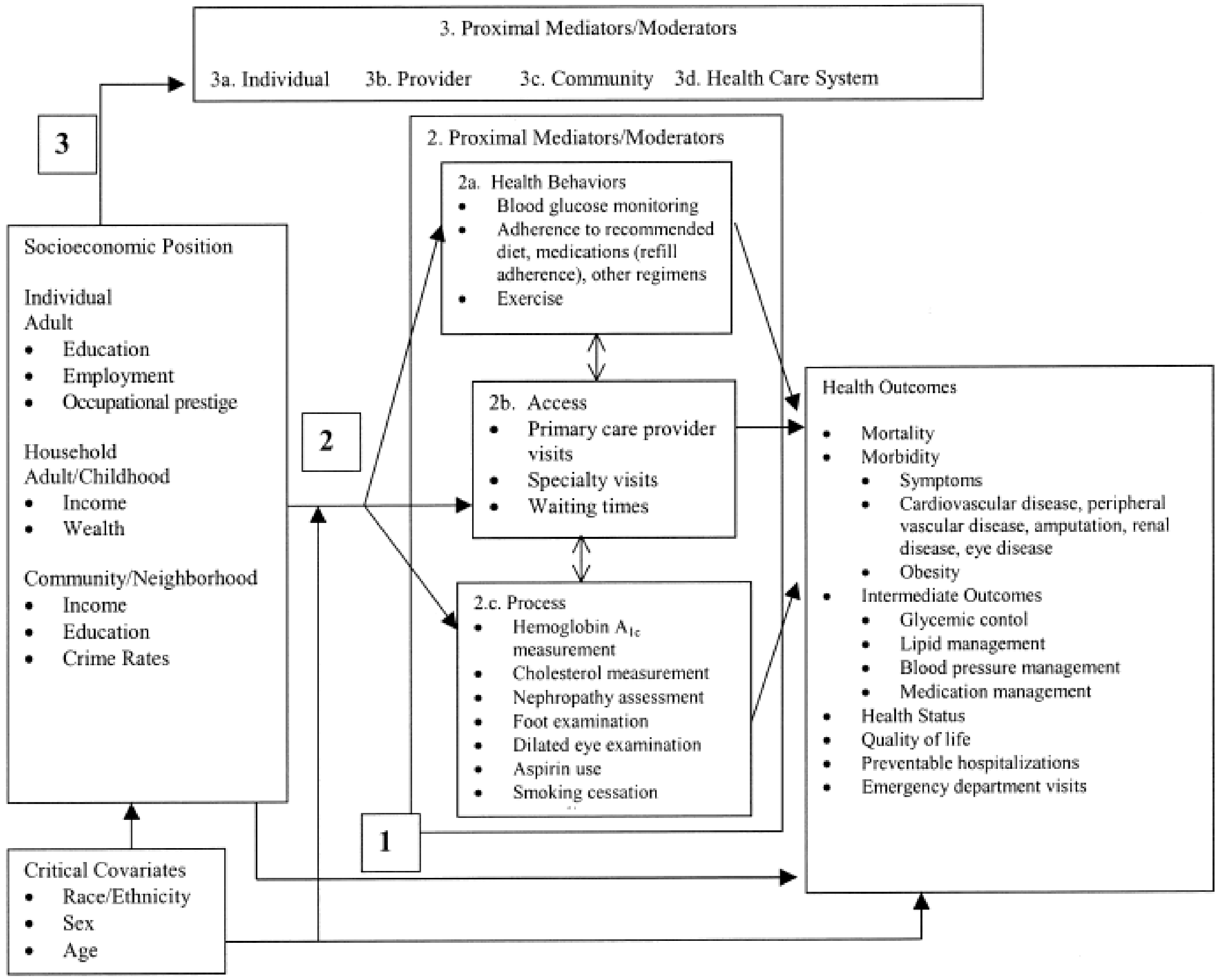
[5] Distal factors, also known as "upstream" factors, influence health through proximal ("downstream") factors (40).

ability to monitor blood glucose may be dependent on SEP whereby individuals from a low SEP (who often have low levels of health literacy[6] due to low education levels) may not be able to follow medical instructions (39).

As seen in pathway 2a of Figure 2.1, health behaviours such as diet, exercise, and adherence to medical advice can affect health outcomes (39). Studies have found that individuals from a low SEP are less physically active, more likely to smoke, and less likely to monitor their blood glucose, which can result in further morbidity, than individuals from a higher SEP (39). In pathway 2b of Figure 2.1, access to health care is an issue for residents who live in areas with fewer and poorer quality health facilities and physicians (39). Finally, pathway 2c of Figure 2.1 depicts the type of care provided by health professionals to the patient. Previous studies have found that individuals from a low SEP receive poorer care than individuals from a higher SEP (39).

[6] Health literacy refers to the ability to read, obtain, and understand medical information, which can allow or hinder an individual from engaging in health-promoting activities to improve health (39).

Figure 2.1: Framework on Socioeconomic Position (SEP) and Diabetes (39).



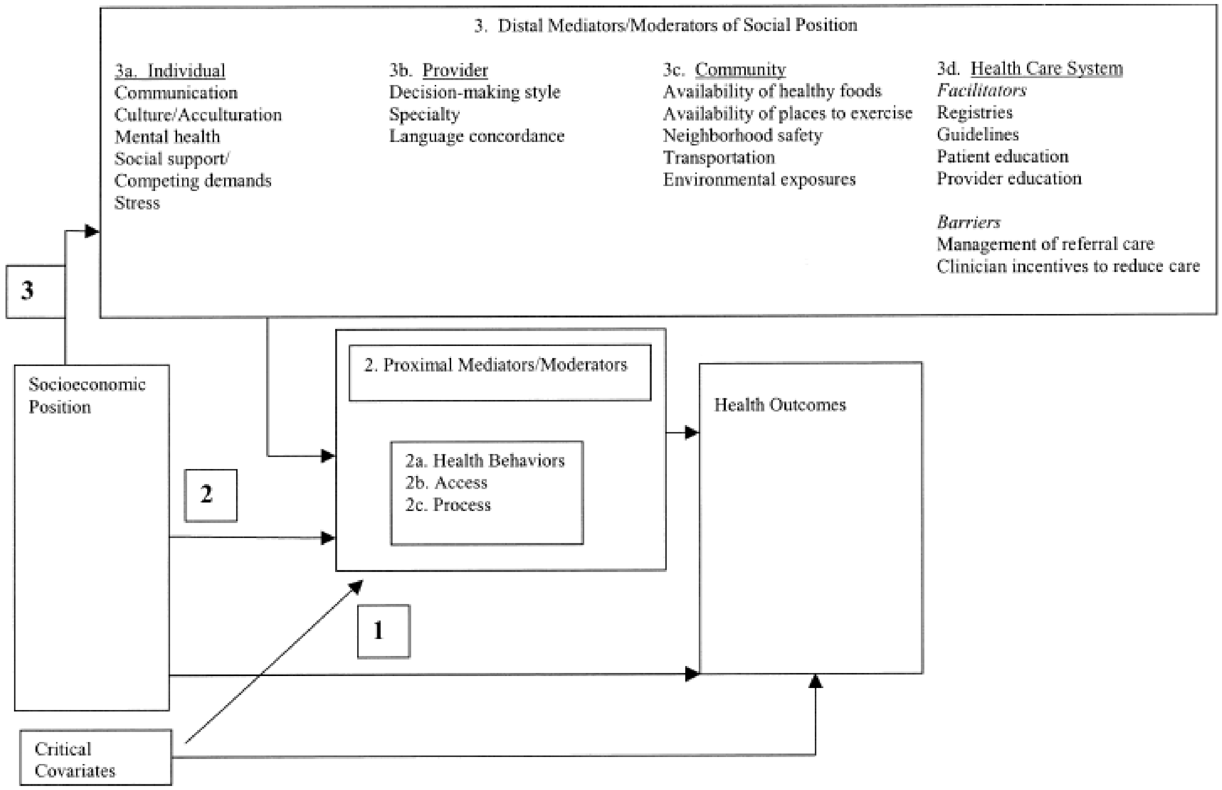
Source: Brown AF, Ettner SL, Piette J, Weinberger M, Gregg E, Shapiro MF, et al. Socioeconomic Position and Health among Persons with Diabetes Mellitus: A Conceptual Framework and Review of the Literature. Epidemiologic Reviews 2004;26:63-77.

Several distal mediators play an important role in health, however the focus of the thesis will center on the individual, the healthcare provider, and the community as seen in pathway 3 of Figure 2.2. At the individual level, barriers that hinder disease management include the patient-provider relationship, language differences, cultural beliefs, mental health, social supports, and stress (39). Communication between a patient and their healthcare provider, as seen in pathway 3a and 3b, can pose a barrier to health if the two parties do not agree on the course of treatment or if the physician takes on a more assertive approach without facilitating patient input, which often occurs for individuals of low SEP (39). Physicians' perceptions on the SEP of patients can also influence the physician's approach and quality of medical advice given to patients (39). Social supports, which can facilitate or hinder healthy behaviours, are often fewer for individuals of low SEP (39). Finally, stress, which is higher in low SEP individuals, can affect glucose levels in individuals through hormonal pathways or via biological processes of allostatic load[7] (39).

At the community level, people of lower SES may rely more heavily on public transportation to travel to health facilities, which may affect their ability to seek treatment, while wealthier individuals are more likely to travel by car (39). Within communities, the composition of individuals as well as the availability and accessibility of health-promoting outlets such as healthy food stores and recreational space for exercise have an impact on health (39). For example, crime rates and the quality of social relations are dependent on the types of individuals living in the community (39). In addition, poor communities may have food stores that are inaccessible, foods that are higher priced, and fewer healthier options compared to wealthier communities (39).

[7] Allostatic load refers to the chronic stress of the body, in which the body becomes unadaptable, leading to wear and tear via biological processes (39).

Figure 2.2: Detailed version on Relationship between Distal Mediators/Moderators and Diabetes (39).



Source: Brown AF, Ettner SL, Piette J, Weinberger M, Gregg E, Shapiro MF, et al. Socioeconomic Position and Health among Persons with Diabetes Mellitus: A Conceptual Framework and Review of the Literature. Epidemiologic Reviews 2004;26:63-77.

2.2 Literature Review

2.2.1 Diabetes

Diabetes, also known as diabetes mellitus, is a growing epidemic in Canada in which the prevalence has nearly doubled from 2000 to 2010, yet approximately 20% of Canadians who have diabetes remain undiagnosed (1,2). As of 2020, about one in three Canadians (4.2 million) are expected to be living with diabetes, prediabetes, and undiagnosed diabetes (2). Saskatchewan is one of the provinces in Canada to have a growing incidence rate of diabetes over the course of a decade from 1998/1999 to 2008/2009 (1). The number of prevalent cases in Saskatchewan was 80,000 in 2012 and it is expected to increase to 111,000 by 2020 (41).

Diabetes develops when the pancreas is unable to properly produce insulin, a hormone to control blood sugar, or when the cells of the body are unable to uptake insulin efficiently (1). Hyperglycemia, which results from high blood sugar levels (1), can lead to short-term effects like fatigue, weight loss, thirst, and frequent urination (42). The long-term effects of hyperglycemia include damage to blood vessels, nerves, kidneys, eyes, and heart (1,42). As a result, complications can arise such as lower limb amputations, blindness, and cardiovascular disease (CVD) (1,42). Severe damage to blood vessels combined with complications can result in mortality (1).

The different types of diabetes include: type 1diabetes, type 2 diabetes, and gestational diabetes (1). Type 1 diabetes, also known as insulin-dependent diabetes or juvenile diabetes, occurs in children or young adults and it develops when the autoimmune system attacks the pancreas cells that produce insulin (1,43) . Since people with type 1 diabetes are unable to produce insulin, they are dependent on other sources such as medication (1). Type 2 diabetes, also known as non-insulin-dependent diabetes or adult-onset diabetes, is the more common type of diabetes that often appears after the age of 40 years (1,43) . It often develops in individuals who are obese, physically inactive, and of certain ethnic origin (1). Finally, gestational diabetes is temporary diabetes that usually begins during the 24th week of pregnancy and lasts until delivery. Approximately 40% of women who have gestational diabetes develop type 2 diabetes in the future (1,43). Another condition related to diabetes is prediabetes, which occurs when blood sugar levels are high, but not high enough to be considered diabetes. If blood sugar levels are not controlled, prediabetes can lead to type 2 diabetes (1).

2.2.1.a Financial Burden

An enormous financial burden is placed on individuals with diabetes, their families, health care providers, the health care system, and the economy (2,44). Some Canadians pay out-of-pocket for diabetes care expenses and the situation becomes inequitable for low-income earners like seniors who are ineligible for social assistance and for individuals who are not fully covered by their public and/or private health plans (2). In Saskatchewan, the cost of diabetes on the health care system was $447 million in 2012 and the cost in 2020 is expected to reach $532 million (41). Approximately 18% of the total cost is directly related to diabetes care such as hospitalizations and medication, while 82% are indirect costs resulting from morbidity and mortality (45). The high costs of the health care system can be attributed to an increasing population, increasing age, lower physical activity levels, lower mortality rates, and higher incidence rates (44,45). The decrease in the number of diabetes cases (as a result of healthy lifestyle choices) can result in a decrease of healthcare costs (2,43). According to the Canadian Diabetes Association, a 2% decrease of prevalent cases would lead to a 9% cost decrease (2).

2.2.1.b Risk Factors and Barriers to Health Equity

A variety of genetic, behavioural, and environmental factors contribute to diabetes development (44). Risk factors for diabetes can be categorized as either modifiable or non-modifiable. Modifiable factors for type 2 and gestational diabetes can be modified through lifestyle choices, which can reduce one's risk of developing diabetes (44). Contrarily, non-modifiable risk factors like family history and ethnicity are inherent to certain individuals, predisposing them to negative health outcomes (2). For the purposes of this study, relevant risk factors pertaining to the study are discussed below.

Risk factors for diabetes that can be modified through health behaviours are body mass index (BMI) and physical activity (44). As lifestyle habits are changing and as fast foods are becoming popular, high levels of BMI have led to obesity, which has become a huge health concern in North America (46). Fast food restaurants and few recreational facilities are characteristic of obeseogenic environments and low SES neighbourhoods (29). Individuals of low SES are also less likely to have the knowledge and income necessary to purchase healthy foods (29). In addition, perceptions of obesity may also not be an issue (29). Physical activity, which is inversely related to BMI, is often higher among high SES individuals than low SES individuals (29), despite low income individuals walking as a form of transportation (46).

Non-modifiable risk factors of diabetes include age, sex, and ethnicity (44). The risk of developing diabetes increases with increasing age, becoming most prevalent after the age of 40 years (1). As the body ages, it becomes more difficult for the pancreas and cells to produce and uptake insulin, respectively (1). Gender is another inherent risk factor in which males have a higher prevalence of diabetes than females in Canada (1,47). Particular ethnic groups like Aboriginal Peoples, South Asians, Africans, and Latin Americans are also at greater risk of developing diabetes than other ethnicities due to a genetic predisposition and low physical activity levels (1,2).

Aboriginal Peoples in Saskatchewan, who make up the second highest concentration of Aboriginal Peoples in Canada, have higher rates of diabetes (type 1, type 2, and gestational) compared to non-Aboriginal Canadians (1,2). Children of women with gestational diabetes are more likely to be obese, insulin resistant, and prediabetic (48). Although the Aboriginal community are younger than the non-Aboriginal population, they still generally have poorer health compared to their counterparts (1,2). The health inequalities that exist between Aboriginals and non-Aboriginal Canadians can be partially attributed to socioeconomic inequalities (43). Risk factors for diabetes among Aboriginals include family history, increased sedentary lifestyle, obesity, consumption of processed foods, diminished access and availability to safe water, and inadequate housing (2). The Aboriginal population face additional barriers specific to rural areas as discussed in the following section (43).

2.2.1.c Urban/Rural Health Inequalities

Over half of the Aboriginal population live on reserves in Saskatchewan and many of them face additional challenges specific to rural areas (43,49) . Rural residents often have lower personal incomes, lower education levels, and higher unemployment rates than their urban counterparts, which is generally associated with negative health outcomes like diabetes (50). In addition to low socioeconomic conditions, individuals living in rural areas have lower physical activity levels, lower consumption of fruits and vegetables, and higher smoking rates compared to individuals from urban areas (51). Contrary to literature, no substantial urban-rural health disparities were found in the Saskatoon Health Region (51).

Another challenge experienced by rural individuals is that many working age individuals migrate from rural communities to urban centers for greater education and employment prospects (50). The resulting older cohort living in rural areas are at a greater risk of diabetes (due to older age) than individuals living in urban areas (50). Rural residents also have difficulty accessing specialized healthcare in their communities, which is partially attributed to a declining rural population (50). As a result, rural residents must travel to urban centers to access specialized health services, which is influenced by their ability to access a vehicle for travel (50). Those unable to own a vehicle may compromise their health by not accessing health facilities. It is important to address rural inequalities, as 40% of the Saskatchewan population live in rural communities (49,51). In addition, rural residents tend to have a high sense of social belonging in their community, which may mitigate negative health outcomes (51). The effects of social belonging and community networks are described in the section 2.2.2.a *Plausible Mechanisms*.

2.2.2 The Relationship between Income and Health

It is well-established that SEP indicators like income, education, and occupation influence health across social status (7). Income is positively correlated with education and occupation status as individuals with higher levels of education are more likely to have a higher employment status and higher income (52). Although material circumstances can be approximated using education and occupation, income would be a better indicator of material deprivation because (absolute) income directly measures the material purchase of health-enhancing resources like food and health services, which can affect health (7). The relationship between income and health is a positive "dose-response" one in which health status increases as income increases, but levels off after a certain point (7).

Two competing theories exist as to how income influences health: the psychosocial environment (relative income) interpretation and the neo-material (absolute income) interpretation (53). The psychosocial environment explanation suggests that an individual's perception of their social position, based on income level, relative to others (as the reference group) affects their health status (53,54). The consequences of these social comparisons are explained in the following section. In contrast to the psychosocial environment theory, the neo-material interpretation states that the combination of negative exposures, lack of resources, low material investment in health, and restrictive societal structures resulting from low material circumstances (i.e., income) can lead to poor health (53). This hypothesis can be explained by the collective resources model and the social inequality model developed by Stafford and Marmot (2003) (55). Based on the collective resources model, poorer neighbourhoods are more likely to have fewer and poorer quality resources than rich neighbourhoods (55). The social inequality model states that individuals living in less egalitarian neighbourhoods are more likely to have poor health compared to egalitarian neighbourhoods (55). Regardless, Kawachi and Berkman (2003) suggested that both psychosocial and material mechanisms can explain the relationship between income and health (32).

Although the relationship between health and absolute income is well-established, controversy exists regarding the effect of relative income on health (54,56,57). The inconsistent findings between relative income and health can be a result of researchers trying to find associations in small areas (57). For example, income inequality can be measured in small areas, however researchers often find weak associations between income and health (due residential segregation and homogeneity among residents) compared to larger areas (where greater homogeneity among residents exists). The weaker associations in small areas are often misinterpreted as the result of absolute income rather than income relative to the broader society outside of these small areas (57). Therefore, no specific income measure can be categorized as either an absolute income measure or a relative income measure because such as measure would depend on the context in which it is used (57).

2.2.2.a Plausible Mechanisms

One plausible mechanism acting on the relationship between income inequality and health is human capital, which is the investment of educational opportunities leading to life opportunities (58). Individuals from affluent neighbourhoods often have greater levels of human capital than individuals from poor neighbourhoods (59). As the rich pay larger taxes than the poor, the amount of public spending on education is greater in affluent neighbourhoods than in poor neighbourhoods (58). Education, which forms one part of human capital, is associated with social support, which occurs at an individual (compositional) level, and social capital, which occurs at a societal (contextual) level (59). Social capital, which is the amount and quality of social relationships based on trust and reciprocity for mutual gain, along with social support are interrelated concepts (59-61). There is a certain level of trust and reciprocity between neighbours in a community if one neighbour borrows another neighbour's car to get to a doctor's appointment, for example (61). The level of trust in a neighbourhood, especially in low income neighbourhoods, may be low if the probability of returning that borrowed car, for example, is low (61).

At the individual level, educated people are more likely to have strong social networks and engage in community organizations compared to less educated people (59). At the community level, high levels of education can lead to healthier behaviours, dissemination of health information, and stronger social supports compared to communities with low levels of education (59). As education is positively correlated with income (52), individuals living in low income areas often have low levels of social support and social capital (informational, emotional, and financial, etc.) in addition to higher mortality rates and poorer psychological health than individuals living in higher income areas with higher levels of social support (59,60,62).

The most widely speculated mechanism acting on the income inequality and health relationship is social capital acting through the "social cohesion and collective social pathway" in which communities with higher income inequality have lower levels of social capital than communities with lower income inequality (58). At the neighborhood level, communities with high levels of social ties may be proactive in deterring unhealthy behaviours like smoking and fighting crime compared to socially isolated communities (60,61). In addition, socially integrated neighbourhoods are likely to have organized social networks ensuring the availability and accessibility of social programs and services. Socially cohesive communities are also more likely to exert effective emotional support, fostering self-esteem and respect (60). At the national level, individuals participating in social organizations like churches and unions develop communication and organization skills that they can use for political engagements like voting for political candidates that will create better opportunities for the disadvantaged, for example (61).

Finally, social comparisons arising from income inequality can lead to frustrations and direct psychosocial and physiological illnesses (58). Wilkinson (1999) stated that the three psychological risk factors mediating the relationship between income inequality and health are low control, income insecurity, and low self-esteem (63). Individuals living in areas characterized by high control, income security, and high self-esteem were more likely to be socially integrated than people living in areas characterized by opposing features (63). Areas characterized by low control, income insecurity, and low self-esteem are likely to have more stressors of greater severity and duration than areas with high control, income security, and high self-esteem (63). Individuals from low SES neighbourhoods experience higher levels of stress for various reasons such as poor quality housing, discrimination, financial insecurity, and poor social ties (64). The accumulation of stress over time, also known as "allostatic load," can lead to negative physiological consequences due to the inability of the body to restore homeostasis (64). One of those stressors such as income security is considered by many to be a source of respect and if not met, individuals can resort to violence in order to suppress feelings of shame and inferiority (57,63).

Overall, the breakdown of social support and social capital is multi-factorial (61). Compared to the past, income inequality between the rich and the poor continues to widen, forcing the disadvantaged to work harder and longer hours to maintain their social status (61). In addition, the emergence of dual-career households has led to less time being devoted to social relationships, which can cause the breakdown of marriages and friendships (61).

2.2.3 Individual versus Area-based Socioeconomic Measures (ABSMs)

There has been debate throughout the literature of whether area-level measures are adequate proxies for individual data (27,33,35). Some researchers have found poor agreement between area-level and individual-based SES measures, while other researchers believe area-based SES can approximate for individual-level SES (27,33-37). Due to the absence of individual-level SES in administrative databases, researchers are necessitated to use area-level measures. Area-based socioeconomic measures (ABSMs) have been compared to individual-level SES for several health outcomes (65), however caution should be exercised when utilizing ABSMs in measuring health inequalities.

2.2.3.a Utilizing ABSMs with Caution

Several issues arise when using ABSMs to measure socioeconomic inequalities such as the modifiable areal unit problem (MAUP), ecological and atomistic fallacies, and the choice of spatial unit. The MAUP arises when different boundaries placed on the same data lead to different results (15). Two aspects of the MAUP are the scale effect and the zonation effect. The scale effect occurs when different results are produced every time the spatial scale (i.e., dissemination areas, census tracts, etc.) changes in size (15). The zoning effect refers to the production of different results from changing the boundaries at a particular scale (15). However, researchers are limited in their ability of changing boundaries due to pre-defined administrative boundaries and the utilization of these areal units in other data sources like the census (15). Overall, the MAUP arises when inferences are made based on data from a different scale (15). For example, the ecological fallacy occurs when inferences on individuals are made from area-level data (15,66). Contrarily, the atomistic fallacy is when group-level inferences are made based on individual-level data (66). These fallacies can be prevented if the level of the data and the level at which the researcher would like to make inferences matches, in the case of this study (66). Although some studies use ABSMs as "proxies" for individual-level SES, it should also be noted that the two measures are different from each other as ABSMs have their own independent contextual effect on health beyond individual factors (22,35,67,68).

Since aggregate data is available at pre-defined spatial units to protect individual identities, these administrative boundaries may not accurately reflect the social, economic, political, and cultural processes influencing health (25). Processes occurring at the neighbourhood-level or at a macro-level do not necessarily coincide with the boundaries set by political authorities because certain processes may only act in particular locations and some processes may be more influential in some parts of society than others (25). In addition, areas that are pre-defined may not accurately reflect the composition of the residents (69). For example, data at the area-level may disregard pockets of disadvantaged residents and may not accurately reflect the negative health outcomes experienced in that particular area (70).

Therefore, the selection of an ABSM spatial unit should be small enough to encompass homogeneity among individuals in that context (65,67). As the size of the spatial unit increases, measurement error increases due to an ABSM approximating for individual characteristics as the heterogeneity of the residents living in that area increases (15,34). Studies have shown smaller spatial units provide less error and larger socioeconomic gradients than bigger spatial units (22,28,35,65,67).

In addition, there is no clear consensus as to which area-level socioeconomic measures best capture health inequalities (22). Krieger et al. (2003) provide criteria for assessing suitable ABSMs such as whether socioeconomic gradients have been established from past research, the gradients occur for several health outcomes, there is limited missing data, and the socioeconomic indicator is easy to interpret (22). Deprivation indices such as the Pampalon deprivation index were created to capture health inequalities using more than one socioeconomic indicator (3). However, the size of socioeconomic gradients depends on the socioeconomic indicator, the size of the spatial unit, and the health outcome (22-24). Since various studies have used different kinds of ABSMs, it has become difficult to compare studies over time and across places (23).

2.2.4 Contextual Effects and Health

Area effects, also known as "place effects" or "contextual effects" in this paper, are regarded as exposures that influence health independent of individual characteristics, however no single established definition exists. Macintyre et al. (2002) defined place effects as "an unspecified black box of mystical influences that remains after investigators have controlled for a range of place characteristics" (71). Although place effects occur at any spatial scale, researchers often identify a small spatial unit, incorporating some degree of homogeneity and heterogeneity among the residents, which is representative of a neighbourhood (34). Different combinations of homogeneity and heterogeneity create neighbourhoods that are different from each other (72). Dissemination areas are small area-level units that maximize homogeneity among the residents, but they are big enough to incorporate some degree of heterogeneity and encompass the social, political, and economic processes that influence a neighbourhood (34).

The term, *neighbourhood*, was referred to by Lebel and colleagues (2007) as a place where people carry out social and economic activities like going to work, caring for family, and shopping (72). A neighbourhood is a place that people are familiar with and it can be a reflection of one’s identity, social values, dreams and goals, and socioeconomic environment (72). Features that should be considered when identifying a spatial unit are the inner characteristics of the neighbourhood and the geographical scale (72). The inner characteristics of a neighbourhood are characterized by homogeneity and heterogeneity of the elements within a neighbourhood like the residents, structural features, demographic features, political features, and sentimental features, for example (72,73). The geographic scale is also important because a neighbourhood can be defined in several ways (72). For example, spatial levels of a typical neighbourhood may consist of the home area in which psycho-social conditions are strong, the locality whereby activities and still nearby and familiar, and the urban district which encompasses the social and economic processes and activities residents participate in (72).

There has been debate as to how spatial units should be defined (74). No single unit can measure all social, economic, or political processes of the environment, and administrative boundaries may not necessarily represent how residents define their neighbourhoods (74). O’Campo (2003) and Lebel et al. (2007) suggest using spatial units that encompass the perspectives of residents, historical boundaries, and socioeconomic status (72,74). However, readily available aggregate data used in the Canadian Census are based on administrative boundaries, which may not represent “neighbourhoods.”

In terms of measuring and conceptualising area effects, several methodological and conceptual issues persist (71). Firstly, contextual effects are difficult to measure due to the interaction between individual and population characteristics (71). It is difficult to determine if area differences are a result of context (of place features) or composition (of types of residents) (32,71). For example, poor health may be a result of low-income individuals eating unhealthy, but it can also be due to areas where grocery stores offer few healthy foods (32). Context and composition have been found to influence several health outcomes such as low birth weight (22), self-rated health (65), mortality (75), cardiovascular disease (17), and diabetes risk factors (17). For example, Gary et al. (2008) found that individuals from low SES neighbourhoods had higher smoking rates, blood pressure, and physical inactivity in addition to neighbourhood problems such as crime, litter, poor night lighting, and inaccessibility to transportation and recreational facilities compared to individuals from higher SES neighbourhoods (76).

2.2.4.a Relation to Diabetes

Although studies have shown an inverse relationship between area-level SES and diabetes or insulin resistance, the mechanisms behind this relationship has been seldom studied (77). Several hypotheses have been proposed for the inverse relationship between area-level SES and diabetes like neighbourhood access to health-promoting or health-limiting resources (77). Studies have shown low SES neighbourhoods have greater difficulty accessing health foods such as fruits and vegetables, while fast-food restaurants are more apparent in these neighbourhoods (77). Neighbourhoods that promote physical activity such as the availability of recreational facilities and neighbourhood green space can lower the risk of obesity and diabetes (77). However, neighbourhood perception is an important factor influencing diabetes development, diabetes management, and insulin resistance (76). Neighbourhood problems like crime may prevent diabetics from engaging in outdoor physical activity even though access to recreational facilities exist (76). Similarly, other characteristics such as public transportation, access to speciality healthcare, or stress may influence individuals with diabetes from engaging in health-related behaviours to lower blood glucose, lipids, blood pressure, and adherence to medication (39,76,77).

CHAPTER 3:
METHODOLOGY

Chapter 3 describes how the research study was conducted. This chapter discusses the data sources, data collection methods, data, analysis techniques, ethics, confidentiality, and limitations. Several data sources were used and data analysis was performed on the 'unweighted' and 'weighted' population as described below.

3.1 Data Sources

The data sources used for the study include health administrative data, the 2006 Canadian Census, and the Canadian Community Health Survey (CCHS). Table 3.1 below describes the purpose of each data source and the variables used for the study. From the administrative data, the hospital discharge abstract database (DAD) and the physician services claims file from the Medical Services Branch (MSB) were used to extract people meeting the Canadian Chronic Disease Surveillance System (CCDSS) case definition for diabetes. The Person Registry System (PRS) database was used to obtain demographic information such as sex and date of birth (birth month and birth year). The CCHS was used to obtain individual-level data, while the 2006 Canadian Census contained average income at the dissemination area (DA) level.

3.1.1 Hospital Discharge Abstract Database (DAD)

The hospital discharge abstract database (DAD) was one source used to extract diabetes cases. The database contains hospital admissions dating back from the 2001/2002 fiscal year (78). The DAD consists of hospital discharges by fiscal year on in-patient separations of acute care, day surgeries, and hospital psychiatric separations (79). The hospital database also contains some hospitalizations of Saskatchewan residents treated out-of-province, however not all records are transferred back to Saskatchewan Health (78).

The variables used from the DAD for the study include the unique person identifier, also known as the "key_hsn," date of admission, date of discharge, and diagnosis codes. The date of admission was used to include individuals who were admitted during the study period of January

1, 2001 to December 31, 2008. The date of discharge was used as a point of separation to count the number of hospital discharges that meet the CCDSS case definition of one or more hospitalizations. The diagnosis codes (ICD-9: 250/ ICD-10-CA: E10, E11, E12, E14) for diabetes were used to extract cases (80).

3.1.2 Medical Services Branch (MSB) Database

The physician services claims file, also known as the medical services branch (MSB) data, was also used to extract diabetes cases (78,79). Service claims from this database date back to the 1998/1999 fiscal year (78). The MSB contains service fee claims made by physicians to the federal government (79). One downfall of the database is that services made by salaried doctors are not included (78).

The variables used from the MSB for the study include the unique person identifier, date of service, and diagnosis codes. The date of service was used to include individuals who were seen by a physician during the study period. It was also used as a point of separation to count the number of physician visits that meet the CCDSS case definition of two or more physician claims (80).

3.1.3 Person Registry System (PRS) Database

The Person Registry System (PRS) is a database that tracks all Saskatchewan residents who have a health card (78). Based on the health card application, demographic and some geographic data from the PRS comes from the form filled out by residents (78). The database contains all past and present residents with their period of residency in Saskatchewan (78).

The variables used from the PRS for the study include the unique person identifier, birth year, and birth month. Since day of birth is not provided in the PRS due to confidentiality reasons, the 15th day was used as a middle ground to estimate the day of birth. Age was calculated on the admission date for all diabetes cases by merging the PRS to the diabetes cohort. The age and sex variables from the PRS were only used to meet the inclusion/exclusion criteria, but these variables were not used for data analysis. The age and sex variables from the CCHS were used instead, due to the study population coming from the CCHS.

3.1.4 Canadian Census (2006)

The Canadian Census is a questionnaire that collects information on demographics, ethnicity, culture, and socioeconomic status on individuals from Canada (81). Some parts of the census consist of responses from the entire population, while other data comes from a 20% sample (81). Individuals included in the census are Canadian citizens, landed immigrants with a Canadian residence, landed immigrants on military or diplomatic duty outside of Canada, landed immigrants on marine duty outside of Canada, and people holding a refugee status, study permit, or work permit with a Canadian residence (81). The 2006 Census excludes government personnel outside of Canada, members of the military outside of Canada, and visitors to Canada without a permit (81).

A Saskatchewan cumulative profile was obtained from the 2006 Census. The variables used from the profile include average income and the DA identification number. Incomes are suppressed by Statistics Canada if there are fewer than 250 people or fewer than 40 private households in a dissemination area in order to maintain confidentiality (82).

3.1.5 Canadian Community Health Survey (CCHS)

3.1.5.a Content Structure

The CCHS is a cross-sectional survey consisting of several components: the common content, the optional content, and the rapid response content (4). The common content are questions asked by all respondents from all provinces and territories. The common content consists of two parts: the core content and the theme content (4). The core content consists of the same questions asked in each cycle for about six years, while the theme content consists of questions on a particular topic. The other component of the CCHS, the optional content, consists of a series of modules that health regions and provinces can select to ask their residents to address region-specific, healthcare needs (4). Finally, the rapid response content, which takes less than two minutes, consists of questions on a particular topic asked to all respondents (4).

3.1.5.b Target Population

The CCHS consists of people aged 12 years and older across Canada (83). Each person sampled represents several people in the population who are not in the sample. People with a variety of characteristics are sampled in order to represent everyone in the population (83). People excluded from the survey include people living on reserves, the Canadian Forces, people living in institutions, and people from Region du Nunavik and Region des Terres-Cries-de-la-Baie-James of Quebec (83). In total, the excluded groups represent approximately 3% of the Canadian population (83).

3.1.5.c Sampling Size and Strategy

The sampling strategy used in the CCHS is a complex multistage one in which approximately 65,000 individuals are sampled to provide reliable estimates of the population at the health region and provincial level (4). In order for the health regions and the provinces to have equal importance, the sample consists of a minimum of 500 individuals from each health region and then a probability proportional to size method (PPS) based on the population size of each province is used (4). The sample size based on province population is then distributed based on the square root of the health region population (4). The entire sample is divided equally for desired sample sizes from the area frame and list frame sampling techniques, which are described below (4). However, a greater number of respondents are sampled to account for non-response and individuals aged 12 to 14 years (who may not have been covered) in the survey (4).

Three random sampling techniques based on PPS are used whereby 49% of respondents come from the area frame, another 50% of respondents come from the list frame, and 1% of respondents come from the random digit dialling (RDD) frame (4). For the area frame, the sample is ensured every six months that it covers all strata representative of the Canadian population, while the list frame sample is checked every two months (4).

In the area frame, three types of regions are selected in each province: urban cities, other cities, and rural areas (4). Within large urban cities, stratum are selected based on geographical or socioeconomic criteria to account for population groups of low number like Aboriginal Peoples and immigrants (4). For the other cities and rural areas, strata are selected based on geographical characteristics and then socioeconomic characteristics (4). Each stratum is then divided into approximately 6 clusters using the probability proportional to size (PPS) method corresponding to the number of households (4).

In the list frame, a list of telephone numbers for each health region is selected based on a simple random sampling method (4). Telephone numbers, which come from directories, are assigned a health region to supplement the area frame sample (4). Since some households may have expired or disconnected telephone numbers, there is under-sampling using the list frame, however Statistics Canada ensures that the use of sampling weights resolves the undercoverage issue (4).

Finally, the random digit dialling (RDD) frame is used as the only frame for 1% of the respondents (4). A list of working and non-working banks is created and using the Elimination of Non-Working Banks (ENWB) technique, banks in each strata that do not have a residential telephone number from a list of one hundred telephone numbers are eliminated (4). The resulting banks are grouped within each health region and a bank is randomly selected. A telephone number is then randomly dialled from each cluster (4).

3.1.5.d Weighting

Based on the sampling strategy above, young individuals aged 12 to 19 years are over-sampled in the CCHS to avoid final sampling weights that are extreme (4). In all households, youth are more likely to be sampled except for households containing more than 5 members or if the number of youth in the household is greater than 2 (4). Despite an over-sampling of youth in the CCHS, an under-representation of youth still persists, which may be due to a combination of factors such as greater cellphone use rather than landline use, unapproachable during interview times, and lack of interest in completing the survey.

The weighting strategy used in the CCHS employs two types of weights: sampling weights and bootstrap weights (4,83). Sampling weights account for the unequal selection of respondents, while the bootstrap weights take into consideration the complex multistage nature of the sample (4). A sampling weight is assigned to each respondent as they represent several other individuals in the population (4). One respondent in the study sample represents approximately 50 individuals in the population. Of the sampling weights, there are two types: household weights and final weights (4). Household weights are based on the area frame and the telephone frame sampling techniques separately (4); however, households weights were not used in this study. Contrarily, final sampling weights are used for population estimates and they are obtained by integrating the area frame weights and the telephone frame weights into a single weight (4). For non-respondents who refused to answer the survey for any reason, weights were re-distributed to respondents who had similar characteristics based on probabilities (4).

Apart from sampling weights, bootstrap weights are used to obtain reliable variance estimates due to the complex sampling design of the CCHS (83). The bootstrap method involves re-sampling (n-1) clusters whereby ‘n’ represents the number of clusters (83). The re-sampling technique randomly selects a random sample with replacement and calculates the weight each time (83). The final bootstrap weights are calculated by stratifying across demographic characteristics (83). The re-sampling and stratifying process is usually repeated 500 times to produce 500 bootstrap weights (83). The Bootvar program, which was developed by Statistics Canada, allows for variance calculation (83).

3.1.5.e CCHS Files

The CCHS files used for the study were the share files in which access to administrative data was granted from respondents who gave permission to share their information (4). Based on the share files, there were two types of CCHS files: the CCHS link file and the CCHS cycle file as seen in Table 3.1. The CCHS link file is used to link the administrative data to the CCHS cycle file. (The descriptions of the data linkages can be found in section 3.3.1 *Data Linkages*).The link file contains 41,468 Saskatchewan respondents from all cycles (cycle 1.1 to cycle 9.1). The variables used from the link file in the study include the household identifier, household member identifier, and unique person identifier.

The CCHS cycle file used in this study is the 4.1 and 5.1 combined cycle from the years 2007 and 2008 with the questions and responses in dataset format. The variables used from the CCHS cycle for the study include the household identifier, household member identifier, age, sex, dissemination area, total household income (individual-level income), employment, education, BMI, Aboriginal identity, physical activity, and marital status.

Table 3.1: Summary of Data Sources Used in the Study

Data Source	Details
Administrative Databases	
Hospital Discharge Abstract Database (DAD)	Identifies people who visited a hospital. The DAD database was used to extract people meeting the diabetes algorithm of having at least one hospital visit for diabetes. The key variables used from the DAD are "key_hsn," date of admission, date of discharge, and diagnosis codes.
Physician Services Claims File: Medical Services Branch (MSB)	Identifies people who visited a physician at a health clinic. The MSB database was used to extract people meeting the diabetes algorithm of having two or more physician visits within two years for diabetes. The key variables used from the MSB are "key_hsn," date of service, and diagnosis codes.
Person Registry System (PRS)	Contains demographic and some geographical information. Data extracted from this database included sex and date of birth (month and year). Age and sex from the PRS was only used to identify the diabetes cohort based on the CCDSS definition. The variables of age and sex from the census were used instead for data analysis.
2006 Canadian Census Saskatchewan Cumulative Profile	Counts on demographic, ethnic, and socioeconomic characteristics are collected on individuals at the dissemination area geography. Variables used from the census were average income (area-level income) and the identification number for each dissemination area in Saskatchewan.
Canadian Community Health Survey (CCHS)	
CCHS Link File (cycles 1.1 to 9.1)	The CCHS link file contains a linking variable (sampleid), which is the household identification number, which is used to merge the CCHS link file to the CCHS cycle file. The link file contains another linking variable (key_hsn), which is used to link the CCHS cycle file to the health administrative data.

CCHS Cycle File (cycle 4.1 & 5.1 combined)	The CCHS cycle file contains variables representing a specific CCHS question and the corresponding responses coded as values. In this study, the variables of interest are: dissemination area, age, sex, total household income (individual-level income), respondent education level (education), working status last week (employment), physical activity level, BMI, and Aboriginal identity.

3.2 Data Collection

3.2.1 CCHS, Census, Administrative Data

Statistics Canada oversees the collection of data for the Canadian Census and the CCHS. For the Census, data collection occurs every five years and all Canadians are required by law to respond to the survey (84). For the CCHS, data collection occurred every year with the survey being released every two years prior to 2005 (4). After 2008, the CCHS began releasing data every year with data collection occurring every two months. In addition, combined files consisting of two cycles are released every two years (4).

CCHS data was collected via computer assisted interviewing (CAI) in which trained interviewers asked the survey questions and input the responses provided by interviewees (4). The CAI method minimizes data entry errors by disallowing out-of-range responses and questions not applicable to the respondent are skipped automatically (4). Two types of computer assisted interviewing were employed: computer assisted personal interviewing (CAPI) and computer assisted telephone interviewing (CATI). Interviews from the area frame were conducted mostly in person using CAPI, while interviews from the list frame and the RDD were conducted by telephone through CATI (4).

For administrative databases, like the hospital discharge abstract database and the medical branch database, these data are submitted to the Ministry of Health by hospital and physician practices, respectively, and assembled in provincial databases by eHealth Saskatchewan on behalf of the Ministry of Health under terms of legislation and data sharing agreements with the relevant source data stewards (78). By means of a data sharing agreement between the Ministry of Health and the Health Quality Council (HQC), extracts of these data were made accessible for this research at HQC after eHealth had removed individual identifiers and encrypted health card numbers to maintain confidentiality (78).

3.2.2 Inclusion and Exclusion Criteria for Diabetes Cohort and Study Population

For the administrative data, the CCDSS definition for diabetes was used to define a case as a person who had at least one hospital discharge record from the hospital discharge abstract database with a diabetes diagnosis (ICD-9: 250, ICD-10-CA: E10, E11, E12, E14), and/or two or more physician service claims with a diabetes diagnosis within two years from the physician services claims database (80,85) . Women aged 10 to 54 who were diagnosed with diabetes 120 days before or 180 days after a gestational event were excluded from the cohort in order to remove women with gestational diabetes (80). Anyone under the age of 1 was also excluded from the cohort (80,85).

The study period for people meeting the case definition was January 1, 2001 to December 31, 2008, respectively. Prevalent cases were counted as existing and pre-existing cases for the year 2008. Cases that died prior to January 1, 2007 were excluded from the cohort. However, cases that died in 2007 (n=2,202) were included in the cohort because some of these individuals (who are diabetic) could have been surveyed in the 2007-2008 CCHS prior to their death. The study population was defined using the CCHS 4.1 and 5.1 combined cycle for the years 2007 to 2008. No individuals from the study population were excluded.

3.3 Data

3.3.1 Data Linkages

Figure A1 and Figure A2 in Appendix A illustrate the data linkages for combining the various data sources. First, the CCHS link file was merged to the CCHS cycle because the link file contained a unique person identifier, “key_hsn,” which was needed to link the CCHS cycle to the administrative data. Second, the resulting merged dataset was linked to the administrative data on the unique person identifier. Third, the diabetes cases were flagged and merged back to the original CCHS cycle. Finally, the census was merged to the CCHS via the DA identification number.

3.3.1.a CCHS Link File and CCHS Cycle 4.1 & 5.1 Combined

The combined cycle of 4.1 and 5.1 (n=7,339) was merged to the CCHS link file (n=41,468) on household identifier ("sampleid") and household member identifier ("personid"). The resulting "merged" dataset contained 6,797 individuals who were linked from the CCHS link file to the CCHS survey. Records with a missing person identifier were excluded (n=1,100), since the person identifier was required to link the CCHS to the administrative data. After exclusions, the final "merged" dataset contained 5,697 observations.

3.3.1.b CCHS Merged File and Administrative Data

The final CCHS "merged" file (n=5,697) was linked to the final diabetes cohort (n=70,722) on the household identifier and the household member identifier. There were 485 people in the diabetes cohort who were successfully matched to the CCHS. The 485 people were flagged as having diabetes and they were merged back to the CCHS cycle of 4.1 and 5.1 (n=7,339). The rest of the population from the CCHS were flagged as non-diabetic (n=6,854).

3.3.1.c CCHS and Canadian Census

After individuals were flagged as either diabetic or non-diabetic, average income (area-level income) from the 2006 Census was merged to the CCHS (n=7,339) via the dissemination area (DA) identification number. The final dataset (n=7,339) contained 1,180 DAs with 402 of those DAs having a missing area-level income.

3.3.2 Individual-level Variables

The individual-level covariates used for analyses are described in Table 3.2 below. The dependent variable of diabetes, which was a binary variable, was obtained from administrative data. All other individual-level covariates were categorical variables from the CCHS. Several variables were re-categorized such as age, total household income, education, employment, BMI for individuals aged 18 years and over, and marital status, while others remained unchanged.

The age variable was originally continuous and it was re-categorized into the following categories: <35, 35-44, 45-59, 60-79, 80+. Total household income, which represented individual-level income, was re-categorized into three categories in the same manner as area-level income to allow for comparisons and the releasability of data (due to sample size) *(see* section 3.6 *Confidentiality and Release Guidelines).* Household income, instead of personal income, was selected as the individual-level income variable since many young individuals and some family members may not have a separate source of income, yet benefit from the total household income.

Respondent education was re-categorized into the following categories: less than high school diploma, high school graduate, and post-secondary degree. Working status was re-categorized as 'employed' and 'unemployed.' In the CCHS, body mass index (BMI) was calculated separately for individuals younger than 18 years of age and for individuals aged 18 years and older. BMI was re-categorized into the following categories: neither overweight nor obese, overweight, and obese. For marital status, married and common-law individuals were categorized as 'partnered,' while the other CCHS categories remained the same. The variables of sex, Aboriginal identity, and physical activity remained unchanged.

Table 3.2: Characteristics of Individual-level Covariates

Variable (Concept)	**Variable Code**	**Original Values**	**Re-Categorized Values**	**New Variable Type**	**Source**
Diabetes mellitus	DISEASE	N/A	0=no diabetes 1=diabetes	Binary	Administrative Data
Age	DHH_AGE	12 to103	<35(reference) 35-44 45-59 60-79 80+	Categorical	CCHS
Sex	DHH_SEX	1=Male 2=Female (reference)	No change	Categorical	CCHS

Total household income (Individual-level income)	INCDHH	1=no income 2=<$15,000 3=$5,000-$9,999 4=$10,000-$14,999 5=$15,000-$19,999 6=$20,000-$29,999 7=$30,000-$39,999 8=$40,000-$49,999 9=$50,000-$59,999 10=$60,000-$79,999 11=$80,000-$99,999 12=$100,000+ 99=not stated	<$30,000 $30,000-$49,999 $50,000+(reference) Missing	Categorical	CCHS
Respondent education level (Education)	EDUDR04	1=<secondary school 2=secondary school grad, no post-secondary 3=some post-secondary 4=post-secondary degree 9=missing	1=<high school (reference) 2,3=high school grad 4=post-secondary degree	Categorical	CCHS
Working Status last week (Employment)	LBSDWSS	1=had a job –at work 2=had a job –absent 3=did not have a job 4=permanently unable to work 6,9=missing	1,2=employed (reference) 3,4=unemployed 6,9=missing	Categorical	CCHS
BMI	HWTDISW (age 18+)	1=underweight 2=normal weight 3=overweight 4=obese class I 5=obese class II 6= obese class III 96,99=missing	1,2=neither (reference) 3=overweight 4,5,6=obese 96,99=missing	Categorical	CCHS
	HWTDCOL (age 12-17)	1=neither (reference) 2=overweight 3=obese 6,9=missing	No change		
Marital Status	DHH_MS	1=married 2=common-law 3=widowed 4=separated 5=divorced 6=single (never married) 9=missing	1,2=partnered (reference) 3=widowed 4,5=separated 6=single 9=missing	Categorical	CCHS
Aboriginal Identity	SDCDABT	1=Aboriginal 2=Not Aboriginal (reference) 9=missing	No change	Categorical	CCHS

Leisure Physical Activity	PACDPAI	1=active (reference) 2=moderately active 3=inactive 9=not stated (missing)	No change	Categorical	CCHS

3.3.3 Area-level Variables

Area-level variables in this study included average income from the 2006 Canadian Census and geography type from the CCHS as seen in Table 3.3 below. Area-level income was represented by an average income assigned to each dissemination area. Average income at the dissemination area-level is based on total household income, which allows for the comparison between individual and area-level income (81).

'Geography type' was a categorical variable from the CCHS in which dissemination blocks were characterized as either urban or rural (86). Inconsistent definitions exist as to how rural areas should be defined (50). Some research studies use a gradient of "rural" in which areas are categorized as: strong metropolitan influenced zone (MIZ), moderate MIZ, weak MIZ, and no MIZ based on the proportion of commuters traveling from rural areas to urban centers (51). However, this study uses the definition used by Statistics Canada whereby urban areas consist of at least 1,000 individuals and 400 individuals per square kilometer within a dissemination block, while rural areas are those that are not urban (86). Dissemination blocks with mixed areas are categorized as either urban or rural based on composition (86). The administrative boundaries of dissemination areas used in the Census and CCHS are based on the year 2006 (4,86).

Table 3.3: Characteristics of Area-level Covariates

Variable	Variable Code	Original Values	Re-Categorized Values	New Variable Type	Source
Geography Type	GEODUR2	1=urban(reference) 2=rural	No change	Categorical	CCHS
Average income	AVG_INC	$2,446 to $95,023	<$30,000 $30,000-$49,999 $50,000+ (reference) Missing	Categorical	2006 Census

3.4 Data Analysis

Data analyses were conducted on the ‘unweighted’(study sample) and ‘weighted’ (estimated population) data. Sampling weights and bootstrap weights were not applied to the ‘unweighted’ population, while they were used on the ‘weighted’ population.

3.4.1 Modeling Strategy

Descriptive analysis was performed on the study sample and the estimated population. The percentage of individuals with diabetes was stratified by income (at the individual and area-levels) and by geography type. Income-specific proportions of diabetics for individual and area-level income were also stratified by geography type. Percentage agreement between individual and area-level income was measured using a weighted Kappa statistic of linear weights.

Multivariable analyses were conducted on the ‘unweighted’ and ‘weighted’ population using a backward elimination strategy. Two multivariable logistic regressions were produced: (1) a saturated model with all covariates, and (2) a final model of significant covariates (with a p-value less than 0.05 alpha). From the saturated models, covariates with a p-value greater than 0.05 alpha were eliminated for the individual and area-level models.

For the saturated models, the following interaction terms of income at the individual and area-levels were found to be insignificant and they were subsequently removed: income*geography type, income*education, income*Aboriginal identity, income*BMI. For all ‘unweighted’ and ‘weighted’ models, statistically significant covariates included age, sex, BMI, employment, physical activity, and Aboriginal identity. Although individual and area-level incomes were not statistically significant, they were included separately in each of the ‘unweighted’ and ‘weighted’ models due to their importance in the study.

3.4.2 Software and Logistics

Data analysis was conducted at the Health Quality Council (HQC) of Saskatchewan. The software used to analyze the CCHS share files and the administrative data was SAS Enterprise

Guide 6.1. Administrative data and the CCHS were made available at the Health Quality Council of Saskatchewan by means of a data sharing agreement between HQC and the Ministry of Health. The Saskatchewan Census community profile was provided by Statistics Canada and imported into HQC via designated HQC analysts who had access to both the internet and SAS Enterprise Guide.

3.5 Ethics

Ethical approval was sought from the University of Saskatchewan Research Ethics Board in conjunction with the Saskatoon Health Region. Approval for the study was also obtained from Statistics Canada, the Saskatchewan Health Quality Council, and the Saskatchewan Population Health and Evaluation Research Unit (SPHERU). Statistics Canada in conjunction with the Saskatchewan Ministry of Health and the Saskatchewan Health Quality Council released de-identified Canadian Community Health Survey (CCHS) data for analysis through data sharing agreements.

3.6 Confidentiality and Release Guidelines

3.6.1 Protection of Data Holdings

Regulations were put into place to physically protect data holdings. For the administrative data, individual identifiers were removed and health cards were encrypted by the Ministry of Health. For the CCHS, a separate link file was used to merge the CCHS cycle to the administrative data. The CCHS link file contained encrypted health card numbers, and no individual identifiers such as personal addresses and names were provided. In order to protect the release of data, all researchers were mandated to sign a confidentiality agreement explaining their responsibilities of protecting the data. Swipe cards were provided to researchers to prevent unauthorized personnel from entering the secure data lab. The internet was inaccessible on computers holding the data and only designated HQC analyst staff were permitted to release data to researchers from the facility after ensuring that confidentiality had not been breached.

3.6.2 Release Guidelines

Several measures have been put in place to maintain confidentiality. Based on the rules of the Health Quality Council of Saskatchewan, cell sizes of less than 6 including analysis on such cells were not releasable. Unique person identifiers, household identifiers, names, and any data used to potentially identify individuals were also prohibited from release.

Non-mandatory guidelines on statistical analysis and release were also set forth by Statistics Canada for the CCHS. There were recommendations on rounding estimates and conducting statistical analysis on at least 10 records for a given characteristic from the study sample. Most of the recommendations for rounding and statistical analysis were followed in the study; exceptions occurred only for a few characteristics that had less than 10 records, but in this case, HQC regulations of cell sizes less than 6 were followed for data release. There were also CCHS sampling variability guidelines for estimates in which there were restrictions on the releasability of results based on the coefficient of variation (CV)[8], which is a measure of sampling error based on the standard deviation (4). Using the appropriate *Approximate Coefficients of Variations Table*, estimates with CVs less than 33.3 are acceptable for release, while CVs greater than 33.3 are unacceptable and unreliable for release (4). For this study, all estimates were acceptable for release.

[8] The coefficient of variation (CV) is a measure of sampling error based on the standard deviation of different results that can be produced from a large sample size (4).

CHAPTER 4:
RESULTS

Chapter 4 consists of statistical analyses performed on the 2007-2008 CCHS study sample and estimated population. The results include descriptive analysis, bivariate analysis, stratified analysis, predicted probabilities, and multivariate analysis.

4.1 Descriptive Analysis

Statistical analysis was performed on the 2007-2008 CCHS 'unweighted' (study sample) and 'weighted' (estimated) population. Table 4.1 illustrates the characteristics of the study population with the use of sampling weights and without the use of sampling weights. The Census was used to compare characteristics of the 2006 population with the 2007-2008 CCHS population in order to examine how well the study sample and estimated population compared to the 'true' Canadian population.

4.1.1 Characteristics of Study Population

As seen in Table 4.1 of population characteristics, the population of the study sample and the estimated population from the 2007-2008 CCHS cycle was 7,339 and 796,207, respectively, living in 1,180 dissemination areas. The Saskatchewan population recorded from the Census in 2006 was 968,157 (87). One reason for the higher population count in the census compared to the 'weighted' frequency in the CCHS is that individuals living on reserves and individuals younger than 12 years of age are excluded from the CCHS (83). Comparing the age category of individuals less than 35 years between the estimated population and the 2006 Census provides additional evidence that excluding individuals younger than 12 years old in the CCHS may be one reason for the lower population counts in the CCHS compared to the 2006 Census (83). Despite the CCHS study sample containing of an over-sample of youth aged 12 to 19 years (4), youth are still under-represented while older individuals are over-represented in the study sample compared to the estimated population and the 'true' Census population as seen in Table 4.1.

The prevalence of diabetes as seen in Table 4.1 was 5.3% in the estimated population and 5.7% in the 2006 Census. There were almost equal proportions of male and female respondents in the estimated population and the 2006 Census, however there were slightly more female than male respondents in the study sample. The greatest proportion of respondents were under the age of 35 years and less than 10% of respondents were greater than 80 years of age.

Table 4.1: Characteristics of the CCHS 2007/2008 Sample and Estimated Population Compared to the 2006 Census of Saskatchewan

Indicator	Study Sample Frequency [N(%)]	Estimated Population Frequency [N(%)]	2006 Census SK. [N(%)]
Total population	7,339 (100%)	796,207 (100%)	968,157[9] (100%)
Diabetes			
Yes	485 (6.6%)	42,053 (5.3%)	44,870 (5.7%)[10]
No	6,854 (93.4%)	754,154 (94.7%)	
Individual-level			
Sex			
Male	3,334 (45.4%)	393,925 (49.5%)	475,240 (49.1%)
Female	4,005 (54.6%)	402,281 (50.5%)	492,915 (50.9%)
Age			
<35	2,323 (31.7%)	299,430 (37.6%)	443,020 (45.8%)
35-44	916 (12.5%)	113,575 (14.3%)	127,875 (13.2%)
45-59	1,661 (22.6%)	201,444 (25.3%)	204,245 (21.1%)
60-79	1,815 (24.7%)	142,416 (17.9%)	145,105 (15.0%)
80+	624 (8.5%)	39,342 (4.9%)	47,920 (4.9%)
Total Household income			$46,705(median)[11]
<$30,000	1,569 (21.4%)	120,208 (15.1%)	
$30,000-$49,999	1,321 (18.0%)	127,627 (16.0%)	
$50,000+	3,279 (44.7%)	421,454 (52.9%)	
Missing	1,170 (15.9%)	126,918 (15.9%)	
Marital Status			
Partnered	3,874 (52.8%)	464,870 (58.4%)	454,035[12] (54.2%)
Widowed	833 (11.4%)	46,916 (5.9%)	56,955 (6.8%)
Separated	620 (8.5%)	49,536 (6.2%)	70,555[13] (8.4%)

[9] Excludes one or more Indian reserves and/or settlements, and population counts are adjusted by Statistics Canada to ensure confidentiality. Statistics Canada, 2006 Census of Population (87).
[10] Statistics Canada CANSIM Table 105-0501 - Diabetes count and rate based on Canadian Community Health Survey for 2007 (88).
[11] Median total household income in 2005 for all private households. Statistics Canada, 2006 Census of Population (87).
[12] "Partnered" marital status refers to individuals 15 years and older in a common-law relationship and individuals who are legally married (and not separated). Statistics Canada, 2006 Census of Population (87).

Single	2,005 (27.3%)	233,155 (29.3%)	256,450 (30.6%)
Missing	7 (0.1%)	1,729 (0.2%)	n/a
BMI			
Neither	2,949 (40.2%)	335,102 (42.1%)	287,939 (41.3%)[14]
Overweight	2,300 (31.3%)	241,935 (30.4%)	226,423 (34.3%)
Obese	1,535 (20.9%)	163,728 (20.6%)	144,742 (21.9%)
Missing	555 (7.6%)	55,441 (7.0%)	n/a
Respondent education			
<High School	2,200 (30.0%)	212,290 (26.7%)	231,730 (34.1%)
High school grad	1,793 (24.4%)	212,862 (26.7%)	205,495 (30.2%)
Post-secondary	3,247 (44.2%)	361,432 (45.4%)	242,705[15] (35.7%)
Missing	99 (1.4%)	9,623 (1.2%)	n/a
Employment			
Employed	4,143 (56.5%)	507,672 (63.8%)	494,900[16] (94.4%)
Unemployed	1,846 (25.2%)	178,411 (22.4%)	29,400[17] (5.6%)
Missing	1,350 (18.4%)	110,124 (13.8%)	n/a
Physical Activity			
Active	1,648 (22.5%)	187,787 (23.6%)	
Moderately active	1,664 (22.7%)	185,072 (23.2%)	
Inactive	3,809 (51.9%)	400,210 (50.3%)	
Missing	218 (3.0%)	23,138 (2.9%)	
Aboriginal Identity			
Aboriginal	746 (10.2%)	74,046 (9.3%)	141,890 (14.9%)
Non-Aboriginal	6,528 (89.0%)	715,457 (89.9%)	811,955 (85.1%)
Missing	65 (0.9%)	6,704 (0.8%)	n/a
Area-level			

[13] "Separated" marital status refers to individuals 15 years and older who are separated, but still legally married and individuals who are divorced. Statistics Canada, 2006 Census of Population (87).
[14] Statistics Canada CANSIM Table 105-0501 and Table 105-4009 –BMI count and rate based on Canadian Community Health Survey for 2007 (88,89).
[15] "Post-secondary" education refers to individuals 15 years and older who earned a college, CEGEP, or other non-university certificate or diploma, individuals who earned a university certificate or diploma below the bachelor level, and individuals who earned a university certificate, diploma, or degree. Statistics Canada, 2006 Census of Population (87).
[16] "Employed" refers to people who participated in paid work, self-employment, family farm work, or a business. Individuals who were absent from work (due to illness or for other reasons) with or without pay were also included (87).
[17] "Unemployed" refers to people who sought paid work, people who were laid off and returning to their job, and people who were set to begin a new job in less than 4 weeks, but did not participate in paid work during the week prior to the census. Counts include people who are non-institutionalized and those who are 15 years and older. Statistics Canada, 2006 Census of Population (87).

Average Income			
<$30,000	3,767 (51.3%)	370,291 (46.5%)	
$30,000-$49,999	2,933 (40.0%)	362,756 (45.6%)	
$50,000+	237 (3.2%)	37,907 (4.8%)	
Missing	402 (5.5%)	25,253 (3.2%)	
Geography			
Urban	4,752 (64.8%)	587,387 (73.8%)	578,068[18] (59.7%)
Rural	2,587 (35.3%)	208,820 (26.2%)	390,089 (40.3%)
Spatial Area			
Dissemination Area	1,180 DAs		

From Table 4.1 above, approximately half of the CCHS respondents in the sample and estimated population had a total household income of more than $50,000. From the 2006 Census, the median income for 2005 was $46,705 (87). Notably, a large number of respondents (about 15%) from the CCHS did not provide a household income for differing reasons.

For marital status, there were similar proportions of individuals in each category for the estimated population and the 2006 Census. Comparing the sample with the estimated population, approximately half of respondents were in partnered relationships, about 30% were single, and even fewer were separated and divorced. For BMI, there were almost equal proportions of individuals in each of the BMI categories in the CCHS and the Census. There were about 40% of individuals who were neither overweight nor obese, about 30% were overweight, and 20% were obese. A large percentage of individuals (about 7%) in the CCHS did not provide weight and height information, possibly due to the sensitive nature of the topic (91).

There were 30% of individuals who had less than a high school diploma in the study sample, while there were slightly fewer individuals (26.7%) with the same credential in the estimated population. There were more individuals who had less than a high school diploma and fewer individuals with a high school diploma or a post-secondary degree in the Census than in the estimated population.

[18] "Urban" was counted as census metropolitan areas (CMAs) and census agglomerations (CAs), while "rural" was counted as strong, moderate, and weak/no influenced zones. Statistics Canada, 2006 Census of Population (90).

For employment, approximately 500,000 individuals were employed according to the Census and the estimated population. The number of individuals who were unemployed in the Census was about 29,000, while the number of unemployed in the estimated population was about 178,000. The low count of unemployed individuals in the Census compared to the estimated population could be attributed to the inclusion and exclusion of unpaid work as a form of employment or unemployment.

In terms of physical activity, approximately half of respondents were inactive, while the other half engaged in some type of physical activity. There was no available data for comparisons from the Census on the specified categories for physical activity.

Self-identified Aboriginals from the sample and the estimated population accounted for about 10% of all respondents. The proportion of self-identified Aboriginals in the 2006 Census was greater than in the CCHS.

For area-level income measured as average income at the dissemination area geography, there were 51.3% and 46.5% of respondents from the study sample and the estimated population, respectively, with an average income of less than $30,000. Approximately 90% of respondents from the study sample and the estimated population had an average income of less than $50,000. Only 3.2% and 4.8% of respondents from the study sample and the estimated population, respectively, had an average income of more than $50,000.

The proportion of individuals living in urban and rural areas was 64.8% and 35.3%, respectively, in the study sample. From the estimated population, there were 73.8% of individuals living in urban areas and 26.2% living in rural areas.

Overall, the proportions for each characteristic were similar in the study sample, estimated population, and the Census.

4.2 Bivariate Analysis

Bivariate analysis was performed whereby each characteristic was stratified by the presence and absence of diabetes as seen in Table 4.2. Figure 4.1 illustrates the percentage of diabetics by income category at the individual and area-level based on Table 4.2. As seen in Figure 4.2, individual-level income was disaggregated to illustrate the trend of individual-level income and diabetes. Figure 4.3 illustrates income-specific proportions of diabetes. Figure 4.4 shows the crude and age-adjusted rates of diabetes. Finally, Table 4.3 and Table 4.4 examine the amount of agreement between individual and area-level income.

4.2.1 Characteristics of Diabetes and Non-Diabetes in Study Population

Research Question #1: What is the type of relationship between diabetes and income (at the individual and area-level) in the study sample and in the estimated population?

As seen in Table 4.2, there was an equal proportion of males and females with diabetes in the study sample, but only slightly more males than females who had diabetes in the estimated population. Based on crude odds ratios of diabetes, males had a higher likelihood of developing diabetes than females. Consistent with literature (1), most diabetic individuals were older in age in which more than 75% of all diabetics were between the ages of 45 and 79 years. The odds ratios of diabetes significantly increased with age in the study sample and estimated population.

Approximately 38% and 28% of individuals with diabetes in the study sample and in the estimated population, respectively, had an individual-level income of less than $30,000. About 23% and 19% of individuals with diabetes from the study sample and the estimated population had an income between $30,000 and $49,999. Finally, 29% and 42% of diabetics from the study sample and the estimated population, respectively, had a household income of greater than $50,000. Crude odds ratios of diabetes decreased with increasing individual-level income relative to the highest income category of '$50,000+.'

For marital status, approximately 58% and 71% of individuals with diabetes were in partnered relationships in the study sample and in the estimated population, respectively.

Although the highest proportion of diabetics were ‘partnered,’ there was also a high proportion of ‘partnered’ individuals without diabetes. ‘Widowed’ individuals had the second highest proportion of diabetics, followed by ‘separated,’ and ‘single’ individuals had the lowest proportion of diabetics in the study sample and in the estimated population.

As seen in Table 4.2, an inverse relationship between body mass index (BMI) and diabetes was found. An increasing BMI led to higher proportions of individuals with diabetes and lower proportions of individuals without diabetes in the study sample and in the estimated population.

For respondent education level, about 42% and 37% of diabetics had less than a high school diploma and a post-secondary degree, respectively, in the study sample. The ‘missing’ category for respondent education was suppressed due to small cell size (<6) and ‘secondary grad’ was also suppressed to ensure confidentiality of the ‘missing’ category. For the estimated population, there were approximately 35% of diabetics with less than a high school diploma, 22% of diabetics with a high school diploma, and 42% of diabetics with a post-secondary degree.

There were approximately 30% and 43% of diabetics who were employed and unemployed, respectively, in the study sample. Similar proportions were found in the estimated population.

Individuals who were less physically active had greater proportions of individuals with diabetes compared to those who were ‘active.’ Although approximately 50% of people without diabetes were also inactive, odds ratios of diabetes revealed a greater likelihood of diabetes for individuals who were inactive compared to individuals who were active.

In terms of ethnicity, approximately 10% and 13% of diabetics were Aboriginal in the study sample and in the estimated population, respectively.

Table 4.2: Association between Each Characteristic and Diabetes

Indicator	Study Sample[19] ('Unweighted')				Estimated Population[20] ('Weighted')			
	With Diabetes [N(%)]	Without Diabetes [N(%)]	X^2 Value (p-value)	Crude OR (95% CI)	With Diabetes [N(%)]	Without Diabetes [N(%)]	X^2 Value (p-value)	Crude OR (95% CI)
Individual-level								
Sex			5.87				1.38	
Male	246 (50.7%)	3,088 (45.1%)	(0.02)	1.3 (1.0-1.5)	22,302 (53.0%)	371,623 (49.3%)	(0.24)	1.2 (1.1-1.2)
Female (Ref.)	239 (49.2%)	3,766 (55.0%)		1.0	19,751 (47.0%)	382,530 (50.7%)		1.0
Age			368.17 (<0.00)				14.69 (<0.00)	
<35 (Ref.)	16 (3.3%)	2,307 (33.7%)		1.0	1,482 (3.5%)	297,948 (39.5%)		1.0
35-44	26 (5.4%)	890 (13.0%)		4.2 (2.2-7.9)	2,997 (7.1%)	110,578 (14.7%)		5.4 (5.1-5.8)
45-59	104 (21.4%)	1,557 (22.7%)		9.6 (5.7-16.4)	12,417 (29.5%)	189,026 (25.1%)		13.2 (12.5-13.9)
60-79	264 (54.4%)	1,551 (22.6%)		24.5 (14.8-40.8)	20,686 (49.2%)	121,730 (16.1%)		34.2 (32.4-36.0)
80+	75 (15.5%)	549 (8.0%)		19.7 (11.4-34.1)	4,469 (10.6%)	34,873 (4.6%)		25.8 (24.3-27.4)
Total Household income			107.93 (<0.00)				41.11 (<0.00)	
<$30,000	183 (37.7%)	1,386 (20.2%)		2.9 (2.3-3.7)	11,839 (28.2%)	108,370 (14.4%)		2.5 (2.5-2.6)
$30,000-$49,999	109 (22.5%)	1,212 (17.7%)		2.0 (1.5-2.6)	7,920 (18.8%)	119,706 (15.9%)		1.5 (1.5-1.6)
$50,000+ (Ref.)	141 (29.1%)	3,138 (45.8%)		1.0	17,541 (41.7%)	403,913 (53.6%)		1.0
Missing	52 (10.7%)	1,118 (16.3%)		--	4,753 (11.3%)	122,165 (16.2%)		--
Marital Status			109.15				56.55	
Partnered (Ref.)	280 (57.7%)	3,594 (52.4%)	(<0.00)	1.0	30,001 (71.3%)	434,869 (57.7%)	(<0.00)	1.0
Widowed	97 (20.0%)	736 (10.7%)		1.7 (1.3-2.2)	4,951 (11.8%)	41,966 (5.6%)		1.7 (1.7-1.8)
Separated	62 (12.8%)	558 (8.1%)		1.4 (1.1-1.9)	3,964 (9.4%)	45,572 (6.1%)		1.3 (1.2-1.3)
Single	46 (9.5%)	1,959 (28.6%)		0.3 (0.2-0.4)	3,137 (7.5%)	230,018 (30.5%)		0.2 (0.2-0.2)
Missing	0 (0%)	7 (0.1%)		--	0 (0%)	1,729 (0.2%)		--
BMI			148.40				83.8	
Neither (Ref.)	93 (19.2%)	2,856 (41.7%)	(<0.00)	1.0	7,199 (17.1%)	327,903 (43.5%)	(<0.00)	1.0
Overweight	163 (33.6%)	2,137 (31.1%)		2.3 (1.8-3.0)	13,311 (31.7%)	228,624 (30.3%)		2.7 (2.6-2.7)
Obese	194 (40.0)	1,341 (19.6%)		4.4 (3.4-5.7)	18,426 (43.8%)	145,302 (19.3%)		5.8 (5.6-5.9)
Missing	35 (7.2%)	520 (7.6%)		--	3,117 (7.4%)	52,325 (6.9%)		--
Respondent education			36.56 (<0.00)				12.53 (0.00)	
<High School (Ref.)	204 (42.1%)	1,996 (29.1%)		1.0	14,789 (35.2%)	197,501 (26.2%)		1.0
Secondary grad	**	**		**	9,294 (22.1%)	203,568 (27.0%)		0.6 (0.6-0.6)
Post-secondary	181 (37.3%)	3,066 (44.7%)		0.6 (0.5-0.7)	17,651 (42.0%)	343,780 (45.6%)		0.7 (0.7-0.7)
Missing	*	**		--	**	9,305 (1.2%)		--
Employment			148.21				55.25	
Employed (Ref.)	147 (30.3%)	3,996 (58.3%)	(<0.00)	1.0	16,778 (40.0%)	490,894 (65.1%)	(<0.00)	1.0
Unemployed	209 (43.1%)	1,637 (23.9%)		3.5 (2.8-4.3)	17,607 (41.9%)	160,804 (21.3%)		3.2 (3.1-3.3)
Missing	129 (26.6%)	1,221 (17.8%)		--	7,668 (18.2%)	102,456 (13.6%)		--
Physical Activity			50.95				44.19	
Active (Ref.)	53 (10.9%)	1,595 (23.3%)	(<0.00)	1.0	4,050 (9.6%)	183,737 (24.4%)	(<0.00)	1.0
Moderately active	98 (20.2%)	1,566 (22.9%)		1.9 (1.3-2.6)	7,506 (17.9%)	177,566 (23.6%)		1.9 (1.8-2.0)
Inactive	314 (64.7%)	3,495 (51.0%)		2.7 (2.0-3.6)	28,551 (67.9%)	371,658 (49.3%)		3.5 (3.4-3.6)
Missing	20 (4.1%)	198 (2.9%)		--	1,946 (4.6%)	21,192 (2.8%)		--
Aboriginal Identity			1.81 (0.40)				2.72 (0.24)	
Aboriginal	57 (11.8%)	689 (10.1%)		n/a	5,648 (13.4%)	68,397 (9.1%)		1.6 (1.5-1.6)
Not Aboriginal (Ref.)	**	**		1.0	35,949 (85.5%)	679,509 (90.1%)		1.0
					**	6,248 (0.8%)		
Missing	*	*		--				--
Area-level								
Average Income			21.25				9.63	

[19] Study sample does not use sampling weights or bootstrap weights.

[20] Estimated population uses sampling weights and bootstrap weights.

<$30,000	297 (61.2%)	3,470 (50.6%)	(<0.00)	1.3 (0.7-2.20)	23,794 (56.6%)	346,497 (46.0%)	(0.01)	1.2 (1.2-1.3)
$30,000-$49,999	155 (32.0%)	2,778 (40.5%)		0.8 (0.5-1.4)	15,287 (36.4%)	347,469 (46.1%)		0.8 (0.7-0.8)
$50,000+ (Ref.)	15 (3.1%)	222 (3.2%)		1.0	2,039 (4.9%)	35,868 (4.8%)		1.0
Missing	18 (3.7%)	384 (5.6%)		--	932 (2.2%)	24,320 (3.2%)		--
Geography			0.79				1.31	
Urban (Ref.)	305 (62.9%)	4,447 (64.9%)	(0.37)	1.0	29,859 (71.0%)	557,527 (74.0%)	(0.25)	1.0
Rural	180 (37.1%)	2,407 (35.1%)		1.1 (0.9-1.3)	12,193 (29.0%)	196,627 (26.1%)		1.2 (1.1-1.2)

Notes:
-- Odds ratios were not calculated for missing categories.
* Cell size less than 6 is not releasable by Health Quality Council of Saskatchewan.
** Cell size is greater than 6, but is not releasable by Health Quality Council of Saskatchewan in order to protect cells less than 6.

For area-level income, a greater proportion of individuals with diabetes had an income of less than $50,000, yet approximately 50% of individuals without diabetes had an income of less than $30,000. Crude odds ratios of diabetes showed a statistically insignificant pattern in which odds ratios were highest for the lowest income category, but smallest for the middle income category relative to the highest income category.

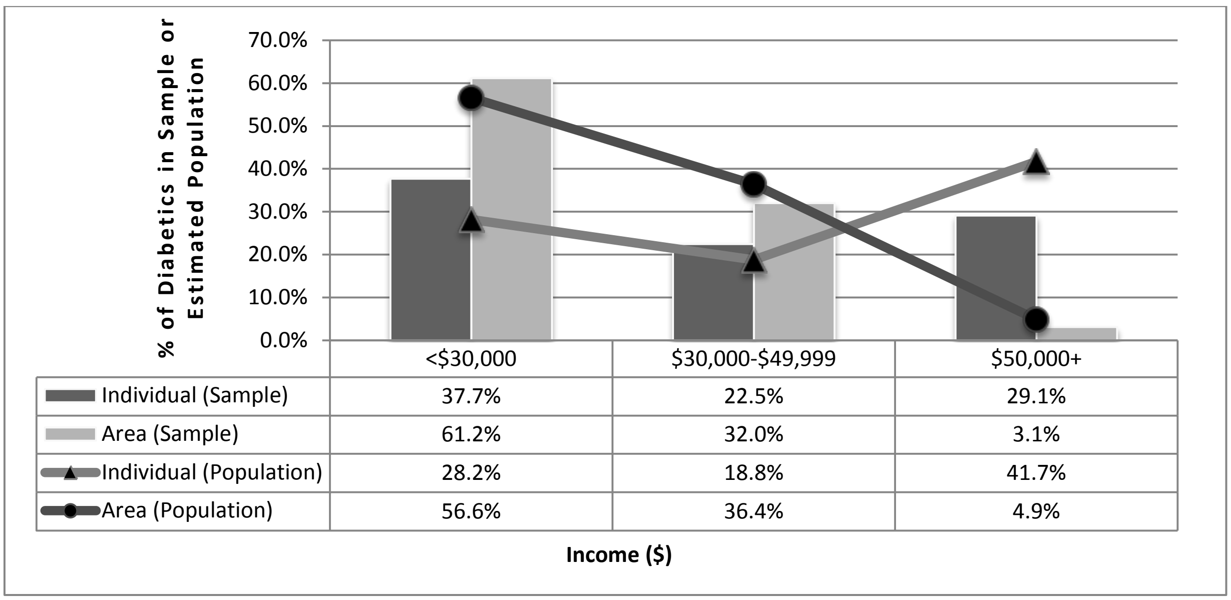
There were 63% and 71% of individuals with diabetes from urban areas in the study sample and in the estimated population, respectively. However, there were similar proportions of non-diabetics living in urban areas as well. Approximately one-third of diabetics and non-diabetics were living in rural areas. Crude odds ratios of diabetes showed that rural residents had a slightly higher likelihood of developing diabetes compared to urban residents, however the odds ratios were statistically insignificant for the study sample.

Based on Table 4.2 above, all covariates with a p-value less than 0.05 alpha were included for further analysis. From the study sample, Aboriginal identity (p-value=0.40) and geography (p-value=0.37) were non-significant, but they were kept for further analysis due to their biological significance. In addition, sex (p-value=0.24), Aboriginal identity (p-value=0.24), and geography (p-value=0.25) were non-significant in the estimated population, but they were included due to their biological significance in the study. Therefore, no variables were excluded at this point.

4.2.2 Relationship between Diabetes and (Individual and Area-Level) Income

Figure 4.1 below, which is based on Table 4.2, illustrates the percentage of individuals with diabetes for individual and area-level income in the study sample and in the estimated population. Individual-level income was aggregated in the same manner as area-level income for comparison purposes. As shown in Figure 4.2, the disaggregation of total household income after the '$30,000' category revealed a downward trend from '<$30,000' to '$50,000-$59,999' with small increases at '$60,000-$79,999' and '$100,000+,' most likely due to the disaggregation of these two income categories.

Figure 4.1: Percent of People with Diabetes in the Study Sample and Estimated Population by Individual and Area-level Income Categories

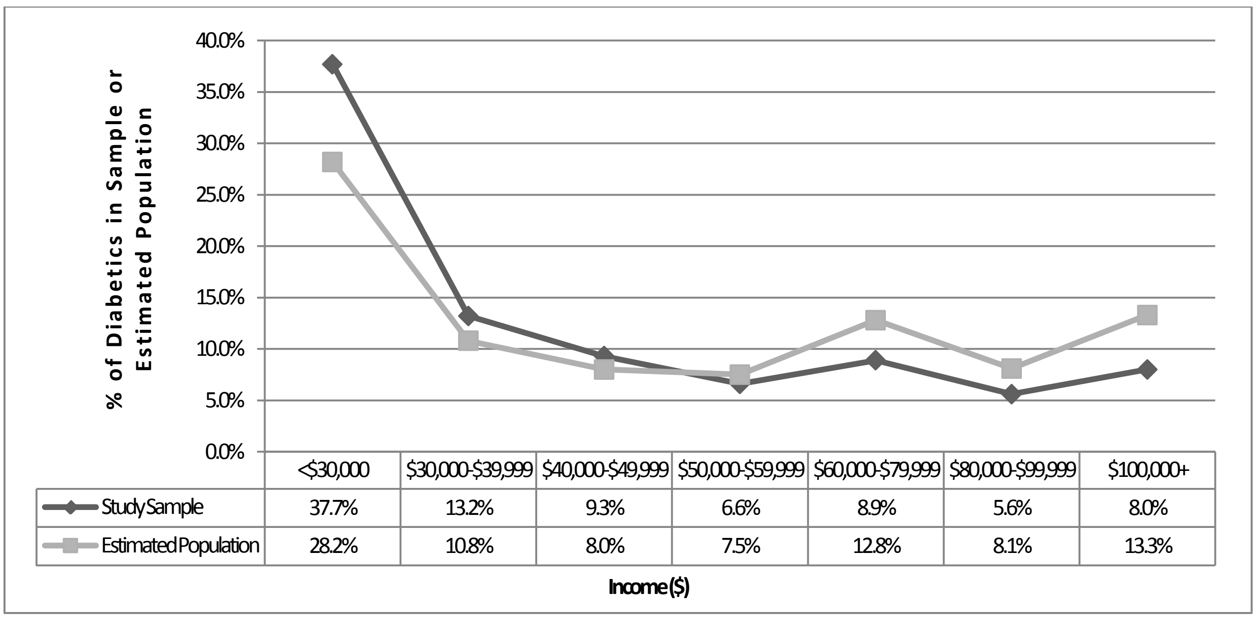


	<$30,000	$30,000-$49,999	$50,000+
Individual (Sample)	37.7%	22.5%	29.1%
Area (Sample)	61.2%	32.0%	3.1%
Individual (Population)	28.2%	18.8%	41.7%
Area (Population)	56.6%	36.4%	4.9%

For area-level income, an inverse relationship was observed in which the percentage of individuals with diabetes decreased with increasing income in the study sample and in the estimated population. However, a different pattern emerged for individual-level income in which the proportion of diabetics decreased from '<$30,000' to '$30,000-$49,999' with an increase in the percentage of diabetics from '$30,000-$49,999' to '$50,000+' in the study sample and in the estimated population. The increase in the percentage of individuals with diabetes in the highest income category for individual-level income can be attributed to limitations of the data in which most individuals had an area-level income of less than $50,000, while individual-level income varied greatly as seen in Figure 4.2.

In Figure 4.2 below, the disaggregation of income groups for individual-level income showed fewer individuals with diabetes as income increased up to $59,999. However, the trend reveals overall stability after $50,000. The small peaks at the income categories of '$60,000-$79,999' and '100,000+' indicate the need for further disaggregation of these pre-determined income groups by the CCHS in order to illustrate any potential trend.

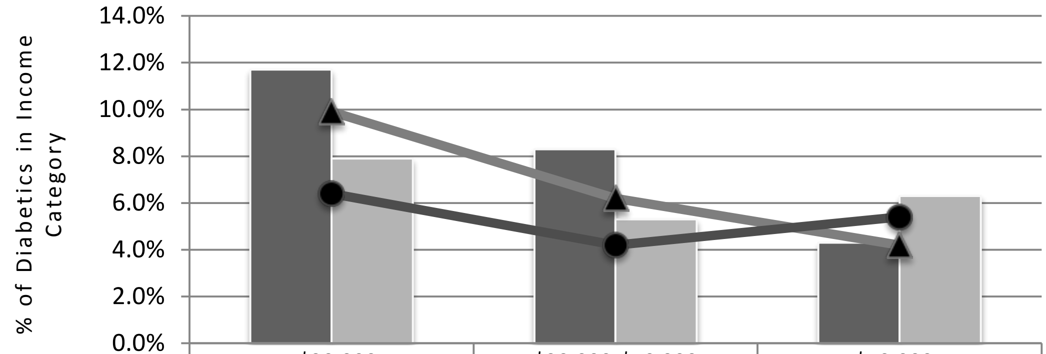
Figure 4.2: Percent of People with Diabetes in the Study Sample and Estimated Population By Stratified Individual-level Income



	<$30,000	$30,000-$39,999	$40,000-$49,999	$50,000-$59,999	$60,000-$79,999	$80,000-$99,999	$100,000+
Study Sample	37.7%	13.2%	9.3%	6.6%	8.9%	5.6%	8.0%
Estimated Population	28.2%	10.8%	8.0%	7.5%	12.8%	8.1%	13.3%

Figure 4.3 illustrates income-specific proportions of diabetics whereby an inverse relationship between individual-level income and diabetes exists. Contrary to Figure 4.1, a slight increase in the percentage of diabetics was observed at an area-level income '$50,000+.' Income-specific proportions were found to be smaller for area-level income than at individual-level income except for the '$50,000+' category.

Figure 4.3: Percent of People with Diabetes within each Income Category for the Study Sample and Estimated Population



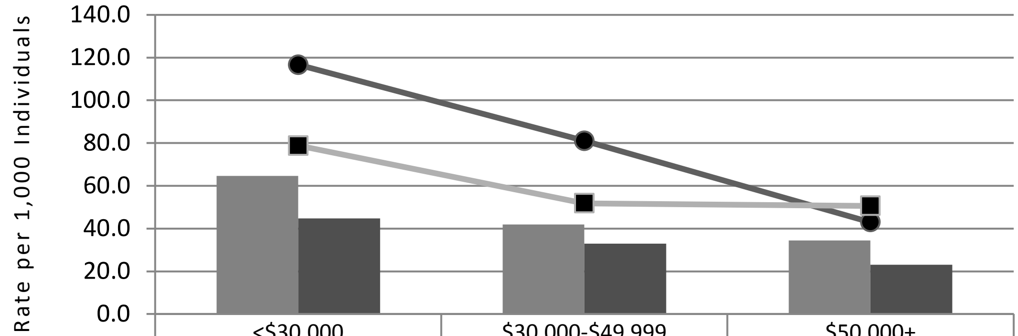
	<$30,000	$30,000-$49,999	$50,000+
Individual (Sample)	11.7%	8.3%	4.3%
Area (Sample)	7.9%	5.3%	6.3%
Individual (Population)	9.9%	6.2%	4.2%
Area (Population)	6.4%	4.2%	5.4%

Income ($)

Contradictory findings from Figure 4.1 and Figure 4.3 may be explained by the confounding effects of age on the income-diabetes relationship. That is, the relatively flat relationship between diabetes prevalence and individual-level income in Figure 4.1 is because both diabetes frequency and income increase with age, such that the relationship of the diabetes and income is confounded by the associations of age with both diabetes and income (1). For this reason, age-adjusted rates may better represent the ‘actual’ effect of income on diabetes. Income-specific crude rates and age-adjusted rates (based on the standard 1991 Canadian population) were conducted via direct standardization as seen in Figure 4.4 below.

Crude and age-adjusted rates decreased as income increased as seen in Figure 4.4. A slight increase from the ‘$30,000-$49,999’ category to the ‘$50,000+’ category was observed for the area-level income crude rate. Crude rates are generally smaller than age-adjusted rates.

Figure 4.4: Crude and Age-Adjusted Rates of People with Diabetes within each Income Category for Individual and Area-level Income in the Study Sample



	<$30,000	$30,000-$49,999	$50,000+
Adjusted Rate (Individual)	64.7	41.9	34.4
Adjusted Rate (Area)	44.7	32.9	23.1
Crude Rate (Individual)	116.6	81.0	43.0
Crude Rate (Area)	78.8	51.8	50.6

Income ($)

Notes:
Crude rates and age-adjusted rates for individual-level income of '$30,000-$49,999' and area-level income of '$30,000-$49,999' are slightly underestimated due to the suppression of one age category with low cell size (<6). Crude rate and age-adjusted rate for area-level income of '$50,000+' is also slightly underestimated due to the suppression of three age categories due to low cell size (<6). As a result, suppressed cells were imputed with a value of 0 to compute crude and age-adjusted rates.

Due to the suppression of some cells for individual-level income, the same cells were suppressed (by the Health Quality Council of Saskatchewan) for the estimated population even though cells sizes were much greater than 6. Therefore, income-specific crude and age-adjusted rates are not shown for the estimated population.

4.2.3 Agreement between Individual and Area-level Income

Research Question #2: How much agreement exists between individual-level income and area-level income in the study sample and in the estimated population?

A cross-examination of individual and area-level income showed a large amount of disagreement between the two measures. There was relatively low agreement between individual and area-level income in the study sample and in the estimated population in which the weighted Kappa statistic was 0.087 and 0.083, respectively. In Table 4.3 below of the study sample, the percentage of individuals with the same individual and area-level income was 64.3% for an income less than $30,000, 37.3% for an income between $30,000 and $49,999, and 4.8% for an income greater than $50,000. Although 64% of individuals had the same individual and area-level income of less than $30,000, there were approximately 55% and 41% of individuals with an individual income of $30,000 to $49,999 and more than $50,000, respectively, categorized with an area-level income of less than $30,000. Contrarily, approximately 29% of individuals with an individual-level income of less than $30,000 had an area-level income between $30,000 and $49,999. There were 48% of individuals with an individual-level income of more than $50,000 who had an area-level income between $30,000 and $49,999. Only 2% of individuals who had an individual-level income of less than $50,000 had an area-level income of more than $50,000.

Table 4.3: Agreement between Individual and Area-Level Income in the Study Sample

Study Sample	**Individual Income**			
	<$30,000	$30,000-$49,999	$50,000+	Missing
Area Income				
<$30,000	1,009 (64.3%)	726 (55.0%)	1,368 (41.7%)	664 (56.8%)
$30,000-$49,999	453 (28.9%)	493 (37.3%)	1,587 (48.4%)	400 (34.2%)
$50,000+	25 (1.6%)	27 (2.0%)	157 (4.8%)	28 (2.4%)
Missing	82 (5.2%)	75 (5.7%)	167 (5.1%)	78 (6.7%)

Similar findings to the study sample were observed in the estimated population as seen in Table 4.4 below. The percentage of agreement between individual and area-level income for an income less than $30,000, $30,000 to $49,9999, and more than $50,000 were 64.7%, 38.9%, and 6.4%, respectively. There were 55% of individuals with an individual-level income of more than $30,000 who were categorized with an area-level income of less than $30,000. Although 30% of individuals with a household income of less than $30,000 had an area-level income between $30,000 and $49,9999, there were approximately 54% of individuals who had a household income of more than $50,000 yet an area-level income between $30,000 and $49,9999. Interestingly, over 50% of individuals who did not report their total household income were categorized as having an area-level income of less than $30,000 in the study sample and in the estimated population as seen in Table 4.3 and Table 4.4.

Table 4.4: Agreement between Individual and Area-Level Income in the Estimated Population

Estimated Population	Individual Income			
	<$30,000	$30,000-$49,999	$50,000+	Missing
Area Income				
<$30,000	77,765 (64.7%)	69,761 (54.7%)	157,651 (37.4%)	65,114 (51.3%)
$30,000-$49,999	35,943 (29.9%)	49,605 (38.9%)	225,893 (53.6%)	51,315 (40.4%)
$50,000+	2,407 (2.0%)	3,289 (2.6%)	27,068 (6.4%)	5,143 (4.1%)
Missing	4,093 (3.4%)	4,971 (3.9%)	10,842 (2.6%)	5,346 (4.2%)

4.3 Multivariable Analysis

Analysis on individual and area-level income stratified by geography type (i.e., urban, rural) was conducted to determine whether income gradients found previously would hold in both urban and rural areas. Table 4.5 below depicts the percentage of diabetics and non-diabetics distributed across income type and geography in the study sample and in the estimated population.

4.3.1 Relationship between Diabetes and (Individual and Area-Level) Income in Urban and Rural Areas

Research Question #2a: How does the relationship between individual-level and area-level income differ in urban and rural areas?

Table 4.5 shows the proportions of diabetics and non-diabetics as well as odds ratios of diabetes by geography and income level. Proportions and odds ratios were not produced for area-level income in rural areas due to low cell counts (<6) in the study sample. Due to the low cell counts in the study sample for area-level income in rural areas, corresponding cells were also suppressed for the estimated population to maintain confidentiality.

For individual-level income in rural areas, the proportion of individuals with diabetes decreased as income level increased in the study sample and in the estimated population. In addition, the proportion of individuals without diabetes increased as individual-level income level increased in rural areas. Although an obvious pattern was not observed for individual-level income in urban areas, the estimated population revealed a greater proportion of diabetics with increasing individual-level income. The percentage of individuals without diabetes also increased as income level increased. Odds ratios of diabetes decreased with increasing individual-level income relative to the highest income category of '$50,000+' in urban and rural areas, however odds ratios in rural areas were slightly higher than odds ratios in urban areas for each individual-level income category.

Area level-income in urban areas showed a downward trend as the proportion of individuals with diabetes increased with increasing income. The proportions of individuals of diabetics with an area-level income of less than $30,000, an income between $30,000 to $49,999, and an income more than $50,000 were approximately 54%, 39%, and 5%, respectively, in the study sample. In the estimated population, the proportions of diabetics with an area-level income of less than $30,000, an income between $30,000 to $49,999, and an income more than $50,000 were 48%, 44%, and 7%, respectively. Odds ratios of diabetes showed an increased likelihood of diabetes for the lowest area-level income category relative to the highest income category. There was a decreased likelihood of diabetes for the middle income category compared to the highest income category, however the odds ratio for the study sample was statistically insignificant.

Table 4.5: Percent of People with and without Diabetes in the Study Sample and Estimated Population by Individual and Area-level Income And Urban or Rural Residence

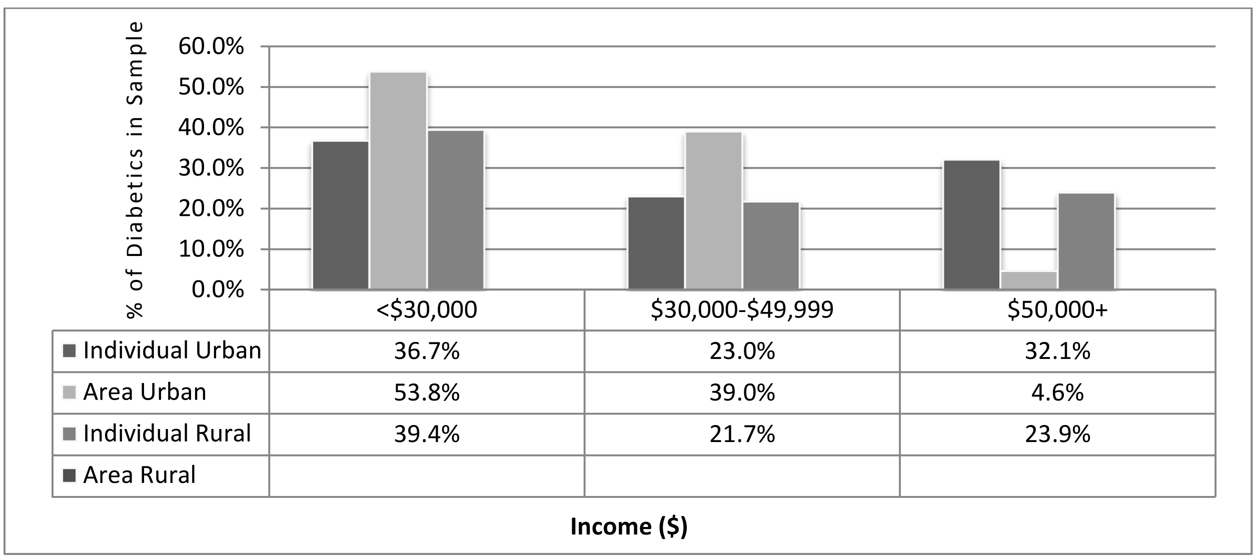
Geography	Income	Study Sample Frequency [N(%)]			Estimated Population Frequency [N(%)]		
		With Diabetes	Without Diabetes	Crude OR (95% CI)	With Diabetes	Without Diabetes	Crude OR (95% CI)
Rural	***Individual-level*** **Total household income**						
	<$30,000	71 (39.4%)	515 (21.4%)	3.2 (2.2-4.8)	4,055 (33.3%)	29,638 (15.1%)	3.7 (3.5-3.9)
	$30,000-$49,999	39 (21.7%)	393 (16.3%)	2.3 (1.5-3.7)	2,377 (19.5%)	32,850 (16.7%)	1.9 (1.8-2.1)
	$50,000+ (Ref.)	43 (23.9%)	1,010 (42.0%)	1.0	3,368 (27.6%)	90,622 (46.1%)	1.0
	Missing	27 (15.0%)	489 (20.3%)	--	2,392 (19.6%)	43,517 (22.1%)	--
Urban	***Individual-level*** **Total household income**						
	<$30,000	112 (36.7%)	871 (19.6%)	2.8 (2.1-3.7)	7,783 (26.1%)	78,732 (14.1%)	2.2 (2.1-2.2)
	$30,000-$49,999	70 (23.0%)	819 (18.4%)	1.9 (1.4-2.5)	5,543 (18.6%)	86,856 (15.6%)	1.4 (1.4-1.5)
	$50,000+ (Ref.)	98 (32.1%)	2,128 (47.9%)	1.0	14,173 (47.5%)	313,291 (56.2%)	1.0
	Missing	25 (8.2%)	629 (14.1%)	--	2,360 (7.9%)	78,648 (14.1%)	--
	Area-level **Average income**						
	<$30,000	164 (53.8%)	1,910 (43.0%)	1.3 (0.7-2.3)	14,342 (48.0%)	216,772 (38.9%)	1.2 (1.1-1.2)

	$30,000-$49,999	119 (39.0%)	2,149 (48.3%)	0.8 (0.5-1.5)	13,136 (44.0%)	299,667 (53.8%)	0.8 (0.7-0.8)
	$50,000+ (Ref.)	14 (4.6%)	210 (4.7%)	1.0	1,967 (6.6%)	34,845 (6.3%)	1.0
	Missing	8 (2.6%)	178 (4.0%)	--	415 (1.4%)	6,243 (1.1%)	--

Notes:
-- Odds ratios were not calculated for missing categories.
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

Based on Table 4.5 above, Figure 4.5 illustrates an inverse relationship between diabetes and area-level income in urban areas of the study sample. With an area-level income less than $30,000, an income between $30,000 to $49,999, and an income more than $50,000, the proportions of individuals with diabetes were about 54%, 39%, and 5%, respectively.

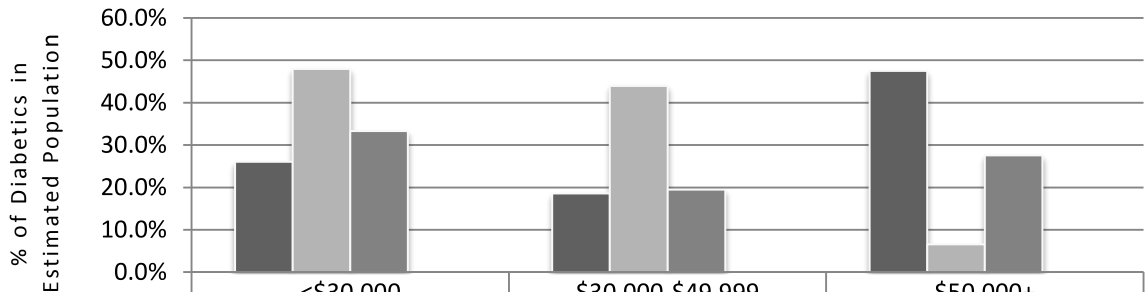
Figure 4.5: Percent of People with Diabetes in the Study Sample by Individual and Area-level Income Categories And Urban or Rural Residence



Notes:
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

Figure 4.6, also based on Table 4.5 above, illustrates a diminished income gradient for area-level income in the urban estimated population. At an area-level income of '<$30,000,' '$30,000-$49,999,' and '$50,000+,' the proportions of diabetics in urban areas are 48%, 44%, and 7%, respectively. For individual-level income, the percentage of individuals with diabetes decreased from the lowest to the middle income category, however a sharp increase was seen at the '$50,000+' category in urban and rural areas.

Figure 4.6: Percent of People with Diabetes in the <u>Estimated Population</u> by Individual and Area-level Income Categories And Urban or Rural Residence



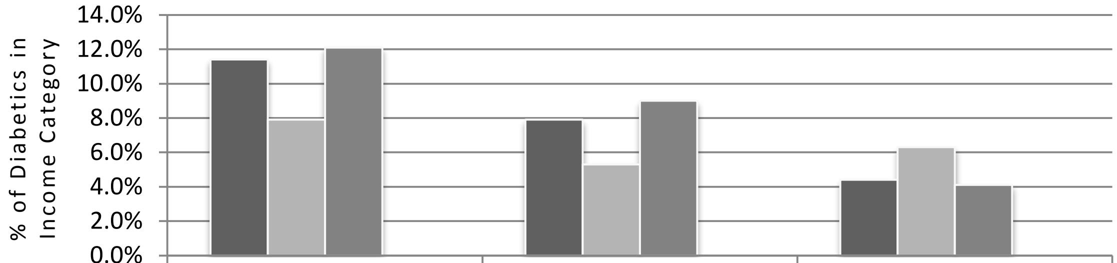
	<$30,000	$30,000-$49,999	$50,000+
Individual Urban	26.1%	18.6%	47.5%
Area Urban	48.0%	44.0%	6.6%
Individual Rural	33.3%	19.5%	27.6%
Area Rural			

Income ($)

Notes:
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

Income-specific proportions of individuals with diabetes were determined for urban and rural areas in the study sample and in the estimated population as seen in Figure 4.7 and Figure 4.8, respectively. As seen in Figure 4.7, an overall downward trend in the proportions of diabetics was observed for individual and area-level income in urban and rural areas. Percentages at the area-level income in rural areas were suppressed due to low cell counts (<6).

Figure 4.7: Percent of People with Diabetes within each Income Category for Individual and Area-level Income And Urban or Rural Residence in the Study Sample



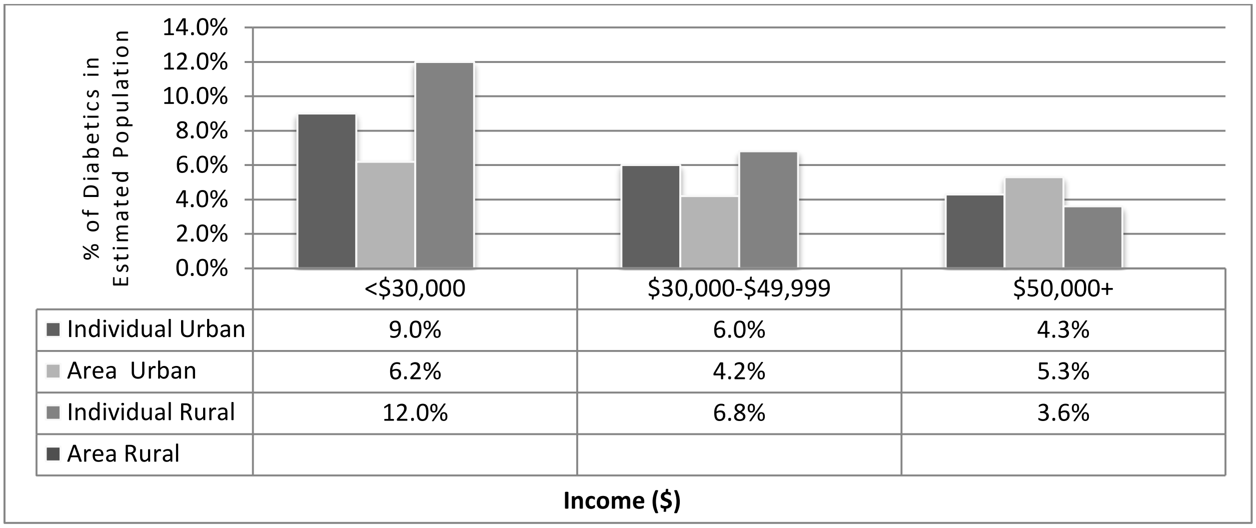
	<$30,000	$30,000-$49,999	$50,000+
Individual Urban	11.4%	7.9%	4.4%
Area Urban	7.9%	5.3%	6.3%
Individual Rural	12.1%	9.0%	4.1%
Area Rural			

Income ($)

Notes:
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

In Figure 4.7, income-specific proportions of diabetics illustrate an overall downward trend for individual and area-level incomes in urban and rural areas. Figure 4.8 below of the estimated population illustrates similar findings to Figure 4.5. Individual-level income proportions were larger than area-level income proportions.

Figure 4.8: Percent of People with Diabetes within each Income Category for Individual and Area-level Income And Urban or Rural Residence in the <u>Estimated Population</u>



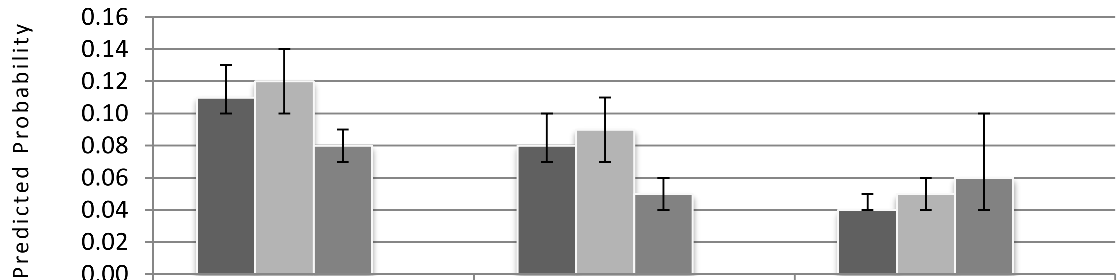
Notes:
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

Research Question #3a: What are the predicted probabilities of diabetes when individual and area-level income are used in urban and rural Saskatchewan?

Diabetes inequalities were primarily examined via predicted probabilities and odds ratios from logistic regression. Predicted probabilities were generated for individual and area-level income in urban and rural areas as seen in Figure 4.9 and Figure 4.10. Since counts from area-level income in rural areas were suppressed due to low cell count (<6), predicted probabilities were also suppressed.

As seen in Figure 4.9 of the study sample, predicted probabilities decreased with increasing income. For individual-level income in urban areas, the predicted probability of diabetes was 0.11 at '<$30,000', 0.08 at '$30,000-$49,999,' and 0.04 at '$50,000+.' In rural areas, the predicted probability at an individual-level income of less than $30,000 was 0.12, 0.09 for an individual-level income between $30,000 and $49,999, and 0.05 for an individual-level income greater than $50,000. In the estimated population as seen in Figure 4.10, predicted probabilities also decreased as income increased.

Figure 4.9: Predicted Probabilities of Diabetes for the Study Sample by Individual and Area-level Income And Urban or Rural Residence

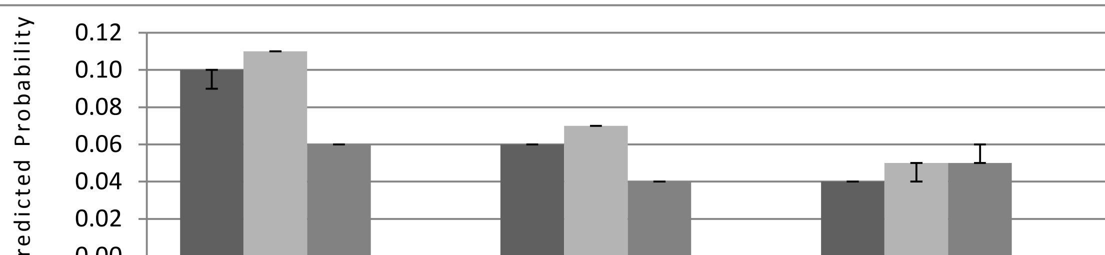


	<$30,000	$30,000-$49,999	$50,000+
Individual Urban	0.11	0.08	0.04
Individual Rural	0.12	0.09	0.05
Area Urban	0.08	0.05	0.06
Area Rural			

Income ($)

Notes:
Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

Figure 4.10: Predicted Probabilities of Diabetes for the Estimated Population by Individual and Area-level Income And Urban or Rural Residence



	<$30,000	$30,000-$49,999	$50,000+
Individual Urban	0.10	0.06	0.04
Individual Rural	0.11	0.07	0.05
Area Urban	0.06	0.04	0.05
Area Rural			

Income ($)

Notes:

Area-level income in rural areas was suppressed for the study sample and the estimated population due to low cell counts (<6) in the study sample. Cell sizes less than 6 are not releasable by Health Quality Council of Saskatchewan in order to maintain confidentiality.

4.3.2 Logistic Regression Analysis

Research Question #3b: How do the odds ratios of diabetes differ between individual and area-level income categories and between urban and rural areas?

Multivariable logistic regression was conducted to examine the odds ratios between individual and area-level income categories after adjusting for confounders. Table 4.6 and Table 4.7 represent the saturated model and the final model, respectively. Based on the bivariate analysis in Table 4.2, all covariates were kept in the saturated model. Covariates with p-values greater than 0.05 alpha in the saturated model were excluded from the final model. As individual and area-level incomes were main predictors of the study, the following interactions were tested: income and geography, income and education, income and Aboriginal identity, income and BMI. However, the above interactions were found to be insignificant and they were excluded from the final models.

4.3.2.a Saturated Model

As seen in Table 4.6, saturated models were conducted for the study population and the estimated population as model-based and design-based models, respectively. The model-based model does not use sampling weights or bootstrap weights, while the design-based model uses sampling weights and bootstrap weights.

From the model-based model in Table 4.8, individuals who have an individual-level income of less than $30,000 are 1.47 times more likely of developing diabetes than individuals who have an individual-level income of more than $50,000. Similarly, individuals who have an individual-level income between $30,000 and $49,999 are 1.30 times more likely to develop diabetes than individuals who have an income of more than $50,000.

Individuals living in rural areas were 14% and 18% less likely to develop diabetes than their urban counterparts in the ‘unweighted’ individual and area-level models, respectively. A lower likelihood of diabetes was found as individuals living in rural areas were 17% and 23%

less likely of developing diabetes than individuals living in urban areas based on the 'weighted' individual and area-level income models, respectively.

Males were approximately 1.5 times more likely of developing diabetes than females in the two 'unweighted' models. A lower likelihood of diabetes was found for males in the design-based models as males were approximately 1.4 times likely of diabetes onset compared to females.

Age was strongly correlated with diabetes as older age revealed a greater likelihood of diabetes. In the two model-based models, odds ratios increased from approximately 4 to 15 as age increased from 35 years. In the two design-based models, odds ratios ranged from about 4 to 24 with a greater likelihood of diabetes development after the age of 60 years compared to an age less than 35 years.

Table 4.6: Saturated Model of Covariates to Examine Diabetes Inequalities

	Model-based (Unweighted)		**Design-based (Weighted)**	
	Model 1 Individual income	**Model 2 Area income**	**Model 3 Individual income**	**Model 4 Area income**
	OR (95% CI)	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Individual-level Income		Not included		Not included
<$30,000	1.47 (1.10-1.97)*		1.21 (0.81-1.80)	
$30,000-$49,999	1.30 (0.98-1.74)		0.95 (0.65-1.38)	
$50,000+	1.00		1.00	
Area-level Income	Not included	.	Not included	
<$30,000		1.12 (0.63-1.98)		1.10 (0.48-2.52)
$30,000-$49,999		0.87 (0.49-1.55)		0.85 (0.35-2.05)
$50,000+		1.00		1.00
Geography				
Urban	1.00	1.00	1.00	1.00
Rural	0.86 (0.70-1.06)	0.82 (0.67-1.02)	0.83 (0.63-1.09)	0.77 (0.59-1.01)
Sex				
Male	1.51 (1.22-1.86)*	1.50 (1.21-1.86)*	1.36 (1.00-1.84)	1.37 (1.02-1.85)*
Female	1.00	1.00	1.00	1.00
Age				
<35	1.00	1.00	1.00	1.00

35-44	3.74 (1.94-7.21)*	3.92 (2.03-7.56)*	4.07 (1.49-11.12)*	4.24 (1.54-11.68)*
45-59	8.38 (4.73-14.84)*	8.49 (4.78-15.09)*	9.66 (3.83-24.33)*	9.77 (3.84-24.87)*
60-79	15.26 (8.66-26.89)*	15.79 (8.96-27.84)*	21.58 (8.31-56.03)*	21.51 (8.27-55.99)*
80+	14.48 (7.38-28.42)*	15.18 (7.75-29.73)*	23.91 (8.35-68.48)*	24.25 (8.37-70.23)*
BMI				
Neither	1.00	1.00	1.00	1.00
Overweight	2.06 (1.57-2.70)*	2.01 (1.53-2.64)*	2.11 (1.45-3.09)*	2.10 (1.43-3.08)*
Obese	3.80 (2.89-4.99)*	3.76 (2.86-4.93)*	4.24 (2.76-6.53)*	4.21 (2.72-6.52)*
Respondent education				
<High School	1.00	1.00	1.00	1.00
Secondary grad	0.88 (0.67-1.17)	0.88 (0.67-1.16)	0.92 (0.63-1.35)	0.92 (0.63-1.34)
Post-secondary	0.84 (0.66-1.06)	0.83 (0.65-1.04)	0.89 (0.65-1.21)	0.91 (0.68-1.22)
Employment				
Employed	1.00	1.00	1.00	1.00
Unemployed	2.14 (1.63-2.81)*	2.27 (1.74-2.96)*	2.11 (1.47-3.02)*	2.08 (1.45-2.99)*
Physical Activity				
Active	1.00	1.00	1.00	1.00
Moderate	1.38 (0.96-1.97)	1.37 (0.96-1.97)	1.42 (0.86-2.35)	1.42 (0.85-2.35)
Inactive	1.63 (1.19-2.24)*	1.61 (1.17-2.21)*	2.16 (1.38-3.37)*	2.09 (1.33-3.28)*
Aboriginal Identity				
Aboriginal	2.30 (1.65-3.21)*	2.41 (1.73-3.36)*	2.70 (1.71-4.25)*	2.79 (1.77-4.42)*
Not Aboriginal	1.00	1.00	1.00	1.00
Marital Status				
Partnered	1.00	1.00	1.00	1.00
Widowed	0.86 (0.64-1.16)	0.93 (0.69-1.24)	0.80 (0.54-1.18)	0.78 (0.54-1.13)
Separated	1.22 (0.88-1.68)	1.39 (1.02-1.90)*	1.16 (0.67-2.02)	1.22 (0.71-2.08)
Single	0.72 (0.50-1.03)	0.78 (0.54-1.11)	0.68 (0.39-1.20)	0.69 (0.39-1.21)

Notes:
* Significant p-value less than 0.05 alpha.
Reference: income -$50,000+, geography –urban, sex –female, age –<35, BMI –neither, respondent education -<high school, employment –employed, physical activity –active, Aboriginal identity –Not Aboriginal, marital status –partnered.
Model 1: ‘Unweighted’ model of individual income adjusted for geography, sex, age, BMI, education, employment, physical activity, aboriginal identity, marital status
Model 2: ‘Unweighted’ model of area income adjusted for geography, sex, age, BMI, education, employment, physical activity, aboriginal identity, marital status.
Model 3: ‘Weighted’ model of individual income adjusted for geography, sex, age, BMI, education, employment, physical activity, aboriginal identity, marital status.
Model 4: ‘Weighted’ model of area income adjusted for geography, sex, age, BMI, education, employment, physical activity, aboriginal identity, marital status.

For the two ‘unweighted’ models, individuals who were overweight were 2 times more likely to develop diabetes, while individuals who were obese were almost 3.8 times more likely of developing diabetes compared to individuals who were neither overweight nor obese. The odds of increasing BMI were slightly higher in the design-based models than the model-based models. In the ‘weighted’ models, individuals who were overweight were 2.1 times more likely of developing diabetes, while individuals who were obese were 4.2 times more likely of developing diabetes than individuals who were neither overweight nor obese.

Although respondent education was statistically non-significant, individuals with a high school diploma were 12% less likely to obtain diabetes and individuals with a post-secondary education were approximately 16% less likely to obtain diabetes compared to individuals with less than high school education in the ‘unweighted’ models. In the design-based models, individuals with a high school diploma were 8% less likely and individuals with a post-secondary education were approximately 10% less likely of developing diabetes than individuals without a high school diploma.

Individuals who were unemployed were 2.1 and 2.3 times more likely of developing diabetes than employed individuals in the ‘unweighted’ individual and area-level income models, respectively. In the design-based models, individuals who were unemployed were 2.1 times more likely of diabetes development compared to individuals who were employed.

Individuals who were less physically active were more likely to obtain diabetes than active individuals. In the ‘unweighted’ models, the odds of diabetes were approximately 1.4 and 1.6 for individuals who were moderately active and individuals who were inactive, respectively, compared to individuals who were active. In the ‘weighted’ models, the odds of diabetes were about 1.4 and 2.1 for individuals who were moderately active and individuals who were inactive, respectively, compared to active individuals.

Aboriginals Peoples were 2.3 and 2.4 times more likely of developing diabetes than non-Aboriginals in the ‘unweighted’ individual and area-level income models, respectively. In the design-based individual and area-level income models, the odds of diabetes for Aboriginals Peoples were 2.7 and 2.8, respectively.

In terms of marital status, individuals who were separated were approximately 1.2 times more likely of obtaining diabetes than partnered individuals, while single people and widowed individuals were less likely of developing diabetes than partnered individuals in the model-based and design-based models. However, the findings for marital status were statistically non-significant.

Based on p-values of less than 0.05 alpha, statistically significant covariates were retained for each of the final models. Statistically significant covariates in all models included age, sex, BMI, employment, physical activity, and Aboriginal identity. Individual and area-level incomes in addition to geography were retained for further analysis in the final models regardless of their p-value due to their biological significance in the study.

4.3.2.b Final Model

Covariates for all final models were the same, allowing for comparisons between individual and area-level income models and the comparisons between 'unweighted' and 'weighted' models to be made.

As seen in Table 4.7, individual-level income in the model-based model showed the odds of diabetes for individuals with an income of less than $30,000 and an income between $30,000 and $49,999 were 1.5 and 1.3, respectively, compared to individuals with an income of more than $50,000. In the 'weighted' individual and area-level income models, however, the odds of diabetes were lower for individuals with an individual-level income of less than $30,000 and an individual-level income between $30,000 and $49,999. Based on the first objective of comparing area-level income estimates to individual-level income estimates (which are bolded in Table 4.7 below), the odds ratios of diabetes are significantly lower for area-level income than individual-level income in both the 'unweighted' and 'weighted' models. Individual-level income in the 'unweighted' model was the only income variable where a clear downward gradient was found between income and odds ratios with statistical significance.

Table 4.7: Final Model to Examine Diabetes Inequalities

	Model-based (Unweighted)		Design-based (Weighted)	
	Model 1 Individual income	**Model 2 Area income**	**Model 3 Individual income**	**Model 4 Area income**
	OR (95% CI)	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Individual-level Income		Not included		Not included
<$30,000	**1.50 (1.15-1.96)***		**1.19 (0.83-1.70)**	
$30,000-$49,999	**1.31 (0.99-1.74)**		**0.94 (0.65-1.37)**	
$50,000+	**1.00**		**1.00**	
Area-level Income	**Not included**		**Not included**	
<$30,000		1.17 (0.66-2.06)		1.12 (0.50-2.52)
$30,000-$49,999		0.90 (0.51-1.59)		0.86 (0.36-2.04)
$50,000+		1.00		1.00
Geography				
Urban	**1.00**	1.00	**1.00**	1.00
Rural	**0.87 (0.71-1.06)**	0.82 (0.66-1.00)	**0.84 (0.64-1.10)**	0.77 (0.59-1.00)
Sex				
Male	**1.54 (1.25-1.89)***	1.52 (1.24-1.85)*	**1.39 (1.03-1.87)***	1.40 (1.05-1.88)*
Female	**1.00**	1.00	**1.00**	1.00
Age				
<35	**1.00**	1.00	**1.00**	1.00
35-44	**4.10 (2.16-7.79)***	4.21 (2.22-7.99)*	**4.60 (1.83-11.56)***	4.82 (1.91-12.18)*
45-59	**9.69 (5.60-16.76)***	9.66 (5.58-16.71)*	**11.25 (4.97-25.46)***	11.44 (5.06-25.86)*
60-79	**17.81 (10.45-30.36)***	18.11 (10.64-30.84)*	**25.61 (11.31-57.96)***	25.54 (11.34-57.51)*
80+	**16.58 (8.77-31.35)***	17.34 (9.20-32.69)*	**28.61 (11.48-71.31)***	29.56 (11.71-74.62)*
BMI				
Neither	**1.00**	1.00	**1.00**	1.00
Overweight	**2.07 (1.58-2.71)***	2.00 (1.53-2.62)*	**2.13 (1.46-3.11)***	2.10 (1.43-3.08)*
Obese	**3.81 (2.90-5.00)***	3.74 (2.86-4.91)*	**4.27 (2.78-6.54)***	4.21 (2.73-6.49)*
Employment				
Employed	**1.00**	1.00	**1.00**	1.00
Unemployed	**2.13 (1.62-2.79)***	2.28 (1.75-2.96)*	**2.06 (1.44-2.95)***	2.01 (1.42-2.86)*
Physical Activity				
Active	**1.00**	1.00	**1.00**	1.00
Moderately	**1.39 (0.97-1.99)**	1.38 (0.97-1.98)	**1.44 (0.87-2.39)**	1.43 (0.86-2.38)
Inactive	**1.66 (1.21-2.27)***	1.65 (1.20-2.25)*	**2.20 (1.40-3.44)***	2.13 (1.35-3.34)*
Aboriginal Identity				
Aboriginal	**2.31 (1.66-3.21)***	2.45 (1.76-3.40)*	**2.69 (1.72-4.21)***	2.78 (1.78-4.35)*
Not Aboriginal	**1.00**	1.00	**1.00**	1.00

Notes:
* Significant p-value less than 0.05 alpha.
Reference: income -$50,000+, geography –urban, sex –female, age –<35, BMI –neither, employment –employed, physical activity –active, Aboriginal identity –Not Aboriginal.

Model 1: 'Unweighted' model of individual-level income adjusted for geography, sex, age, BMI, employment, physical activity, aboriginal identity.
Model 2: 'Unweighted' model of area-level income adjusted for geography, sex, age, BMI, employment, physical activity and Aboriginal identity.

Based on the 'unweighted' individual-level income model, individuals living in rural areas were 13% less likely of developing diabetes than their urban counterparts. For the 'unweighted' area-level income, individuals from rural areas were 18% less likely of developing diabetes than individuals from urban areas. Similarly, individuals from rural areas were 16% and 23% less likely of diabetes development compared to individuals from urban areas in the 'weighted' individual and area-level income models, respectively. However, these findings were statistically insignificant.

In the 'unweighted' models, males were approximately 1.5 times more likely of obtaining diabetes than females. Compared to the 'unweighted' models, the odds of males developing diabetes was lower (OR 1.4) compared to females in the design-based models.

According to age, the odds of diabetes increased from 4 to 17 as age increased from the '25-44' age category to the '80+' age category compared to the '<35' age category in the 'unweighted' models. In the 'weighted' models, the odds ratios increased with increasing age ranging from approximately 4 to 29.

In all models, odds ratios were higher for individuals with increasing BMI. Overweight individuals were approximately 2 times more likely of developing diabetes than individuals who were neither overweight nor obese. Obese individuals were about 4 times more likely of developing diabetes than individuals who were neither overweight nor obese.

For employment, individuals who were unemployed were at greater risk of diabetes than those who were employed. In all models, unemployed individuals were approximately 2 times more likely of obtaining diabetes than employed individuals.

Greater physical inactivity was associated with a greater likelihood of diabetes. In the 'unweighted' models, individuals who were moderately active had a 1.4 greater likelihood of diabetes than individuals who were physically active. Similarly, individuals who were inactive had a 1.7 greater likelihood of diabetes than individuals who were active. In the 'weighted' models, individuals who were overweight were approximately 1.4 times more likely of developing diabetes than individuals who were neither overweight nor obese. Similarly, individuals who were obese were nearly 2 times more likely of developing diabetes than individuals who were neither overweight nor obese.

With regards to Aboriginal identity, the 'unweighted' individual-level income model showed Aboriginal Peoples to be 2.3 times more likely of developing diabetes than non-Aboriginals. The 'unweighted' area-level income model revealed a slightly higher odds ratio of 2.5 for Aboriginals compared to non-Aboriginals. In the 'weighted' individual-level income model, Aboriginal Peoples had a higher odds of developing diabetes (OR 2.7) compared to the 'unweighted' models. Finally, the 'weighted' area-level income model showed Aboriginal Peoples were 2.8 times more likely of developing diabetes than non-Aboriginals.

CHAPTER 5:
DISCUSSION

Chapter 5 summarizes the findings of the research study and its relevance the context of the research topic. Evidence is provided from previous studies to support the findings of this study. The strengths and limitations are discussed and concluding remarks are made.

5.1 Summary of Findings and Discussion

The purpose of conducting this research study was to determine whether an area-based socioeconomic measure like income can be used as a proxy for individual-level data. Due to the difficulty in accessing readily available SES information in administrative databases, researchers are compelled to use area-based measures in place of individual-level SES despite some of the limitations in doing so (3,22-24). Nonetheless, this research study sought to compare the two measures using descriptive and multivariable analyses. The research questions of the study included: (1) What is the relationship between diabetes and income (at the individual and area-level) in the study sample and in the estimated population, (2) How much agreement exists between individual-level income and area-level income in the study sample and in the estimated population? How does the relationship differ in urban and rural areas? (3) Are diabetes inequalities underestimated when measured using area-level income compared to individual-level income in urban and rural Saskatchewan?

Based on Table 4.1 of the study population, a greater proportion of individuals in the study sample and in the estimated population were less than 35 year of age, non-Aboriginal, employed, partnered, physically inactive, neither overweight nor obese, living in urban areas, had a post-secondary education, had an area-level income of less than $30,000, and an individual-level income of more than $50,000. Although over half of individuals had a household income of more than $50,000, the areas in which they lived were characterized by low income.

5.1.1 Research Question #1: What Is The Relationship between Diabetes and Income (at the Individual and Area-Level) in the Study Sample and in the Estimated Population?

When each characteristic was stratified by the presence and absence of diabetes, proportions of diabetes were higher for males, individuals who were older, partnered, unemployed, obese, physically inactive, and living in urban areas. For individual-level income, a larger proportion of individuals with diabetes in the study sample had an income of less than $30,000, while a greater proportion of diabetics had an income of more than $50,000 in the estimated population. For area-level income, most individuals with diabetes were categorized as having less than $50,000.

Based on Table 4.2, graphs illustrated the percentage of individuals with diabetes for each income category. Although clear income gradients were observed for area-level income, individual-level income did not show the same patterns. For individual-level income, there was a decrease in the proportion of diabetics from the lowest income category to the middle income category, however there was an increase from the middle income category to the highest income category for the ‘unweighted’ and ‘weighted’ population as seen in Figure 4.1. The pattern could be attributed to a combination of factors such as older individuals having a higher proportion of diabetes than younger individuals, employment earnings increasing with age, and seniors having low income.

In contrast to Figure 4.1, gradients were observed for individual-level income, while patterns were present for area-level income in the study sample and in the estimated population as seen in Figure 4.3 on the percentage of diabetics within each income category. After the adjustment of the age variable, strong gradients were observed for individual and area-level income in the study sample and in the estimated population as seen in Figure 4.4 on age-adjusted rates. Despite area-level income having a different influence on diabetes than individual-level income, it is difficult to determine whether contextual effects exist as this study did not tease apart area-based gradient effects from individual income effects. Data limitations could have also contributed to the spike seen at the highest income category of ‘$50,000+.’

As seen in Figure 4.1 and Figure 4.3, observed patterns could have been due to data limitations in the way income categories were created. The highest income category of '$50,000+' is a very heterogeneous group and the aggregation of individual-level income could partially explain the patterns. As seen in Figure 4.2, the stratification of individual-level income into smaller income categories revealed a downward gradient until '$50,000-$59,999' and then small peaks at '$60,000-$79,999' and '$100,000+,' which could have been further disaggregated into smaller income categories. Since more than 90% individuals had an area-level income of less than $50,000 and about 5% had an income of more than $50,000 as seen in Table 4.1, individual income categories were defined in the same manner for comparison purposes and to comply with the cell size release regulations by the Health Quality Council of Saskatchewan.

When comparing income gradients between the study sample and the estimated population, gradients were larger in the study sample, which could be due to the under-representation of young individuals and the over-representation of older individuals in the study sample compared to the estimated population and the 'true' Census population. Although all groups of the Canadian population are covered and youth aged 12 to 19 years are over-sampled (4,83), the over-representation of older individuals (who have a greater proportion of diabetes) than younger individuals in the study sample is likely to be the driving force for the income-diabetes relationship. The utilization of sampling weights for the estimated population partially account for the selection differences in the study sample (4), which is why gradients are smaller in the estimated or 'weighted' population compared to the study sample or 'unweighted' population.

Overall, the patterns observed between the income and diabetes relationship in the study sample and in the estimated population are due to a combination of factors such as older individuals having a higher proportion of diabetes than younger individuals, employment income increasing with age, and seniors having low household incomes. Data limitations as mentioned previously may also play a role in the patterns observed. These factors in addition to the study sample containing selection bias have led to larger gradients in the study sample or 'unweighted' population compared to the estimated or 'weighted' population.

5.1.2 Research Question #2: How much agreement exists between individual-level income and area-level income in the study sample and in the estimated population? How does the relationship differ in urban and rural areas?

As seen in Table 4.3 and Table 4.4, there was a high discrepancy between individual and area-level income. The Kappa statistic for the study sample and the estimated population revealed relatively low agreement between the two measures. Although most individuals had an area-level income of less than $50,000, approximately half of individuals had an individual-level income of more than $50,000. This finding reiterates the notion that individuals live in low income areas, while maintaining higher household incomes. One possible explanation for the discrepancy between individual and area-level incomes is social desirability bias among those earning low incomes (91). For example, most individuals who responded with a total household income greater than $50,000 were classified as having an area-level income of less than $30,000 as seen in Table 4.3 and Table 4.4. Survey questions pertaining to sensitive topics like income may result in social desirability bias[21] in which low income individuals may respond with a higher income level (91). Although the huge difference seen in individual and area-level incomes cannot be fully attributed to social desirability bias, it may explain some of the disagreement. In addition, the CCHS is vulnerable to non-response bias in which non-respondents may possess characteristics not present in the responding group, however Statistics Canada takes steps to minimize non-response as outlined in section 5.2.2.a *Canadian Community Health Survey (CCHS)* (4).

The high discrepancy between individual and area-level income can also be attributed to measurement error (15,34). Area-based income is measured at an aggregate level in which the dissemination area, which contains approximately 125 to 144 households (26), may consist of residents who are very heterogeneous in terms of total household income (5,34). The dissemination area may not be an adequate spatial unit in which residents share similar SES characteristics, although these administrative boundaries may affect individual health via contextual effects (5). Other studies have also found poor agreement between individual and area-level SES measures (31,37,67). A study by Pardo-Crespo et al. (2013) found poor

[21] Social desirability bias is a form of response bias arising from survey data in which respondents respond untruthfully to present a favourable view of themselves to others (91).

agreement between individual and area-level SES measures in which the smallest spatial unit of the U.S. census, the block group, was used (5). The block group, which consists of about 1,000 individuals, is similar in size to the smallest spatial unit in the Canadian Census, the dissemination area (26,31).

When comparing urban and rural areas in the study sample and estimated population, patterns were seen for individual-level income in both urban and rural areas, while downward-sloping gradients were observed for area-level income in urban areas as seen in Figure 4.5 and Figure 4.6. In Figure 4.7 and Figure 4.8 on the percentage of diabetics within each income category in the study sample and in the estimated population, respectively, reverse patterns were observed contrary to Figure 4.5 and Figure 4.6. Income gradients were shown for individual-level income in urban and rural areas, while different patterns were seen for area-level income in urban areas. Contradictory to previous research (3,33,38), income had a significant effect in rural areas in which statistically significant odds ratios of diabetes were higher for each individual-level income category compared to odds ratios in urban areas as seen in Table 4.5. The proportion of diabetics within each individual-level income category for rural areas was generally greater than the proportion of diabetics within each individual-level income category for urban areas, which is consistent with other research that rural residents experience poorer health than their urban counterparts (7,8). One possible explanation could be the greater number of physicians and health facilities in urban areas compared to rural areas, allowing for greater accessibility to medical care in urban areas (77). However, the final model showed a decrease in odds of diabetes for rural residents compared to urban residents, but these findings were statistically insignificant.

Overall, there was relatively low agreement between individual and area-level income in the study sample and in the estimated population based on the Kappa statistics. As previously observed in Figure 4.1 and Figure 4.3, patterns and downward-sloping gradients were seen for both individual and area-level income when geography type was stratified. As mentioned previously, a combination of factors could have led to the patterns that were found such as individuals earning higher individual income with age, older individuals having the greater proportion of diabetes, and seniors having low income. As rural areas consist of an older population than urban areas (50), the higher proportion of diabetes in rural areas is consistent with previous research (7,8). One limitation of this study was the inability to compare area-level income between urban and rural areas due to low cell size in rural areas.

5.1.3 Research Question #3: Are Diabetes Inequalities Underestimated When Measured Using Area-Level Income Compared to Individual-Level Income in Urban and Rural Saskatchewan?

Based on Figure 4.9 and Figure 4.10 of the predicted probabilities, strong individual-level income gradients were seen in urban and rural areas for the study sample and the estimated population. However, patterns were observed for area-level income in urban areas. In summary, a strong association between individual-level income and diabetes was observed in urban and rural areas, however a weaker association was found with area-level income. Consistent with other research, area-based SES measures provide weaker associations between socioeconomic status and health compared to individual-level measures (27-30).

To achieve the primary and secondary objectives, multivariable analyses were conducted. The first and second objectives were as follows: (1) To determine whether area-based income can be used as a proxy for individual-level data, and (2) To determine whether area-level income has an effect on diabetes. Results of the final multivariable model showed a strong inverse relationship between individual income and the odds ratios of diabetes in the study sample, while a weaker relationship was found in the estimated population. The stronger association in the study sample could be a result of the over-representation of older individuals (who have a higher burden of diabetes) and the under-representation of individuals less than 35 years of age, despite the final model adjusting for age. For the area-level income models, a weaker gradient appeared whereby odds ratios were highest in the lowest income category, but were smaller in the middle income category compared to the highest income (reference) category, however the results were statistically insignificant.

Although area-level income influences diabetes, the pattern of influence varies from individual-level income. Other research has also shown area-based socioeconomic measures to have different patterns of association between SES and health compared to individual-level SES (5,27-31). However, it is difficult to determine whether a contextual effect (if any) exists since individual-level effects have not been separated from area effects in this study.

The relationship between individual and area-level income is a complex one. Individual and area-level SES have independent and may have interactional effects on health, however it is difficult to tease apart these interactions occurring at different levels (92). There is a debate as to whether area-level associations on health are due to individual-level characteristics of residents or area-level characteristics of the place in which residents live (92). Nonetheless, the SES of individuals and the neighbourhoods in which they live has an effect on individual health. A study by Winkleby et al. (2006) found low SES individuals living in high SES neighbourhoods had higher rates of mortality than better off individuals living in high SES neighbourhoods (75). The finding contradicts the collective resources theory that suggests low SES individuals living in high SES neighbourhoods may benefit from greater high-quality resources (75). One plausible mechanism suggested by Winkleby et al. (2006) is that low SES individuals may have high housing costs, resulting in less disposable income for goods and services such as healthy foods, transportation, and medical care (75). Secondly, the relative low social position of low SES individuals in high SES neighbourhoods may lead to greater stressors and fewer social supports to manage stress (75). Similarly, greater stressors exist with living in a low SES neighbourhood compared to a higher SES neighbourhood (64). Age-standardized death rates per 100,000 were higher for low SES individuals living in low SES neighbourhoods than for high SES individuals living in low SES neighbourhoods (75). Nonetheless, further investigation is needed to determine how area-level income influences diabetes.

Based on the entirety of the results, it can be suggested that there is relatively low agreement between individual and area-level income. Most individuals were categorized as having a low area-level income, although individual-level income showed otherwise. The high amount of disagreement between the two measures may have resulted in varying patterns in the association between income and diabetes, as suggested by Pardo-Crespo et al. (2013).

5.2 Study Strengths and Limitations

5.2.1 Strengths

There are several strengths of this research study. First, this study provides researchers with greater insight into using ABSMs like income as a proxy for individual-level income in Saskatchewan. It provides further evidence in the debate between the use and disuse of using area-level SES measures in the absence of individual-level data. This study examines diabetes inequalities in urban and rural Saskatchewan in which future research can examine the mechanisms leading to such inequalities. Second, this research study uses health administrative data to identity diagnosed diabetes cases, which does not contain response bias associated with survey data. Third, an ‘unweighted’ and ‘weighted’ CCHS population were compared to determine whether the sampling methods employed by Statistics Canada to obtain the study sample are generalizable to the entire Saskatchewan population. To the researcher’s knowledge, no studies have compared the study sample and the estimated population of the CCHS for Saskatchewan.

5.2.2 Limitations

Some of the limitations of this study include limitations of the data and the availability of data. One main limitation of this study is the way income categories were categorized. Since the purpose of this study was to compare individual and area-level income, income categories were categorized in the same manner for both measures. However, over 90% of individuals were categorized as having an area-level income of less than $50,000 and due to data release regulations, income categories were defined in the manner in which they were. The highest income category of $50,000 or more is a very heterogeneous income group, which may be one reason for the results seen between income and diabetes. Another confinement when categorizing income was using the pre-defined income categories for individual-level income in the CCHS.

Another limitation of this study is the unavailability of data to differentiate between type 1 and type 2 diabetes. To compensate for this shortcoming, the CCDSS definition excluded individuals younger than 1 years of age who are more likely to be diagnosed with type 1 diabetes (1,43). Second, only diagnosed individuals are identified in the database as having diabetes. Although it is possible that individuals who have not visited a physician may have diabetes, the CCDSS definition of diabetes has been found to be reliable at identifying diabetes cases with 86% sensitivity for Newfoundland and Labrador (93,94) .

A limitation synonymous to the Canadian Census and the CCHS is the cross-sectional nature of both surveys. Despite the biases arising from self-report survey data such as response bias (4), the cross-sectional design of both surveys leads to temporality issues in which there is uncertainty of whether exposures precede health outcomes (95,96). Another limitation apparent in the Census is the suppression of area-level income for dissemination areas with a population of less than 250 (82). These dissemination areas may provide valuable information on socioeconomic inequalities in health, which vary from the other dissemination areas.

5.2.2.a Canadian Community Health Survey (CCHS)

The Canadian Community Health Survey (CCHS) has its own limitations. The survey contains sampling error as well as non-sampling errors. Sampling error refers to the error arising from sampling a portion of the population rather than surveying the entire population (4). The CCHS minimizes sampling error by using an adequate sample, a sample containing a variety of characteristics, a complex sampling design and estimation method in order to infer the target population (4). To measure the size of sampling error for each estimate, coefficient of variation tables are used in conjunction with release guidelines (4).

The CCHS is also vulnerable to non-sampling errors, which can arise at any stage of the survey process (4). For example, such errors include incorrect data entry by the interviewer, recall bias and response bias by the respondent, and errors occurring from data processing (4). Non-sampling errors are minimized in the CCHS through interviewer training, monitoring of interviewers, interviewers following-up with non-respondents, and protocols that prevent out-of-range values in the computer-assisted interviewing (CAI) application (4). One source of non-sampling error is non-response bias, which is more difficult to control. Non-response bias arises from non-response of single CCHS questions or total non-response in which the respondent either refused to participate in the survey or the respondent cannot be contacted (4). Partial non-response is unlikely since survey participants usually completed the questionnaire with very little item-non-response (4). At the national level, the response rate for 2007 was 78%, while the response rate was 75% in 2008 (97).

5.3 Conclusion and Study Implication

This study provides greater insight into using ABSMs as a proxy for individual-level data. Specifically, individual-level income should be used (if available) when measuring diabetes inequalities in Saskatchewan. If individual-level data is absent, researchers investigating the effects of individual-level income on diabetes in Saskatchewan should exercise caution when using area-level income. Several biases arise from utilizing area-based measures as proxies for individual-level data such as the modifiable areal unit problem, fallacies arising from making particular inferences, and the misclassification of individuals when assigning individuals to area-level income (15,66). In this study, area-level income was found not to be a good proxy for individual-level income in which estimates were significantly lower for area-level income. Since socioeconomic status changes across time and place (24), it is suggested that researchers conduct a study prior to utilizing area-based measures to determine whether area-level SES significantly differs from individual-level data.

It was also found that area-level income has a varying effect on diabetes development separate from individual-level income. Although the mechanisms behind the place effects (if any) are unknown, qualitative research can provide a greater understanding between the income-diabetes relationship. Since dissemination areas are based on administrative boundaries, qualitative research can offer greater insight into how residents define their neighbourhood or surrounding areas (58). In addition, qualitative research provides a historical account of places, which plays an important role in health (58). Future research comparing individual and area-based socioeconomic measures would benefit from incorporating a mixed-methods approach to understand the mechanisms influencing health.

REFERENCES

(1) Public Health Agency of Canada. 2011 Diabetes in Canada: Facts and Figures from a Public Health Perspective. 2011:1-112.

(2) Canadian Diabetes Association, Diabète Québec. Diabetes: Canada at the Tipping Point: Charting a New Path. 2011:3-58.

(3) Pampalon R, Hamel D, Gamache P, Raymond G. A Deprivation Index for Health Planning in Canada. Chronic Disease in Canada 2009;29(4):178-191.

(4) Statistics Canada. Canadian Community Health Survey (CCHS): 2007 Microdata Files User Guide. 2008; Available at: http://www23.statcan.gc.ca/imdb-bmdi/pub/document/3226_D7_T9_V4-eng.pdf. Accessed June/17, 2014.

(5) Pardo-Crespo M, Narla N, Williams A, Beebe T, Sloan J, Yawn B, et al. Comparison of individual-level versus area-level socioeconomic measures in assessing health outcomes of children in Olmsted County, Minnesota. Journal of Epidemiology and Community Health 2013;67(4):305-310.

(6) Ross CE, Wu C. Education, Age, and the Cumulative Advantage in Health. Journal of Health and Social Behaviour 1996;37(1):104-120.

(7) Galobardes B, Shaw M, Lawlor DA, Lynch JW, Smith GD. Indicators of Socioeconomic Position (Part 1). J Epidemiol Community Health 2006;60:7-12.

(8) Galobardes B, Lynch J, Smith GD. Measuring Socioeconomic Position in Health Research. British Medical Bulletin 2007;81,82:21-37.

(9) Braveman P. Monitoring Equity in Health and Healthcare: A Conceptual Framework. J Health Popul Nutr 2003;21(3):181-192.

(10) Braveman P, Kumanyika S, Fielding J, LaVeist T, Borrell NL, Manderscheid R, et al. Health Disparities and Health Equity: The Issue is Justice. American Journal of Public Health 2011;101(S1):S149-S155.

(11) Braveman P, Gruskin S. Defining Equity in Health. J Epidemiol Community Health 2003;57:254-258.

(12) Kawachi I, Subramanian SV, Almeida-Filho N. A Glossary For Health Inequalities. J Epidemiol Community Health 2002;56:647-652.

(13) Kawachi I, Subramanian SV, Almeida-Filho N. A Glossary for Health Inequalities. J Epidemiol Community Health 2002;56:647-652.

(14) Marmot M. Black to Acheson: Two Decades of Concern With Inequalities in Health. A Celebration of the 90th Birthday of Professor Jerry Morris. International Journal of Epidemiology 2001;30:1165-1171.

(15) Schuurman N, Bell N, Dunn JR, Oliver L. Deprivation Indices, Population Health and geography: An Evaluation of the Spatial Effectiveness of Indices at Multiple Scales. Journal of Urban Health 2007;84(4):591-603.

(16) Pampalon R RG. A Deprivation Index for Health and Welfare Planning in Quebec. Chronic Diseases in Canada 2000;21(3):104-113.

(17) Metcalf P, Scragg R, Schaaf D, Dyall L, Black P, Jackson R. Comparison of different markers of socioeconomic status with cardiovascular disease and diabetes risk factors in the Diabetes, Heart and Health Survey. The New Zealand Medical Journal 2008;121(1269):45-56.

(18) Dinca-Panaitescua S, Dinca-Panaitescub M, Bryantc T, Daiskid I, Pilkingtond B, Raphaela D. Diabetes Prevalence and Income: Results of the Canadian Community Health Survey. Health Policy 2011;99:116-123.

(19) Rabi D, Edwards A, Southern D, Svenson L, Sargious P, Norton P, et al. Association of socio-economic status with diabetes prevalence and utilization of diabetes care services. BMC Health Services Research 2006;6(124):1-7.

(20) Booth G, Hux J. Relationship Between Avoidable Hospitalizations for Diabetes Mellitus and Income Level. Arch Intern Med 2003;163(1):101-106.

(21) Braveman P, Cubbin C, Egerter S, Chideya S, Marchi K, Metzler M, et al. Socioeconomic Status in Health Research: One Size Does Not Fit All. Journal of the American Medical Association 2005;294(22):2879-2888.

(22) Krieger N, Chen J, Waterman P, Soobader M, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). Journal of Epidemiology and Community Health 2003;57:186-199.

(23) Krieger N, Chen J, Waterman P, Rehkopf D, Subramanian S. Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures—The Public Health Disparities Geocoding Project. American Journal of Public Health 2003;93(10):1655-1671.

(24) Krieger N, Chen J, Waterman P, Soobader M, Subramanian SV, Carson R. Geocoding and Monitoring of US Socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-based Measure and Geographic Level Matter? American Journal of Epidemiology 2002;156(5):471-482.

(25) Pickett KE PM. Multilevel Analyses of Neighbourhood Socioeconomic Context and Health Outcomes: A Critical Review. J Epidemiol Community Health 2001;55:111-122.

(26) Godley J, Haines VA, Hawe P, Shiell A. Small Area Contextual Effects on Self-Reported Health: Evidence from Riverside, Calgary. BMC Public Health 2010;10(264).

(27) Pampalon R, Hamel D, Gamache P. A Comparison of Individual and Area-Based Socio-Economic Data for Monitoring Social Inequalities in Health. Health Reports 2009;20(3):85-94.

(28) Soobader M, LeClere F, Hadden W, Maury B. Using Aggregate Geographic Data to Proxy Individual Socioeconomic Status: Does Size Matter? 2001;91(4):632-636.

(29) Janssen I, Boyce W, Simpson K, Pickett W. Influence of individual- and area-level measures of socioeconomic status on obesity, unhealthy eating, and physical inactivity in Canadian adolescents. American Journal of Clinical Nutrition 2006;83:139-145.

(30) Cesaroni G, Farchi S, Davoli M, Forastiere F, Perucci C. Individual and Area-Based Indicators of Socioeconomic Status and Childhood Asthma. European Respiratory Journal 2003;22(4):619-624.

(31) Geronimus A, Bound J. Use of Census-based Aggregate Variables to Proxy for Socioeconomic Group: Evidence from National Samples. American Journal of Epidemiology 1998;148(5):475-486.

(32) Kawachi I BL editor. Neighbourhoods and Health. 1st ed. New York, USA: Oxford University Press Inc.; 2003.

(33) Pampalon R, Hamel D, Gamache P. Health Inequalities in Urban and Rural Canada: Comparing Inequalities in Survival According to an Individual and Area-Based Deprivation Index. Health & Place 2010;16:416-420.

(34) Mustard CA, Derksen, S., Berthelot J, Wolfson M. Assessing Ecologic Proxies for Household Income: A Comparison of Household and Neighbourhood Level Income Measures in the Study of Population Health Status. 1999;5:157-171.

(35) Subramanian S.V., Chen J.T., Rehkopf D.H., Waterman P.D., Krieger N. Comparing individual- and Area-Based Socioeconomic Measures for the Surveillance of Health Disparities: A Multilevel Analysis of Massachusetts Births, 1989-1991. American Journal of Epidemiology 2006;164:823-834.

(36) Hanley GE MS. On the Validity of Area-Based Income Measures to Proxy Household Income. BMC Health Services Research 2008;8(79).

(37) Marra CA, Lynd LD, Harvard SS, Grubisic M. Agreement between Aggregate and Individual-Level Measures of Income and Education: A Comparison Across Three Patient Groups. BMC Health Services Research 2011;11(69).

(38) Pampalon R, Hamel D, Gamache P. Recent changes in the geography of social disparities in premature mortality in Quebec. Social Science & Medicine 2008;67(8):1269-1281.

(39) Brown AF, Ettner SL, Piette J, Weinberger M, Gregg E, Shapiro MF, et al. Socioeconomic Position and Health among Persons with Diabetes Mellitus: A Conceptual Framework and Review of the Literature. Epidemiologic Reviews 2004;26:63-77.

(40) Krieger N. Proximal, Distal, and the Politics of Causation: What's Level Got to Do With It? Public Health 2008;98(2):221-230.

(41) Canadian Diabetes Association, Diabète Québec. At the Tipping Point: Diabetes in Saskatchewan. 2011:1-3.

(42) Cox M., Boyle P.J., Davey P.G., Feng Z., Morris A.D. Locality Deprivation and Type 2 Diabetes Incidence: A Local Test of Relative Inequalities. Social Science & Medicine 2007;65(9):1953-1965.

(43) Garner R., Carriere G., Sanmartin C., Longitudinal Health and Administrative Data Research Team. The Health of First Nations Living Off-Reserve, Innuit, and Metis Adults in Canada: The Impact of Socio-economic Status on Inequalities in Health. Statistics Canada: Health Analysis Division 2010;82-622-X(004):1-19.

(44) Saskatchewan Health. The Provincial Diabetes Plan. 2004:1-55.

(45) Canadian Diabetes Association. The Cost of Diabetes in Saskatchewan. 2011:1-6.

(46) Hearst M, Sirard J, Forsyth A, Parker E, Klein E, Green C, et al. The relationship of area-level sociodemographic characteristics, household composition and individual-level socioeconomic status on walking behavior among adults. Transportation Research Part A 2013;50:149-157.

(47) Tang M, Chen Y, Krewski D. Gender-related differences in the association between socioeconomic status and self-reported diabetes. . International Journal of Epidemiology 2003;32(3):381-385.

(48) Dyck R., Klomp H., Tan L.K., Turnell R.W., Boctor M.A. A Comparison of Rates, Risk Factors, and Outcomes of Gestational Diabetes Between Aboriginal and Non-Aboriginal Women in the Saskatoon Health District. Diabetes Care 2002;25(3):487-493.

(49) Saskatoon Health Region. Saskatoon Health Region Annual Report: 2013-2014. 2014:1-69.

(50) How Healthy Are Rural Canadians? An Assessment of Their Health Status and Health Determinants. Ottawa: Canadian Institute for Health Information; 2006.

(51) Rural Health Status in Saskatoon Health Region. Saskatoon: Saskatoon Health Region; 2009.

(52) Stronks K, van de Mheen H, van den bos J, Mackenbach J. The Interrelationship between Income, Health and Employment Status. International Journal of Epidemiology 1997;26(3):592-600.

(53) Lynch JW, Davey Smith G, Kaplan GA, et al. Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. BMJ 2000;320:1200-1204.

(54) Lorgelly P, Lindley J. What is the Relationship Between Income Inequality and Health? Evidence from the BHPS. Health Economics 2008;17:249-265.

(55) Stafford M. MM. Neighbourhood Deprivation and Health: Does It Affect Us All Equally? International Journal of Epidemiology 2003;32:357-366.

(56) Kawachi I, Adler N, Dow D. Momey, Schooling, and Health: Mechanisms and Causal Evidence. Annals of the New York Academy of Sciences 2010;1186:56-68.

(57) Wilkinson R, Pickett K. Income Inequality and Population Health: A Review and Explanation of the Evidence. Social Science & Medicine 2006;62:1768-1784.

(58) Kawachi I. Income Inequality and Health. In: Berkman L, Kawachi I, editors. Social Epidemiology New York, United States: Oxford University Press Inc.; 2000. p. 76-94.

(59) Scheffler, R.M., Petris, N.C., Borgonovi, F., Brown T.T., Sassi, F., Dourgnon, P., Sirven, N. Social Capital, Human Capital and Health: What is the Evidence? : Centre for Educational Research and Innovation; 2010. p. 1-44.

(60) Kawachi I, Kennedy B, Glass R. Social Capital and Self-Rated Health: A Contextual Analysis. American Journal of Public Health 1999;89(8):1187-1193.

(61) Kawachi I. Social Capital and Community Effects on Population and Individual Health. Annals New York Academy of Sciences 1999;896:120-130.

(62) Subramanian S, Kawachi I. Income Inequality and Health: What Have We Learned So Far? Epidemiologic Reviews 2004;26:78-91.

(63) Wilkinson R. Income Inequality, Social Cohesion, and Health: Clarifying The Theory –A Reply to Mutaner and Lynch. International Journal of Health Services 1999;29(3):525-543.

(64) Adler N, Rehkopf D. U.S. Disparities in Health: Descriptions, Causes, and Mechanisms. Annual Review of Public Health 2008(29):235-252.

(65) Oliver LN HM. Does Choice of Spatial Unit Matter for Estimating Small-area Disparities in Health and Place Effects in the Vancouver Census Metropolitan Area? Canadian Journal of Public Health 2007;98(1):S27-S34.

(66) Diez Roux AV. Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis. American Journal of Public Health 1988;88(2):216-222.

(67) Diez Roux AV, Kiefe CI, Jacobs JDR, Haan M, Jackson SA, Nieto FJ, et al. Area Characteristics and Individual-Level Socioeconomic Position Indicators in Three Population-Based Epidemiologic Studies. Annals of Epidemiology 2001;11(6):395-405.

(68) Geronimus A, Bound J, Neidert L. On the Validity of using Census Geocode Characteristics to Proxy Indiviudal Socioeconomic Characteristics. Journal of the American Statistical Association 1996;91(434):529-537.

(69) Riva M, Gauvin L, Barnett TA. Toward the Next Generation of Research into Small Area Effects on Health: A Synthesis of Multilevel Investigations Published Since July 1998. Journal Epidemiol Community Health 2007;61:853-861.

(70) Cummins J, Skyers S, Oosthuizen R, Poorman J. Building Community Cohesion Into Area Based Initiatives. United Kingdom: Community Cohesion Unit; 2004.

(71) Macintyre S., Ellaway A., Cummins S. Place Effects on Health: How Can We Conceptualise, Operationalise, and Measure Them? Social Science & Medicine 2002;55:125-139.

(72) Lebel A, Pampalon R, Villeneuve PY. A Multi-Perspective Approach for Defining Neighbourhood Units in the Context of a Study on Health Inequalities in the Quebec City Region. International Journal of Health Geographics 2007;6(27).

(73) Galster G. On the Nature of Neighbourhood. Urban Studies 2001;38(12):2111-2124.

(74) O'Campo P. Invited Commentary: Advancing Theory and Methods for Multilevel Models of Residential Neighborhoods and Health. American Journal of Epidemiology 2003;157(1):9-13.

(75) Winkleby M, Cubbin C, Ahn D. Effect of Cross-level Interaction between Individual and Neighborhood Socioeconomic Status on Adult Mortality Rates. American Journal of Public Health 2006;96(12):2146-2153.

(76) Gary T, Safford M, Gerzoff R, Ettner S, Karter A, Beckles G, et al. Perception of Neighborhood Problems, Health Behaviors, and Diabetes Outcomes Among Adults With Diabetes in Managed Care. Diabetes Care 2008;31(2):273-278.

(77) Bocquier A, Cortaredona S, Nauleau S, Jardin M, Verger P. Prevalence of treated diabetes: Geographical variations at the small-area level and their association with area-level characteristics. A multilevel analysis in Southeastern France. Diabetes and Metabolism 2011;37:39-46.

(78) Saskatchewan Health Quality Council. Data Quality of the Data Holdings From SK Health. 2013:1-5.

(79) Saskatchewan Ministry of Health -Epidemiology & Research Unit (Population Health Branch). Health Services Databases: Information Document. 2010:1-17.

(80) Saskatchewan Health Quality Council. Quality of Diabetes Management in Saskatchewan: Technical Appendix on Methods and Analysis. 2006:7.

(81) Statistics Canada: Census Operations Division. 2006 Census Dictionary. 2010;Catalogue no. 92-566-X:1-578.

(82) Statistics Canada. Data Quality and Confidentiality Standards and Guidelines (Public): Data Quality Practices. 2008; Available at: http://www12.statcan.gc.ca/census-recensement/2006/ref/notes/DQ-QD/DQPractices-QDPratiques-eng.cfm. Accessed June/17, 2014.

(83) Statistics Canada. Canadian Community Health Survey -Annual Component. 2013; Available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226. Accessed 4/4, 2014.

(84) Statistics Canada. Overview of the Census: Census Year 2011. 2011;98-302-XIE:1-46.

(85) Public Health Agency of Canada -Chronic Disease Surveillance Division. Report from the Canadian Chronic Disease Surveillance System: Hypertension in Canada, 2010. 2010;HP32-4/2010E-PDF:1-26.

(86) Statistics Canada. Canadian Community Health Survey (CCHS) 2008 (Annual Component) and 2007-2008 Derived Variable (DV) Specifications: Master and Share Files. 2009:1-186.

(87) Statistics Canada. *Saskatchewan (Code47)* (table). *2006 Community Profiles*. 2006 Census. Statistics Canada Catalogue no. 92-591-XWE. . Ottawa. March 13, 2007; Available at: http://www12.statcan.ca/census-recensement/2006/dp-pd/prof/92-591/index.cfm?Lang=E. Accessed 05/22, 2014.

(88) Statistics Canada. Table 105-0501 - Health indicator profile, annual estimates, by age group and sex, Canada, provinces, territories, health regions (2012 boundaries) and peer groups, occasional, CANSIM (database). 2009; Available at: http://www5.statcan.gc.ca/cansim/a47. Accessed 05/22, 2014.

(89) Statistics Canada. Table 105-4009 Body mass index (BMI), by sex, household population aged 18 and over excluding pregnant females, Canada, provinces and territories, occasional(1,2,3,4,5,6), CANSIM (database). 2009; Available at: http://www5.statcan.gc.ca/cansim/a05. Accessed 05/22, 2014.

(90) Statistics Canada. Population and dwelling counts, for Canada, provinces and territories by the Statistical Area Classification, 2011 and 2006 censuses (table). Population and Dwelling Count Highlight Tables. 2011 Census. Statistics Canada Catalogue no. 98-310-XWE2011002.

2012; Available at: http://www12.statcan.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=601&PR=47&S=0&O=D&RPP=25. Accessed 05/22, 2014.

(91) Podsakoff P, MacKenzie S, Podsakoff N, Lee J. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. . Journal of Applied Psychology 2003;88(5):879-903.

(92) Jackson C, Richardson S, Best N. Studying place effects on health by synthesising individual and area-level outcomes. Social Science & Medicine 2008;67(12):1995-2006.

(93) Roche M, Wang P. Sex Differences in All-Cause and Cardiovascular Mortality, Hospitalization for Individual With and Without Diabetes, aand Patients Diagnosed with Diabetes Early and Late. Diabetes Care 2013;36(9):2582-2590.

(94) Public Health Agency of Canada. Diabetes in Canada:Highlights from the National Diabetes Surveillance System 2004 – 2005. 2008;HP32-2/2005E-PDF:1-19.

(95) Cummins S, Macintyre S, Davidson S, Ellaway A. Measuring Neighbourhood Social and Material Context: Generation and Interpretation of Ecological data from Routine and Non-Routine Sources. Health and Place 2005;11(3):249-260.

(96) Kestenbaum B. Chapter 4: Cross-sectional Studies. Epidemiology and Biostatistics New York: Springer; 2009.

(97) Murray C, Walsh G, Gorber S. A Comparison between Atlantic Canadian and National Correction Equations to Improve the Accuracy of Self-Reported Obesity Estimates in Atlantic Canada. Journal of Obesity 2012:1-7.

APPENDIX A

Figure A1: Data Linkages –CCHS and Admin Data

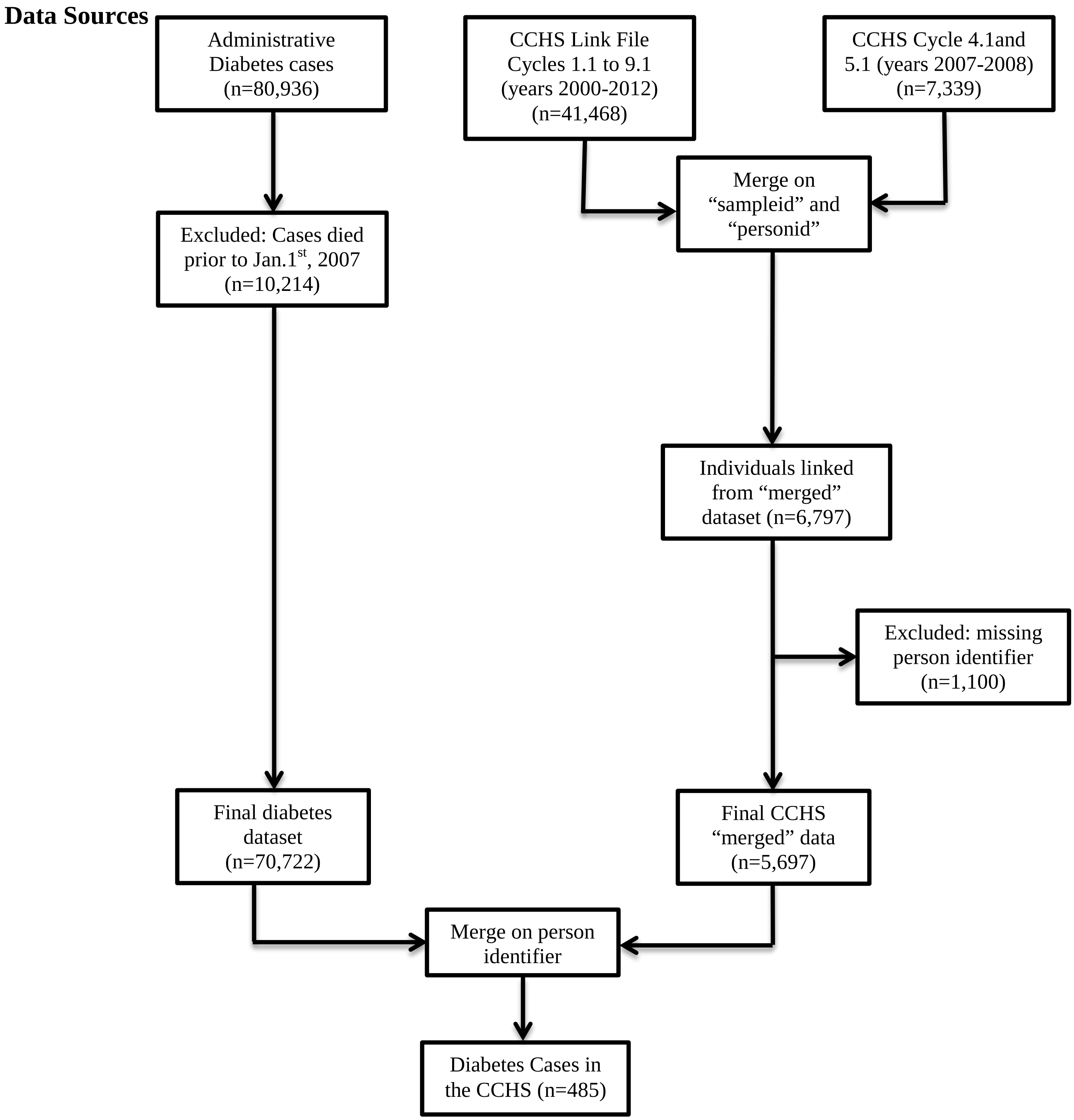


Figure A2: Data Linkages –Census and CCHS-Admin Data