The Impact of Sustainable Transit Availability on Health Inequality in Canadian Cities, 2006 to 2016

Emma Gunther

A Thesis in the Department of Economics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Economics) at

Concordia University

Montréal, Québec, Canada

August 2023

© Emma Gunther, 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the Thesis prepared

By: Emma Gunther

Entitled: The Impact of Sustainable Transit Availability on Health Inequality in Canadian Cities, 2006 to 2016

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Economics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

______________________________ Examiner
Dr. Panagiotis Margaris

______________________________ Supervisor
Dr. Axel Watanabe

Approved by: ______________________________
Dr. Christian Sigouin
Graduate Program Director

Date: _______________ ______________________________
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

The Impact of Sustainable Transit Availability on Health Inequality in Canadian Cities, 2006 to 2016

Emma Gunther

Concordia University

August 30, 2023

Abstract

Focus on sustainable transit has grown in recent years, as Canada invests in active and public transport (Infrastructure Canada 2023), and plans to reduce CO2 pollution (Government of Canada 2022). While the health benefits of commuting by walking, biking or even public transit may seem clear, the impact on health inequality within cities is less so; some researchers claim that strong transit systems equalize access to health care (Abu-Qarn and Lichtman-Sadot 2022), while others argue that uneven implementation of sustainable transit may lead to gentrification in transit-accessible neighbourhoods, leading to worse outcomes for vulnerable residents (Tehrani, Wu, and Roberts 2019). This investigation seeks to determine the impact of sustainable transit availability on health inequality in Canadian cities using cross-sectional regression analysis. We use the gap in the hospitalization rate between the highest and lowest income quintiles as a proxy for health inequality. Walkability is found to be related to a smaller gap, while bikability is associated with a wider gap. This may be explained by bikability and transit being associated with gentrification, mitigating any positive effects they may have had on the gap in hospitalizations.

Acknowledgements

I would like to acknowledge and express thanks for the guidance and support of my supervisor Dr. Axel Watanabe. His willingness to share time and knowledge to help me with this project has greatly enhanced the quality of this work. I have been fortunate to work under his supervision and benefit from his feedback.

Contents

List of Figures

List of Tables

1 Introduction

Interest in sustainable transit (public transit, walking or biking) is increasing for a multitude of reasons. During the COVID-19 pandemic, Montreal added several hundred kilometers of active transit corridors (Remiorz 2020), and many other cities did the same (The Centre for Active Transportation 2020). These are also attractive as Canada plans to lower carbon emissions by 40% by 2030 (Tasker 2022). As the world moves to sustainable transit, it is important to understand what some downstream impacts might be. Given the longer wait times and even closures occuring in emergency rooms across Canada (Miller and Bamaniya 2023), it is of paramount importance to consider how increased reliance on sustainable transit may impact Canadians' health.

Transit accessibility is also an equity issue. In the first quarter of 2023, the average price of a new car in Canada was \$60,000, while a used one was \$39,000 (Friedman 2023). The median after-tax income for a Canadian family was \$68,400 in 2021 (Statistics Canada 2023a), making a vehicle a difficult purchase, which presents a major problem in a car-dependent area. Significant savings can be achieved by using public transportation, however. In 2019, the average expenditure by a Canadian household was \$11,258 for private transportation, or \$1,479 for public transportation (Statistics Canada 2021) (households probably use one mode or the other, so these numbers should be interpreted with some caution). Transportation is hard on carbon budgets as well, contributing the equivalent of 5,000kg of carbon dioxide to the average Canadian's footprint. This is compared to 1,970kg for someone living in Japan, which has a higher quality public transit system (Bernstien 2021).

A desire for less car dependance and carbon emissions motivates investment in sustainable transit infrastructure. This infrastructure may impact health in several ways. It can induce more physical activity or allow easier access to health inputs for low income individuals. It can also increase exposure to pollution, and adding such infrastructure to neighbourhoods may prompt gentrification. These impacts may be felt differently by individuals at varying income levels.

This thesis will test how sustainable transit impacts the gap in the hospitalization rate for low- and high-income individuals. The hospitalization rate may be impacted in two ways by transit infrastructure. Transit infrastructure may decrease the probability of being ill by prompting individuals to engage in more physical activity and improving access to health inputs such as grocery stores, leading to fewer hospitalizations. Transit infrastructure may also increase the likelhiood of seeking healthcare by making it easier to get to a healthcare provider. By examing only hospitalizations for severe health problems, we attempt to eliminate the second effect and focus here on the first. Some attention will be paid to the mechanisms by which transit impacts the hospitalization rate (i.e. physical activity and gentrification) to the extent that it supports the understanding of how this gap is determined. However, it must be acknowledged that health is extremely complex. It is determined by many factors, some measurable and some not, of which transit infrastructure is only one. No model can single-handedly account for all these factors and explain something so complicated.

The motivation behind this paper is to understand the impact of sustainable transit availability on health inequality in Canadian cities. This study uses the gap in hospitalization rate between the lowest and highest income quintiles as a proxy by which to measure health. Previous work on the subject has largely been theoretical, or merely grazed over the topic as part of a much larger study.

The most comprehensive examination so far is the work of Abu-Qarn and Lichtman-Sadot (2022), which studies the impact of the implementation of new bus lines in Isreal on socioeconomic and geographic health disparities. The authors measure health status using data from a repeated cross-sectional survey which provides self-reported assessments of health. They find that the bus lines improve access to secondary health care, and result in better health among the region's disadvantaged. This research will differ from that of Abu-Qarn and Lichtman-Sadot (2022) in several ways. More precise data on health outcomes is used; while Abu-Qarn and Lichtman-Sadot (2022) examine "chronic health conditions" as a group, we use data on hospitalizations by type. Although they are able to directly observe the health of a cross-section of the population, where we extrapolate from observations of those reporting to hospital, there is an advantage of relying only on diagnoses determined and reported by health professionals. The weakness of relying on self-reported information is evidenced by the results of Abu-Qarn and Lichtman-Sadot (2022) which show that diagnoses increase after access to secondary health care is improved, implying that diagnoses were previously underreported as individuals did not know that they were suffering from a given condition. We also include measures such as walkability, bike infrastructure, and access to transit stops, whereas they focused exclusively on bus lines.

It is important to include walkability in the study, as it is uncorrelated with transit availability, yet both are needed to paint a picture of how one might get around without a private vehicle. There is a moderate correlation between bikeability and transit availability, but both will be analyzed to gain an understanding of each. In part of a larger study, Badland and Pearce (2019) find different directions of impact for different modes of transportation, so it is important to understand the impacts of each type. It is especially interesting because walking and biking have a more direct impact on a person's health than public transit use, although all three are associated with higher activity levels according to Freeman et al. (2013).

Walkability is found to be associated with a reduced gap, while bikeability is shown to have the opposite direction of impact and transit has no relationship at all. It is possible that any potential positive effect that transit and bike infrastructure may have had on the hospitalization gap is negated by their impact on affordability, as both are found to be associated with gentrification.

This paper will procede as follows: Section 2 explores the literature. Section 3 presents the theoretical model. Section 4 outlines the methodology and Section 5 the data used. Section 6 explores the results, which are discussed in Section 7. Section 8 concludes.

2 Literature Review

Despite the noted importance of physical environment for healthcare consumption, Duru and Paelinck (1990) "insist on the limitation of the present approach" of macro models in health economics to explain the relationship (pg. 4). More recently, Deryugina and Molitor (2021) have noted that the relationship between place and health is established, but not fully understood. Although previous literature examines how transit options affect communities, and how neighbourhood characteristics impact health inequality, results vary and a clear understanding has not emerged.

This paper follows Abu-Qarn and Lichtman-Sadot (2022) in seeking to determine how transport options impact health disparities. The authors examine the impact of expanded healthcare access due to implementation of new bus lines in Israel. Using regression analyses, they show that self-reported diagnoses among the disadvantaged increase shortly after new bus lines are introduced, but decrease in the long term. The authors suggest that individuals initially are unaware of undiagnosed conditions, leading to more reported health problems in the short run, but eventually become healthier. As our study relies on diagnoses from health professionals rather than self-reporting, we do not expect to see such under or overreporting distortions arising from individuals not reporting a condition that they are not aware they have or misdiagnosing themselves. Their investigation shows that in the long term, reported health outcomes are the same or better after the bus line introduction, meaning that there is not a strong negative health effect from public transportation due to, for example, exposure to pollution. They conclude that introducing new bus lines improved healthcare access, thereby reducing health inequality.

I follow the theory of Bernard et al. (2007), who find that neighbourhoods facilitate access to health inputs. These inputs shape residents' health and social functioning by determining their exposure to parks and healthy food stores or pollution and liquor stores. Fuller, Gauvin, and Kestens (2013), however, note that while a relationship between access to healthy food stores and physical activity facilities has been found, the result has also been questioned.

Other authors have examined similar questions, with varying results. Gorman et al. (2003) finds positive health impacts when a high level of funding is put towards public transit, and Mueller et al. (2020) found that the shift to active and public transit prompted by the implementation of Superblocks in Barcelona added 200 days to residents' life expectancy. Maizlish et al. (2013) predict a 13% reduction in premature deaths under their active transit scenario, although they also estimate a health risk from road traffic injuries and pollution exposure. Prince et al. (2022)

finds an association between active transit and physical activity in OECD countries. Turrell et al. (2013) argue that it is "plausible" that walkability contains health inequality. Cullen, Cummins, and Fuchs (2012) note that relying heavily on cars can also increase mortality, due to harmful small particulates generated by vehicles.

On the other hand, Badland and Pearce (2019) find different directions of impact for different modes of transportation. They conclude that there is a need to monitor the impact on the vulnerable when considering liveability policies, and that uneven implementation may exacerbate inequalities. Tehrani, Wu, and Roberts (2019) also question whether such policies have the intended impact on vulnerable residents. They find that public transit infrastructure meant to reduce inequality may reinforce it, by causing gentrification and leading to mixed impacts on health. Fuller, Gauvin, and Kestens (2013) find that low income residents of Montreal tend to live significantly closer to public transit and bike share programs, although they note that Montreal has relatively small disparities in access to transport compared to other cities.

A potential shortcoming of this study is that it relies on the presence of transit amenities, rather than the rate at which they are used. Shareck, Frohlich, and Kestens (2014) theorize that walkability may not translate into a high walking rate. Herrmann et al. (2017) present supporting evidence, arguing that walking rate depends on factors such as tree canopy cover, the number of parking lots in the area and sidewalk setback from roads. However, Brown et al. (2013) and Hirsch et al. (2013) both present evidence that higher WalkScores lead to higher walking. Freeman et al. (2013) also find that walkability has a positive impact on engagement in active travel, but note an inverse relationship for non-Hispanic whites in high-income neighbourhoods.

Freeman et al. (2013) and Deryugina and Molitor (2021) note a risk of using cross-sectional analysis for such a study. Both acknowledge that geographic variation in health may be due to the personal characteristics of residents (for example, someone who values physical activity will choose a neighbourhood that facilitates this). However, Freeman et al. (2013) explain that this risk is partially mitigated as neighbourhood preferences are constrained by factors such as socioeconomic, ethnic and racial housing patterns. This issue is even smaller in this study, as census subdivisions are used as a unit of analysis, which are larger areas than neighbourhoods, meaning that residential patterns due to personal preferences will be less pronounced.

Previous work has employed a variety of methods for estimating levels of health and health inequality, including self-reported surveys (Abu-Qarn and Lichtman-Sadot 2022), life expectancy (Deryugina and Molitor 2021; Cullen, Cummins, and Fuchs 2012), and measuring physical activity (Prince et al. 2022; Turrell et al. 2013). Others use health impact assessments (Gorman et al. 2003; Mueller et al. 2020) or conceptual framework to project changes in health (Badland and Pearce 2019), meaning they estimate expected changes in health in response to policy changes based on previous work. This study aims to add empirical evidence to the existing body of literature.

We use data on hospitalization rates by income quintile, which has the advantage of being more objective than self-reported health levels, and more complete than measuring physical activity. Life expectancy is a fairly robust measure with a low risk of underreporting, but we argue that there is a clearer link between hospitalization in a given year and an individual's current place of residence, whereas early mortality may be related to a condition developed due to an individual having lived in poor conditions decades prior.

There is ambiguity in the literature that we hope to resolve here. Although the work of Badland and Pearce (2019) and Tehrani, Wu, and Roberts (2019) provides useful insight about what may impact health inequality, their lack of empirical evidence is a shortcoming that this work will address. This study will provide new evidence to the knowledge base created by Abu-Qarn and Lichtman-Sadot (2022) and Gorman et al. (2003), while expanding on their findings by including walking and biking.

3 Theoretical Model

I consider a government with progressive preferences rewarding an equitable distribution of health profiles in their population. This is reasonable as one of Quebec's policy objectives is to attain "comparable standards of health and welfare in the various strata of the population and in the various regions" (Québec 2020), and most countries with public healthcare systems have stated similar goals (Glied and Smith 2013). As such, we assume the government wishes to minimize the gap in health between high- and low-income individuals.

The government provides transit infrastructure, t, which can be decomposed into walking, biking and public transit infrastructure. Here we wish to examine how health inequality varies with the level of transit infrastructure available. To this end, it is assumed that cities are endowed with different, exogenously determined, levels of t, and that migration between cities is not possible. We also assume that each city has the same number of residents.

There are two types of consumers, with exogenously determined income levels, m_{high} and m_{low}. They have homogenous preferences regarding health $h_i \in \mathbb{R}_{++}$ (i = low or $high$), and other goods $z_i \in \mathbb{R}_+$. They live in cities to facilitate consumption (Glaeser, Kolko, and Saiz 2001), and derive utility from consumption according to the function:

$$u(h_i, z_i), \tag{1}$$

where preferences are well-behaved. An individual's demand for health is described by:

$$h(m_i, t), \tag{2}$$

where h_i is increasing in m_i. The nonlinear relationship between income and health is well-documented (Wilkinson 1992; Rehkopf et al. 2008; Rehnberg and Fritzell 2016), as such h_i is expected to increase with m_i at a decreasing rate.

The cost of health to an individual is determined by the level of t available in their city and the level of health they desire, where expenditure is not necessarily linear in either h_i or t. This is described by the expenditure function $x(h_i, t)$. The price of z_i is normalized to 1. Higher t may also impact the price of z, but for tractability we keep this price fixed here. Consumption choices are therefore constrained by consumers' budgets as follows:

$$m_i = z_i + x(h_i, t). \tag{3}$$

The level of t available may impact $x(h_i, t)$ in several ways:

1. Walkability, bikeability and public transit infrastructure can all increase an individuals' level of physical activity (Brown et al. 2013; Hirsch et al. 2013; Freeman et al. 2013; Maizlish et al. 2013), improving their health regardless of whether they pay for access to a gym or similar facility.

2. Transit facilitates access to primary health care (Abu-Qarn and Lichtman-Sadot 2022) and other health inputs such as grocery stores (Bernard et al. 2007) without the need for a private vehicle.

In both cases, an increase in t reduces the level of expenditure required to achieve a given level of health, or $\partial x(h_i, t)/\partial t < 0$.

3. There is an increased risk of traffic-related injuries and exposure to air pollution, which may lead to higher health-related expenses.

4. Implementation of transit infrastructure has also been theorized to lead to gentrification, thereby increasing the effective price of health. In reality this might also impact the price of z, but as mentioned above we fix this price here for tractability.

The latter two cases mean that higher expenditures are required to achieve a given level of health in cities with more transit infrastructure t (i.e. $\partial x(h_i, t)/\partial t > 0$). Empirically, the positive impacts of t seem to outweigh the negative (Abu-Qarn and Lichtman-Sadot 2022; Freeman et al. 2013; Gorman et al. 2003). As such, we assume the price to an individual of achieving their desired level of health decreases as t increases ($\partial x(h_i, t)/\partial t < 0$).

The level of t available is common among residents of the same city, but income levels are heterogenous. This study seeks to understand the impact transit infrastructure t has not on health achievement of individuals, but rather the inequality in health achievement I between high and low income individuals who reside in the same city. This gap is measured by:

$$I(t) = \frac{h(m_{high}, t)}{h(m_{low}, t)}. \tag{4}$$

The government wishes to minimize this gap, as discussed above.

The data available to model health inequality measures hospitalization rates per 100,000 residents by income quintile. This is not a measure of health $h(m_i, t)$, but rather the rate of residents' hospital visits $v(m_i, t)$. Hospital visits occur according to two factors. The first is the likelihood $p(s, m_i, t)$ of an individual suffering from an illness of (self-diagnosed) severity $s \in S \subseteq \mathbb{R}$ in a given year[1]. Health generally increases with income (Wilkinson 1992; Rehkopf et al. 2008; Rehnberg and Fritzell 2016), and therefore we assume that $p(s, m_{high}, t) < p(s, m_{low}, t)$. The second is the likelihood $q(s, m_i, t)$ of that individual seeking treatment when they are suffering from the aforementioned health condition. The expected rate of hospital visits is therefore:

$$\bar{v}(m_i, t) = \int_{s \in S} p(s, m_i, t) q(s, m_i, t) ds. \tag{5}$$

The rate of visits $v(\cdot)$ is impacted by the level of t in two ways. First, as discussed by Point 1 and 2 above, a higher level of t in a city induces higher levels of physical activity among its residents. This reduces the likelihood of illness $p(\cdot)$ and therefore is expected to reduce visits $v(\cdot)$. We assume that this will have a higher impact among residents with m_{low}, who become able to increase their physical activity without needing to pay for access to a fitness facility. A high level of transit infrastructure t also increases the likelihood of visiting the hospital when sick $q(\cdot)$, particularly among low income residents who rely on public transit. The rate of hospital visits $v(\cdot)$ does not differentiate between these opposing effects. As such, it is unclear how $v(\cdot)$ relates to $h(\cdot)$ above; less visits may indicate better health h due to increased physical activity (increased $p(\cdot)$), or more visits may mean that consumers are choosing to consume more healthcare due to easier access (increased $q(\cdot)$), thereby improving their health.

To isolate the effect of $p(\cdot)$, we assume there exists a threshold severity level $\underline{s}$ above which $q(s, m_i, t) = 1$ for any m_i and t. Healthcare visits for illness with $s \geq \underline{s}$ are assumed to be less sensitive to transit access than those for routine care where $s < \underline{s}$ (consider an individual who may skip appointments to monitor a chronic health issue because it is too difficult to get to the appointment, but who will call an ambulance if they are experiencing a heart attack). This also limits distortions that may arise from certain types of individuals having an intrinsic desire to consume more healthcare (an individual who has a preference for presenting frequently for routine

[1]For simplicity we assume consumers only suffer from a given condition once per year.

care will nevertheless only be recorded as being hospitalized for a heart attack if they meet the criteria for this diagnosis). We assume the effect of $q(\cdot)$ on $v(\cdot)$ to be absent in this case, as an individual is unlikely to use public transit for urgent healthcare. As such, when removing hospital visits for $s < \underline{s}$, we can compute $v(\cdot)$ as:

$$\hat{v}(m_i, t) = \int_{s \geq \underline{s}} p(s, m_i, t) ds. \tag{6}$$

To further control for severity, we assign a weight $f(s)$ to each diagnosis type. The empirical values for these weights are given in Table 4 in Section 5. The function $v(m_i, t)$ is modified as:

$$\tilde{v}(m_i, t) = \int_{s \geq \underline{s}} p(s, m_i, t) f(s) ds \tag{7}$$

to better represent individual's health status. Health $h(m_i, t)$ is expected to be negatively correlated with $\tilde{v}(m_i, t)$. We assume the gap in health among high and low income quintiles to be the inverse of the same gap in hospitalization rates, as follows:

$$\tilde{I}(t) = \frac{\tilde{v}(m_{high}, t)}{\tilde{v}(m_{low}, t)} = \frac{1/h(m_{high}, t)}{1/h(m_{low}, t)} = \frac{1}{I(t)}. \tag{8}$$

The gap in health achievement is low when $I(t)$ is low, or $\tilde{I}(t)$ is high. This is a restrictive assumption; despite efforts to minimize distortions, $\tilde{v}(m_i, t)$ remains an imperfect estimate of health inequality. Note also that this assumes that the number of residents is the same across all cities, which is too strong an assumption. This will be addressed in the empirical model by regressing the gap on population size.

We made the assumption above that increasing the level of t led to a lower likelihood of illness $p(\cdot)$ and fewer visits $v(\cdot)$, and that this was more impactful for residents with m_{low}, who became able to increase their physical activity at a low cost. If this is not the case, we expect $\tilde{I}(t)$ to be a constant function, as t influences hospital visits of low- and high-income individuals equally, leaving the ratio unchanged.

The remainder of this thesis concerns itself with assessing the null hypothesis that $\tilde{I}(t)$ is a constant function. Where data availability permits, some attention will be paid to determining the mechanisms behind this impact, as enumerated by Points 1-4 above.

4 Data

To approximate health achievement, we use a dataset from the Canadian Institute for Health Information (2019a), or CIHI, titled "Measuring Trends in Health Inequalities in Cities: Hospitalization and Day Surgery Indicator Results, by Census Metropolitan Area and Census Subdivision - Data Tables". For grouped fiscal years, this dataset provides hospitalization rates by type and income quintile. The data is available at either the Census Metropolitan Area (CMA) level, or the Census Subdivision (CSD) level, which is smaller. For each income quintile and type of hospitalization, the dataset provides the crude number of people hospitalized, the number of people in the quintile, and adjusted indicators including the rate per 100,000 people, the age-standardized rate per 100,000 people (described below) along with confidence limits, the rate ratio with confidence limits, and the rate difference with confidence limits. CSD inclusion in the CIHI dataset depends on a population based cut-off. There are 263 CSDs in the dataset, 165 of which do not report by quintile and were dropped from the analysis. All observations marked by CIHI as "Use with caution" were also removed. These observations are those for which the coefficient of variation (ratio of the standard deviation and the rate) is deemed to be high (Canadian Institute for Health Information 2019b). It should be noted that hospitalization rate is influenced by many factors and therefore is not a perfect standard by which to measure health. This is discussed further in Sections 3 and 5.

We use the age-standardized rate of hospitalizations per 100,000 residents, which are based on five-year age groupings using 2011 Census population data[2]. The crude hospitalization rate per age group is calculated as:

$$\text{Crude rate per age group} = \frac{\text{Numerator}}{\text{Denominator}}, \tag{9}$$

where the numerator is the number of people hospitalized in a given income quintile in a CSD, and the denominator is the number of people in that quintile overall, according to Canadian Institute for Health Information (2019b). The crude rate for each age group is adjusted by the weight of that age group in the standard population, and multiplied by 100,000 as:

$$\text{Age-specific weighted rate} = \text{Crude rate per age group} \times \text{Weight of age group} \times 100,000. \tag{10}$$

The age-standardized rate is then:

$$\text{Age-standardized rate} = \sum \text{Age-specific weighted rates}, \tag{11}$$

according to Canadian Institute for Health Information (2022).

Income quintiles are based on average before-tax income per single-person equivalent in a dissemination area. Dissemination areas are the smallest geographic areas for which census infomation is disseminated, and represent "a block in urban areas or an area bounded by roads in rural areas" (Statistics Canada 2020). Income levels for dissemination areas were calculated using 2006

[2]For heart attack and stroke indicators, the youngest age group is 18-24, followed by 5 year intervals onwards

and 2016 census data (income information was not collected in the 2011 census), but only income quintiles, not levels, are included in the data. Note that CSDs do not each report all hospitalization indicators for all time periods; however what they report is always reported for each income quintile. To avoid inconsistency, the method for calculating the hospitalization gap only compares income quintiles within one CSD.

Table 1 shows age-standardized hospitalizations by type and CSD. This table was created using CIHI variables *CSD name*, *Indicator*, *Neighbourhood income quintile* and *Age-standardized rate (per 100,000 population)*. All fiscal years are included, and observations for which the CSD name was not available are dropped.

Table 1: Age-standardized number of hospitalizations per 100,000 residents by type and by income quantile in Canadian census subdivisions

Hospitalization type	Income Quantile				
	1	2	3	4	5
Childhood Dental Caries	53,104	35,376	28,496	22,581	26,262
Falls Injury Hospitalization	105,989	59,833	56,014	47,623	53,460
Opioid Poisoning	777	433	303	261	299
Alcohol	20,654	10,507	8,284	6,333	6,246
Angina	3,194	2,366	1,957	1,825	1,599
Asthma	4,010	2,687	2,169	1,851	1,628
Chronic Obstructive Pulmonary Disease	20,928	11,197	8,199	5,914	4,417
Congestive Heart Failure	6,575	3,962	2,930	2,545	2,076
Diabetes	5,893	4,036	2,817	2,246	2,240
Grand Mal & Epileptic Convulsions	2,600	1,850	1,528	1,327	1,411
Hypertension	491	348	263	209	163
Heart Attacks	18,440	16,200	13,971	12,978	13,296
Stroke	7,721	6,075	5,592	4,862	4,856
Injury	167,127	102,995	95,236	85,219	97,439
Motor Vehicle Traffic Injury	3,873	2,979	2,715	2,350	2,498
Self-Injury	4,180	2,646	1,926	1,558	1,507

Note: The above table shows the age-standardized number of hospitalizations by 100,000 residents divided by reason for hospitalization and income quintile of resident.

To measure the gap in hospitalization rates $\tilde{I}(t)$, the age-standardized hospitalization rate per 100,000 residents in the highest income quintile is divided by the same value for the lowest for each hospitalization type in each census subdivision. This number includes all fiscal years. A higher ratio $\tilde{I}(t)$ represents a smaller gap between income quintiles, where a value closer to zero represents a larger gap. Approximately 7% of the observations in the sample (including all hospitalizations) have a ratio above 1, meaning high-income residents are hospitalized at a higher rate than low-income residents. Of these, 67% have above-median walkability and 72% have above-median median income in the CSD. Their bikeability and transit scores are distributed similarly to the sample as a whole.

Data from The Canadian Urban Environmental Health Research Consortium (CANUE) provides indicators describing urban environments. This study includes their neighbourhood datasets for bikeway comfort and safety (DMTI Spatial Inc, various years; Beairsto et al. 2022), public transportation (DMTI Spatial Inc 2019; OpenStreetMap contributors 2019), proximity measures (DMTI Spatial Inc, various years; Statistics Canada, Canadian Mortgage and Housing Corporation 2020), active living environments (Ross et al. 2018), and gentrification (Firth et al. 2020) which are described in more detail in Table 2. With the exception of the active living environment (ALE) data, which considers intersection density, dwelling density and points of interest, all datasets were only available for the year noted in Table 2, which unfortunately do not match CIHI years (2006-2016). ALE data was available for 2006 and 2016, and 2016 data is used as it was closer to the reporting years for other CANUE variables.

Briant, Combes, and Lafourcade (2010) warn of distortions that may arise due to choice of geographic unit. They find that the size and shape of zones employed influences results, particularly for large zones. The "Modifiable Areal Unit Problem" (MAUP) is reduced when data is spatially auto-correlated and averaged, but exacerbated when regression variables are not computed under the same aggregation process. The unit of analysis in this study is census subdivision (CSD), as this is the finest geographic region available in the CIHI data.

To connect health data at the CSD level with transporation data at the postal code level, we use Statistics Canada's Postal Code Conversion Files (PCCF) (Statistics Canada 2017). The PCCF links postal codes to CSDs using November 2021 postal codes and 2016 census geography. Postal codes do not respect census geography, so in some cases they are matched to multiple CSDs, and may even cross provincial boundaries.

CSD median income is included in the model as income quintiles are calculated at the CSD level rather than the national level. Smaller hospitalization gaps are expected in wealthier areas due to the nonlinear relationship between income and health (Wilkinson 1992; Rehkopf et al. 2008; Rehnberg and Fritzell 2016). Income data comes from 2016 Census data on household income statistics (Statistics Canada 2016b). Since CIHI income quintiles are calculated before tax, we use median total income of households before taxes. Statistics Canada reports median income for two adjacent CSDs with the same name for Langley (British Columbia), Moncton (New Brunswick) and North Vancouver (British Columbia) (Statistics Canada 2016b), whereas CIHI combines these adjacent CSDs into one observation (Canadian Institute for Health Information 2023). In these

Table 2: CANUE variables

Variable	Description
Low Comfort Infrastructure	Kilometers of low comfort bike infrastructure in 1km buffer (2021)
Medium Comfort Infrastructure	Kilometers of medium comfort bike infrastructure in 1km buffer (2021)
High Comfort Infrastructure	Kilometers of high comfort bike infrastructure in 1km buffer (2021)[1]
Bus Stops at 1000m	Number of bus stops within a 1000 meter buffer of a postal code (2019)[2]
Pharmacy or drug store	Binary variable indicating a pharmacy or drug store within 1km on foot (2019)
Grocery store	Binary variable indicating a grocery store within 1km on foot (2019)
Childcare	Binary variable indicating a childcare facility within 1.5km on foot (2019)
Primary Education	Binary variable indicating a primary education facility within 1.5km on foot (2019)
Secondary Education	Binary variable indicating a secondary education facility within 1.5km on foot (2019)[3]
ALE Class	Categorical value representing Active Living Environment characteristics, based on intersection and dwelling density and points of interest (2016)[4]
Gentrified census tract	Binary variable for gentrification according to Freeman (2016)[5]

[1]Canadian Bikeway Comfort and Safety Classification System (Can-BICS) metrics by CANUE. Bikeway data by OpenStreetMap contributors (2022).
[2]Accessed via the CANUE Data Portal: https://www.canuedata.ca/
[3]Statistics Canada/CMHC proximity data by DMTI postal code provided by CANUE
[4]Canadian Active Living Environments Index (Can-ALE) by DMTI postal code provided by CANUE.
[5]GENUINE: Gentrification, Urban Interventions, and Equity data were provided by CANUE.

cases, a population-weighted average of the two median incomes reported by Statistics Canada is calculated, using their population counts (Statistics Canada 2016a). We also include a dummy variable for province to capture the impact of differing healthcare systems and tax schemes across the country.

Finally, the impact of CSD population on the hospitalization gap is tested, as we otherwise assume the gap to be the same regardless of the number of residents if the regressors take the same value. The amount of transit infrastructure available may be correlated with population, as such we test which better explains the gap in hospitalizations. CSD population is provided by a Statistics Canada (2023b) dataset. As above, in some cases CIHI combines CSDs that Statistics Canada reports separately. In these cases, the population estimates have been added together to match the CIHI list of CSDs.

5 Methodology

The consideration of health inequality is guided by Asada (2005). This framework includes three requirements for measuring health inequality; defining when a distribution becomes inequal, defining a strategy to understand the chosen concepts of equity, and defining how to quantify equity.

We make no judgement regarding what is an acceptable level of inequality; the goal is to understand why CSDs have larger or smaller hospitalization gaps. As required by the CIHI data, cross-sectional analysis is performed on hospitalization rates in census subdivisions across Canada. To quantify health inequality, a ratio was calculated by dividing the hospitalization rate among the top income quintile by that of the lowest. We take the natural logarithm of this ratio to normalize the distribution.

Postal code and census subdivision data are matched by generally following the method of "Matching Census Data to Postal Codes using SPSS" (n.d.), allowing CANUE and CIHI datasets to be combined. From the CIHI (hospitalization) data, we drop observations that do not report the census subdivision name as this is the geographic unit used for the analysis. Summary statistics for all model variables are shown in Table 3.

To create a walkability index, we combine variables indicating proximity to pharmacies, grocery stores, childcare, and primary and secondary education facilities. These variables come from a dataset created by Statistics Canada and the Canadian Mortgage and Housing Corporation (CMHC). This results in an index from 1 to 5 for each postal code, but as the analysis is at the census subdivision level, the average is calculated for each CSD. Some of these variables measure presence of the amenity within 1km of the dissemination block, and some 1.5km (see Table 2). There is fuzziness in the distances regardless, as distance is measured from the dissemination block, not an individual's address. While no justification is offered by StatsCan or CMHC, generally it is reasonable to travel longer for a twice daily commute (to reach childcare or an education facility) than for a trip to the pharmacy or grocery store. Similar indicators are available for public parks and libraries, but these were not included as they are not essential to everyday life as the others are. It is important to use composite measures for walkability as "single components of neighborhood walkability often show fewer effects and are much more inconsistent in terms of significance and direction of effect" (Duncan 2013, pg. 245).

A variable was created for kilometers of bike infrastructure using indicators "Kilometers of Low Comfort Infrastructure in 1km buffer", "Kilometers of Medium Comfort Infrastructure in 1km buffer" and "Kilometers of High Comfort Infrastructure in 1km buffer" described in Table 2. These indicators were added together to get the number of kilometers available to each postal code, and the average number of kilometers available to a postal code was calculated for each CSD. Note that this yields the average kilometers within the buffer of each postal code, not the average number of kilometers in the postal code. This means that some stretches of bike infrastructure are double counted as they are in the buffers of two or more postal codes.

The CANUE variable “Number of Bus Stop at 1000m” was used to calculate the average number of bus stops within 1000m of a postal code in each CSD. As above, some bus stops belong to the buffers of several postal codes.

Similarly, the ALE class variable from CANUE was used to calculate the average ALE class per postal code in a CSD.

The relationship between the regressors and gentrification is also modeled, as we hypothesized in Point 4 of Section 3 that implementing transit infrastructure may trigger gentrification, thereby impacting the level of expenditure $x(h_i, t)$ required to achieve the desired health status. The gentrification variable needs no manipulation. Four gentrification variables are available from CANUE, each a binary variable decided by a different academic. We use gentrification according to Freeman (2016) as it has the highest number of observations.

CIHI calculates income quintiles at the CSD level, meaning someone in the lowest income quintile in a wealthy CSD may be better off than someone in the same quintile in a poorer CSD. A nonlinear relationship has been found between income and health in high-income countries (Wilkinson 1992; Rehkopf et al. 2008; Rehnberg and Fritzell 2016), with a stronger association between health and income at the low end of the income distribution. As such, a variable for median income of each CSD is included in the model. A smaller gap in hospitalizations (or higher $\tilde{I}$) is expected in CSDs with high median incomes. The natural logarithm of median income is taken to normalize the distribution.

The impact of CSD population size is also tested as transit infrastructure is expected to vary between bigger and smaller population centers. The natural logarithm of population size is taken to normalize the distribution. A smaller gap in hospitalizations is predicted in larger population centers, as a larger quantity and spread of health inputs are expected.

A dummy variable for the province in which the CSD is located is included, as healthcare is a provincial responsibility in Canada, meaning that services vary by province. Income distributions also vary by provincial tax regime. For example, the lowest tax rate in Quebec in 2020 was 15%, but only 5.05% in Ontario (Milligan 2021). Differences in mortality have been shown to be driven by regional differences in healthcare delivery and quality (Cullen, Cummins, and Fuchs 2012).

There is a potential issue with this study, in that health is estimated by hospitalization rate. A low rate may mean individuals are healthy, or that they face barriers to accessing care, such as a lack of transit infrastructure making it too costly to go to a hospital. A high rate may mean individuals are sick, or that their transportation costs are low so they seek care frequently. A lack of transit infrastructure or low transportation costs impact the likelihood of seeking treatment $q(\cdot)$. To control for this, we separate out high severity hospitalizations, assuming that severe hospitalizations are

Table 3: Descriptive Statistics for model variables

Variable	Minimum	Mean	Median	Maximum	Std. Dev.
Ratio	0.012	0.565	0.517	3.908	0.389
Median Income	50,227	76,276	77,282	119,900	14,523
Walk	0.188	0.503	0.487	1.043	0.132
Transit	0	18.107	14.373	63.021	14.436
Bike	0.019	4.019	3.644	9.712	1.996
ALE Class	0.893	2.368	2.309	3.839	0.565
CSD size	8,401	444,990	223,180	2,819,399	563,960

the least elastic. It is expected that someone experiencing a severe health issue will find a way of presenting to hospital, such as calling an ambulance. Similarly, this method should eliminate the effect of excess visits from individuals with a preference for receiving frequent care.

We adapt the weights published by Salomon et al. (2012), who surveyed the public regarding their judgements of health losses for a variety of diagnoses. To observe only severe hospitalizations, observations for hospitalization types with a below-median hospitalization weight are removed (this corresponds to restricting observations to those for which $s \geq \underline{s}$ as described in Section 3). This leaves a sample with only the more urgent health conditions. Table 4 shows the weights assigned by Salomon et al. (2012), which range from 0 (full health) to 1 (death), and those used in the regressions presented in this study. Salomon et al. (2012) uses more precise diagnoses categories than are available from CIHI, so the mean value of all relevant diagnoses is used for the analysis. In some cases there was no match between CIHI hospitalization types and the diagnoses published by Salomon et al. (2012), in which case the closest reasonable diagnoses are used.

Table 4: Severity weights s of health issues

Salomon et al. (2012) diagnosis	Weight	CIHI hospitalization type	Mean weight
Angina pectoris: mild, moderate, severe	0.037, 0.066, 0.167	Angina	0.09
Asthma: partially controlled	0.027	Asthma	0.0795
Asthma: uncontrolled [1]	0.132		
Heart failure: mild, moderate, severe	0.037, 0.070, 0.186	Congestive Heart Failure	0.0977
Diabetic foot	0.023	Diabetes	0.061
Diabetic neuropathy	0.099		
Epilepsy: treated, recent seizures	0.319	Grand Mal Status, Epileptic Convulsions	0.465
Epilepsy: untreated	0.420		
Epilepsy: severe [2]	0.657		
Heart failure: severe	0.186	Hospitalized Heart Attacks	0.186
Stroke: mild	0.021	Hospitalized Strokes	0.303
Stroke: moderate	0.076		
Stroke: moderate, cognition problems	0.312		
Stroke: severe	0.539		
Stroke: severe, cognition problems	0.567		
Injuries [3]	0.003 - 0.673	Falls Injury Hospitalization	0.15384
		Injury Hospitalization	0.15384
		Self-Injury Hospitalization	0.15384
		Motor Vehicle Traffic Injury Hospitalization	0.15384
Heroin and other opioid dependence	0.641	Hospitalizations Due to Opioid Poisoning	0.641
Alcohol use disorder: mild, moderate, severe	0.259, 0.388, 0.549	Hospitalizations Entirely Caused by Alcohol	0.3987
Chronic respiratory diseases: mild, moderate, severe	0.015, 0.192, 0.383	COPD	0.197
Dental caries: symptomatic	0.012	Day Surgery for Childhood Dental Caries	0.012
Heart failure: mild	0.037	Hypertension	0.037

[1] Controlled asthma is not included, as CIHI data only includes asthma leading to hospitalization.
[2] Seizure free epilepsy is not included, as CIHI data only includes epilepsy leading to hospitalization.
[3] All diagnoses in this category are included, except those relating to burns, crush injury and drowning.

A linear regression model is used to assess the impact of regressors listed above. There is a moderate correlation between bike kilometers and bus stops within 1000m, so the two variables are not modelled together. The model for the hospitalization gap is therefore given by:

$$ln(\tilde{I}) = \beta_0 + \beta_1 Walk + \beta_2 Transit + \beta_3 ln(Income) + \beta_4 Dummy_{Province} + \epsilon \tag{12}$$

$$ln(\tilde{I}) = \beta_0 + \beta_1 Walk + \beta_2 Bike + \beta_3 ln(Income) + \beta_4 Dummy_{Province} + \epsilon, \tag{13}$$

where β_0 is the intercept, β_i are coefficients and ϵ is the error term. $\tilde{I}$ is the proxy for health inequality, given by:

$$\tilde{I} = \frac{\text{Hospitalization rate in highest income quintile}}{\text{Hospitalization rate in lowest income quintile}}. \tag{14}$$

Hospitalizations are age-standardized and given as rates per 100,000 residents.

We wish to determine if the coefficients for transit indicators in Equations 12 and 13 simply indicate an effect due to living in a big city, where transit is presumed to be more available. We therefore run a similar regression which includes CSD population size instead of transit infrastructure as given by:

$$ln(\tilde{I}) = \beta_0 + \beta_1 ln(Size) + \beta_2 ln(Income) + \beta_3 Dummy_{Province} + \epsilon \tag{15}$$

Positive coefficients for the Walk, Bike and Transit indicators are expected, following from the research of Abu-Qarn and Lichtman-Sadot (2022). The ratio $\tilde{I}$ given by Equation 14 is expected to increase with ln(Income) and ln(Size) as well. A higher magnitude of coefficients is expected for severe hospitalizations compared to all hospitalizations, as there are likely to be some distortions relating to the likelihood of seeking care $q(\cdot)$ in the results for lower-elasticity hospitalizations due to lack of hospital accessibility.

To determine to what extent improvements to the hospitalization gap are brought about by increased physical activity (as hypothesized in Point 1 in Section 3), $\tilde{I}$ is regressed on a variable describing the active living environment (ALE) class. This measures neighbourhood features found to induce physical activity according to Ross et al. (2018), such as proximity to points of interest. This relationship is modelled by:

$$ln(\tilde{I}) = \beta_0 + \beta_1 ALE + \beta_2 ln(Income) + \beta_3 Dummy_{Province} + \epsilon. \tag{16}$$

Positive impacts from transportation infrastructure not attributable to ALE class are assumed to be due to improved access to health inputs.

To test the relationship between the above regressors and gentrification (Point 4 from Section

3), they are regressed on a binary indicator for gentrification using a probit model. This can be done at the postal code level as this model only includes CANUE variables. It is therefore not necessary to take the average of the walk index across a subdivision, allowing it to be treated as a categorial variable counting the amenties available within walking distance from zero to five.

$$G = \beta_1 Walk + \beta_2 Transit + \epsilon \tag{17}$$

$$G = \beta_1 Walk + \beta_2 Bike + \epsilon, \tag{18}$$

where G is the measure representing gentrification.

Unfortunately the data to directly measure health impacts from pollution exposure or traffic related injuries is not available (Point 3 from Section 3). Any negative impacts not explained by gentrification are assumed to be attributable to these factors.

6 Results

Table 5 shows that bike kilometers and bus stops within 1000m are moderately correlated, so they cannot be included in the same regression equation. As such, Table 6 shows results for Equations 12 and 13 separately.

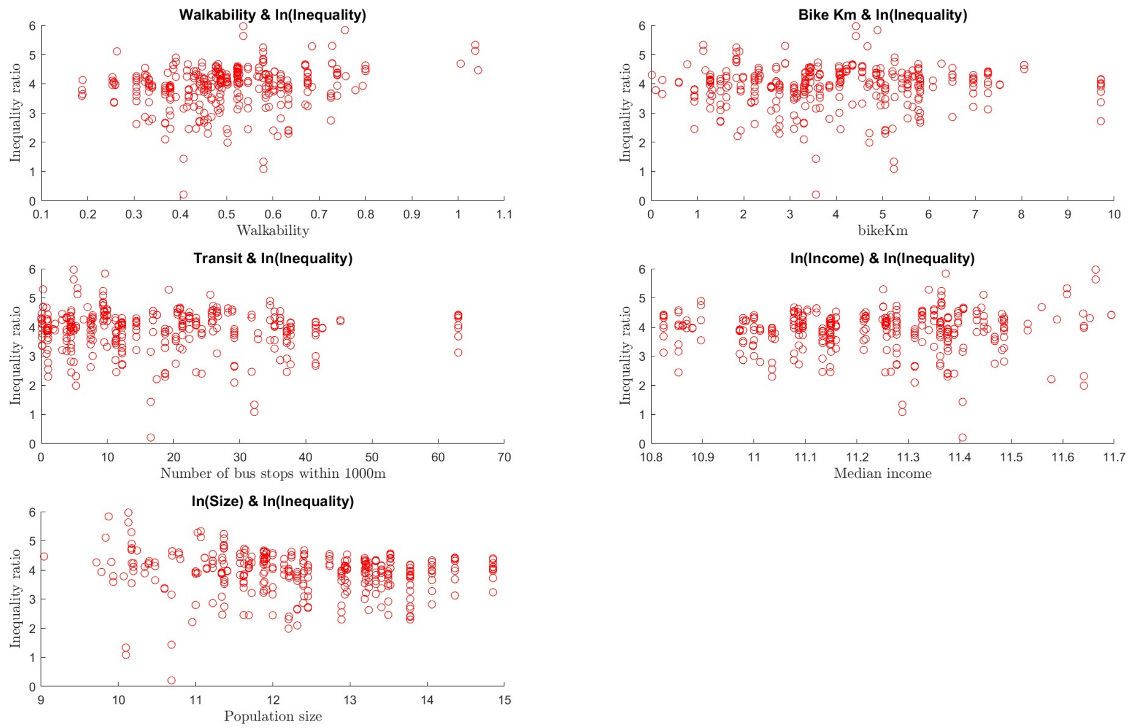
Table 5: Correlation Matrix of regressors

	Bus	Walk	Bike	ln(Income)	ln(Size)	ALE Class
Bus	1	-0.12	0.50*	-0.26	0.25	0.64*
Walk	-0.12	1	0.02	0.40	0.10	-0.04
Bike	0.50*	0.02	1	-0.07	0.28	0.57*
ln(Income)	-0.26	0.40	-0.07	1	-0.09	-0.39
ln(Size)	0.25	0.10	0.28	-0.09	1	0.50*
ALE Class	0.64*	-0.04	0.57*	-0.39	0.50*	1

A coefficient of 0.5-0.7 is considered to be moderate positive (negative) correlation (Mukaka 2012). These values are marked by *.

Table 1 shows that hospitalizations tend to be highest among the lowest income quintile, and they decrease as income rises, although they level off between the fourth and fifth quintiles. This supports research that shows that the impact of income on health dimishes as income rises (Rehkopf et al. 2008). Figure 1 shows relationships between the regressors and the hospitalization gap, using all hospitalization types.

Figure 1: Scatter plots for variables and $ln(\tilde{I})$



The above are basic scatter plots showing values for one variable and the corresponding values for ln(Inequality).

The ratio of age-standardized hospitalizations per 100,000 residents between individuals in the 1st and 5th income quintiles was regressed on built environment characteristics, median income, and province of the CSD, as represented by Equations 12 and 13. Each model was tested using all hospitalization types and with severe hospitalizations only. The results are shown in Table 6.

These results suggest that having more walkable amenties is associated with narrowing the gap in hospitalizations between the highest and lowest income quintiles. Conversly, an additional kilometer of bike infrastructure nearby is associated with a wider gap, although we do not take this to imply that reducing bike infrastructure will diminish the gap in hospitalizations. This will be further explained once the gentrification models are presented. Adding a transit stop within one kilometer has no strong impact on the hospitalization gap. The walkability coefficients when modelling only severe hospitalization types have larger magnitudes than the models using all hospitalizations, which was expected given the presumed lower elasticity of high severity hospitalizations. The opposite direction of impact for walkability and bikability is unexpected, but will become clearer when examining the results for Equations 17 and 18.

We have previously considered the possibility than an association between transit and the hos-

Table 6: Results for Equations 12 and 13

	All hospitalizations, unweighted		Severe only, weighted	
Model	(12)	(13)	(12)	(13)
Intercept	-4.472	-4.166	1.917	2.232
	(0.098)	(0.119)	(0.531)	(0.462)
ln(Median Income)	0.279	0.275	0.109	0.105
	(0.240)	(0.241)	(0.686)	(0.693)
Walk index	0.838	0.759	0.957	0.868
	(0.003***)	(0.006***)	(0.003***)	(0.006***)
Transit stop within 1km	-0.004		-0.004	
	(0.100)		(0.128)	
Bike infrastructure		-0.052		-0.054
		(0.008***)		(0.017**)
Number of observations	434	434	333	333
Adjusted R-squared	0.111	0.120	0.102	0.112

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

pitalization gap might actually indicate a relationship between population size and this gap. However, the correlation matrix given by Table 5 shows no significant correlation between the transit indicators and population size. The results from Table 7 further test this hypothesis.

Table 7 does not indicate that the positive association observed between walkability and the hospitalization gap is explained by population size, which supports the hypothesis that transit is a factor in explaining the gap in hospitalizations between the lowest and highest income quintiles. However, the R^2 values for Table 6 are only marginally higher than those for Table 7.

Given the lack of any interesting impact of transit availability, it is possible that the improvements to the hospitalization gap associated with walkability comes from increased levels of physical activity rather than due to increased access to health inputs (i.e. Point 1 from Section 3 is more important for explaining the relationship than Point 2). The results in Table 8, which show the impact of active living environment class on the hospitalization gap, will inform what is behind this relationship.

Table 8 shows the impact that active living opportunities has on the hospitalization gap. Given

Table 7: Results for Equation 15

Model	All hospitalizations, unweighted	Severe only, weighted
Intercept	0.693	2.057
	(0.824)	(0.563)
ln(Median Income)	0.338	0.206
	(0.181)	(0.476)
ln(CSD size)	-0.056	-0.051
	(0.087*)	(0.184)
Number of observations	434	333
Adjusted R-squared	0.094	0.077

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the weak results, it cannot be concluded that the association between neighbourhood walkability and the improved gap is due to increased physical activity. This relationship must be due to other factors; we suggest that low-cost access to other health inputs such as healthy food stores and pharmacies may be a possible explanation. However, the bikability results shown in Table 6 tell a different story. It is therefore necessary to look at the results for Equations 17 and 18 to test the theory that transit infrastructure prompts gentrification, and therefore moderates or mitigates any improvements to the hospitalization gap.

As a probit model is used to show the relationship between neighbourhood characteristics and the gentrification indicator, only the marginal effects are shown in Table 9. The magnitude of coefficients in a probit regression do not have the same meaning as they would with a linear regression. In the probit case, only the direction of impact and p-value of the coefficients are meaningful.

Table 9 shows that more bike and transit infrastructure is associated with gentrification. Walkability generally tends to be linked to neighbourhoods that are not considered to be gentrified, although interestingly have no amenities within walking distance is positively and statistically significantly associated with a postal code having been gentrified. Given that increased access to health inputs may improve the hospitalization gap as previously discussed, Table 9 may indicate that bikability and transit would also have this effect, but that any positive impact is mitigated by gentrification, creating a difficult cost of living for those in the lowest income quintile. Bikeability and transit may have a different direction of impact from walkability as they typically connect a

Table 8: Results for Equation 16

	All hospitalizations, unweighted	Severe only, weighted
Intercept	-6.765	-0.411
	(0.042)	(0.914)
ln(Median Income)	0.523	0.362
	(0.063*)	(0.258)
ALE class	0.001	-0.001
	(0.987)	(0.988)
Number of observations	434	322
Adjusted R-squared	0.088	0.072

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

neighbourhood to the city's downtown core, increasing access to the labour market. Badland and Pearce (2019) make a similar claim.

The regressions for Equations 12 and 13 were run again while dropping observations from one province at a time, with little change in the results. This suggests that the patterns observed here are consistent across the country and no single region drives the overall results. The same regressions are also run after splitting the CSDs by above- and below-median median incomes. This yielded more interesting results, shown in Table 10.

Table 10 shows that coefficients for all transit types are positive in lower income CSDs, while the strength of the relationship between transit infrastructure and the hospitalization gap is weaker for high income CSDs. This suggests that different census subdivisions cannot be assumed to react in the same way to transit-related interventions.

Table 6 indicates a relationship between sustainable transit infrastructure and the hospitalization gap, meaning that we reject the null hypothesis that $\tilde{I}(t)$ is a constant function. This means that transit infrastructure appears to be more impactful for the hospitalization rate of low income individuals than high income individuals. The mechanism behind this relationship does not appear to be increased physical activity, as shown in Table 8, but may be better explained by low-cost access to health inputs such as grocery stores and pharmacies. However, Table 9 indicates that these benefits are mitigated by bike and transit infrastructure also being associated with gentrification.

Table 9: Marginal Effects for Equations 17 and 18

Model	(12)	(13)
Transit stop within 1km	0.002 (0.000***)	
Bike infrastructure		0.010 (0.000***)
Walk: 0	0.045 (0.000***)	0.037 (0.001***)
Walk: 1	-0.027 (0.000***)	-0.026 (0.000***)
Walk: 2	-0.045 (0.000***)	-0.044 (0.000***)
Walk: 3	-0.067 (0.000***)	-0.065 (0.000***)
Walk: 4	-0.067 (0.000***)	-0.073 (0.000***)
Walk: 5	0.009 (0.588)	0.008 (0.630)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 10: Results for Equations 12 and 13 using weighted severe hospitalizations only, and splitting by median income

	CSD median income $\leq$ \$77,282		CSD median income $>$ \$77,282	
Model	(12)	(13)	(12)	(13)
Intercept	-2.976	-1.203	1.107	1.014
	(0.509)	(0.789)	(0.909)	(0.916)
ln(Median Income)	0.582	0.431	0.203	0.252
	(0.152)	(0.285)	(0.811)	(0.764)
Walk index	0.852	0.887	0.740	0.659
	(0.005***)	(0.003***)	(0.266)	(0.322)
Transit stop within 1km	0.004		-0.017	
	(0.184)		(0.010**)	
Bike infrastructure		0.000		-0.122
		(0.998)		(0.007***)
Number of observations	173	173	160	160
Adjusted R-squared	0.137	0.128	0.114	0.118

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

7 Discussion

In Section 1, it was suggested that sustainable transit infrastructure could impact health inequality through several avenues, including physical activity, access to health inputs, exposure to pollution and gentrification. This hypothesis was tested using the gap in hospitalizations between low- and high-income residents as a proxy for health inequality.

Access to health inputs appear to be the strongest explaination for the association between walkability and the improved hospitalization gap shown in Table 6. Living in an area rated as an active living environment is not strongly related to an improved gap, while the walkability index was. Although both may contribute to lowering health inequality, the results suggest that it is the access to amenities that is the most important mechanism through which this manifests itself.

Table 6 does not show that transit infrastructure uniformly reduces the hospitalization gap. In Section 3, it was hypothesized that there were two avenues by which transit infrastructure could increase the expenditure required to achieve a given level of health. The data that would be required to test whether pollution exposure or road traffic injuries are a factor is not available, but these have previously been shown to have minimal impacts (Abu-Qarn and Lichtman-Sadot 2022; Freeman et al. 2013). The gentrification hypothesis does appear to hold some weight, however. This will mostly occur when transit improvements are not evenly implemented across a city (Tehrani, Wu, and Roberts 2019). Walkability may have a different relationship with gentrification than bikeability or transit as the latter two likely connect a neighbourhood with the rest of the city, including the downtown core, allowing access to jobs and other amenties (Badland and Pearce 2019).

The work presented here would be a valuable foundation for further research. In the model, it is assumed that individuals are unable to migrate between cities, although in the long term this would likely happen and it would be useful to incorporate this possibility into the model. We also do not consider how transit infrastructure is financed, and simply include it as an endownment. In reality, such infrastructure would depend on taxation, thereby influencing the budgets of consumers. Higher levels of transit infrastructure may lower health expenditure requirements, but also decrease consumers' budgets by increasing their tax burdens. That considered, car infrastructure is also paid for by public funds and to some extend this funding may be reallocated. This question requires further examination to be properly addressed. Finally, we ignore how transit infrastructure may impact the cost of other consumption z_i. Taking this into consideration could provide valuable insights into the question we are trying to answer.

Another weakness is that the CIHI hospitalization data is not ideally suited to what we measure. It is not possible to observe the incomes of hospitalized individuals, only the CSD income quintile to which they belong. This issue is mitigated as much as possible by including a term in the regression for CSD median income. Using hospitalization data rather than measuring objective indicators such as blood pressue of everyone in a sample also means that we are not perfectly able to observe health levels, as discussed in Sections 3 and 5. This is mitigated as much as possible by limiting the analysis to severe hospitalizations, which is expected to reduce the distortions in health measurement due to certain individuals having easier access to hospitals or preferring to use more healthcare services. Regardless, hospitalization rates cannot be a perfect proxy for health status. An advantage of the CIHI data, however, is that it standardizes hospitalization rates by age makeup, which allows us to better compare CSDs.

Of course, health inequality is determined by a wide variety of factors, observable and otherwise, of which transit infrastructure is only one. As a result, the explanatory power of the model is relatively low. This paper does not address race, which is an important determinant of health inequality within cities (De Maio et al. 2020). We do not have data on all types of hospitalizations, some of which might be more prevalent amongst certain demographics, which creates further bias. Finally, while the relationship between health and income is generally explained as more income allowing individuals to choose better health (Fritzell, Nermo, and Lundberg 2004; Ecob and Smith 1999), we cannot dismiss the possibility that intrinsically healthier people are able pursue higher paying jobs. If this is the case, it changes the interpretation of the results presented.

The issues of race and demographics noted above could be better controlled for by a difference-in-difference model examining health inequality in each city before and after a change in transit availability. The cross-sectional approach used was necessary due to the nature of the data available, but is limited in that we cannot control for many other factors that vary across cities. An attempt was made to limit this issue by including a dummy variable for province in the model, but this does not completely resolve it.

This investigation has rejected the null hypothesis that $\tilde{I}(t)$ is a constant function. Improvements to the hospitalization gap appear to be due more to increased access to health inputs than higher levels of physical activity. The positive benefits of increasing sustainable transit infrastructure appear to be negated to some extent by this infrastructure also causing gentrification in some neighbourhoods. A possible way to mitigate this effect would be to implement transit infrastructure evenly in a city.

8 Conclusion

This paper has examined whether sustainable transportation availability impacts health inequality in Canadian cites. By using linear regression analysis, we have shown that walkability is associated with reduced hospitalization gap in a census subdvision, which is not attributable to increased physical activity. Bikeability is related to larger gaps, while transit does not appear to impact the gap. However, both bike and transit infrastruture are associated with gentrification. This suggests that sustainable transit infrastructure must be implemented evenly across a city to avoid adverse effects. In future research, it would be interesting to use a less sensitive measure of health, further investigate why walkability and bikeability have opposite direction of impact on health inequality, and to relax some of the assumptions made.

References

Abu-Qarn, Aamer, and Shirlee Lichtman-Sadot. 2022. “Can greater access to secondary health care decrease health inequality? Evidence from bus line introduction to Arab towns in Israel.” *Economic Modelling* 106:105695.

Asada, Yukiko. 2005. “A framework for measuring health inequity.” *Journal of Epidemiology & Community Health* 59 (8): 700–705.

Badland, Hannah, and Jamie Pearce. 2019. “Liveable for whom? Prospects of urban liveability to address health inequities.” *Social science & medicine* 232:94–105.

Beairsto, Jeneva, Colin Ferster, Karen Laberee, Trisalyn Nelson, Kevin Manaugh, and Meghan Winters. 2022. *Canadian Bikeway Comfort and Safety Classification System (Can-BICS) Metrics User Manual.*

Bernard, Paul, Rana Charafeddine, Katherine L Frohlich, Mark Daniel, Yan Kestens, and Louise Potvin. 2007. “Health inequalities and place: a theoretical conception of neighbourhood.” *Social science & medicine* 65 (9): 1839–1852.

Bernstien, Jaela. 2021. “Canadians are among the world’s worst carbon emitters. Here’s what we can do about it.” *CBC News* (October 8, 2021). Accessed June 15, 2023. https://www.cbc.ca/news/science/how-canadians-can-cut-carbon-footprints-1.6202194.

Briant, Anthony, P-P Combes, and Miren Lafourcade. 2010. “Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?” *Journal of Urban Economics* 67 (3): 287–302.

Brown, Scott C, Hilda Pantin, Joanna Lombard, Matthew Toro, Shi Huang, Elizabeth Plater-Zyberk, Tatiana Perrino, Gianna Perez-Gomez, Lloyd Barrera-Allen, and José Szapocznik. 2013. “Walk Score®: associations with purposive walking in recent Cuban immigrants.” *American journal of preventive medicine* 45 (2): 202–206.

Canadian Institute for Health Information. 2023. private communication by e-mail. Corresponded with Erica Withey.

———. 2022. *Measuring Health Inequalities: A Toolkit.* Ottawa, ON.

———. 2019a. *Measuring Trends in Health Inequalities in Cities: Hospitalization and Day Surgery Indicator Results, by Census Metropolitan Area and Census Subdivision - Data Tables.* Ottawa, ON.

Canadian Institute for Health Information. 2019b. *Measuring Trends in Health Inequalities in Cities: Hospitalization and Day Surgery Indicators - Technical Notes.* Ottawa, ON.

Cullen, Mark R, Clint Cummins, and Victor R Fuchs. 2012. "Geographic and racial variation in premature mortality in the US: analyzing the disparities." *PLoS One* 7 (4): e32930.

De Maio, Fernando, Raj C Shah, John Mazzeo, and David A Ansell. 2020. *Community health equity: a Chicago reader.* University of Chicago Press.

Deryugina, Tatyana, and David Molitor. 2021. "The causal effects of place on health and longevity." *Journal of Economic Perspectives* 35 (4): 147–170.

DMTI Spatial Inc. various years. *CanMap Postal Code Suite [compute file].*

———. 2019. *CanMap Postal Code Suite v2019.3 [compute file] DMTI Spatial Inc.*

Duncan, Dustin T. 2013. "What's your Walk Score®?: Web-based neighborhood walkability assessment for health promotion and disease prevention." *American journal of preventive medicine* 45 (2): 244–245.

Duru, Gérard, and Jean H Paul Paelinck. 1990. *Econometrics of health care.* Vol. 20. Springer Science & Business Media.

Ecob, Russell, and George Davey Smith. 1999. "Income and health: what is the nature of the relationship?" *Social science & medicine* 48 (5): 693–705.

Firth, C.L., B. Thierry, D. Fuller, M. Winters, and Y. Kestens. 2020. *GENUINE: Gentrification, Urban Interventions, and Equity Technical Document.*

Freeman, Lance, Kathryn Neckerman, Ofira Schwartz-Soicher, James Quinn, Catherine Richards, Michael DM Bader, Gina Lovasi, Darby Jack, Christopher Weiss, Kevin Konty, et al. 2013. "Neighborhood walkability and active travel (walking and cycling) in New York City." *Journal of Urban Health* 90:575–585.

Friedman, Gabriel. 2023. "Why are cars so expensive in Canada today?" *Financial Post* (May 10, 2023). Accessed June 12, 2023. https://financialpost.com/transportation/autos/why-cars-so-expensive-canada-today.

Fritzell, Johan, Magnus Nermo, and Olle Lundberg. 2004. "The impact of income: assessing the relationship between income and health in Sweden." *Scandinavian journal of public health* 32 (1): 6–16.

Fuller, Daniel, Lise Gauvin, and Yan Kestens. 2013. "Individual-and area-level disparities in access to the road network, subway system and a public bicycle share program on the Island of Montreal, Canada." *Annals of behavioral medicine* 45 (suppl_1): S95–S100.

Glaeser, Edward L, Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of economic geography* 1 (1): 27–50.

Glied, Sherry, and Peter C Smith. 2013. *The Oxford handbook of health economics.* Oxford University Press.

Gorman, Dermot, MJ Douglas, L Conway, P Noble, and P Hanlon. 2003. "Transport policy and health inequalities: a health impact assessment of Edinburgh's transport policy." *Public health* 117 (1): 15–24.

Government of Canada. 2022. *2030 Emissions Reduction Plan: Clean Air, Strong Economy.* Accessed: 2023-06-26. https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/climate-plan-overview/emissions-reduction-2030.html.

Herrmann, Thomas, Geneviève Boisjoly, Nancy A Ross, and Ahmed M El-Geneidy. 2017. "The missing middle: Filling the gap between walkability and observed walking behavior." *Transportation research record* 2661 (1): 103–110.

Hirsch, Jana A, Kari A Moore, Kelly R Evenson, Daniel A Rodriguez, and Ana V Diez Roux. 2013. "Walk Score® and Transit Score® and walking in the multi-ethnic study of atherosclerosis." *American journal of preventive medicine* 45 (2): 158–166.

Infrastructure Canada. 2023. *Building Canada's public transit future: Healthy and sustainable modes of transportation for all.* Accessed: 2023-06-26. https://www.infrastructure.gc.ca/transit-transport/index-eng.html.

Maizlish, Neil, James Woodcock, Sean Co, Bart Ostro, Amir Fanai, and David Fairley. 2013. "Health cobenefits and transportation-related reductions in greenhouse gas emissions in the San Francisco Bay area." *American journal of public health* 103 (4): 703–709.

"Matching Census Data to Postal Codes using SPSS." n.d. *University of Toronto Libraries,* https://mdl.library.utoronto.ca/technology/tutorials/matching-census-data-postal-codes-using-spss.

Miller, Adam, and Prapti Bamaniya. 2023. "Why ERs are under intense pressure across Canada — and how to help fix them." *CBC News* (June 24, 2023). Accessed June 26, 2023. https://www.cbc.ca/news/health/canada-er-pressure-health-care-system-solutions-1.6885257.

Milligan, Kevin. 2021. "Average Tax Rates in the Canadian Personal Income Tax." *National Tax Journal* 74 (2): 513–527.

Mueller, Natalie, David Rojas-Rueda, Haneen Khreis, Marta Cirach, David Andrés, Joan Ballester, Xavier Bartoll, Carolyn Daher, Anna Deluca, Cynthia Echave, et al. 2020. "Changing the urban design of cities for health: The superblock model." *Environment international* 134:105132.

Mukaka, Mavuto M. 2012. "A guide to appropriate use of correlation coefficient in medical research." *Malawi medical journal* 24 (3): 69–71.

OpenStreetMap contributors. 2019. *Planet OSM.* Retrieved May 2021. https://planet.openstreetmap.org/.

Prince, Stephanie A, Samantha Lancione, Justin J Lang, Nana Amankwah, Margaret de Groh, Alejandra Jaramillo Garcia, Katherine Merucci, and Robert Geneau. 2022. "Examining the state, quality and strength of the evidence in the research on built environments and physical activity among adults: An overview of reviews from high income countries." *Health & Place* 77:102874.

Québec, Légis. 2020. *Act respecting health services and social services. S-4.2, 118.1.*

Rehkopf, David H, Lisa F Berkman, Brent Coull, and Nancy Krieger. 2008. "The non-linear risk of mortality by income level in a healthy population: US National Health and Nutrition Examination Survey mortality follow-up cohort, 1988–2001." *BMC Public Health* 8:1–11.

Rehnberg, Johan, and Johan Fritzell. 2016. "The shape of the association between income and mortality in old age: A longitudinal Swedish national register study." *SSM-Population Health* 2:750–756.

Remiorz, Ryan. 2020. "Montreal reassigns streets for pedestrians, bicycles during pandemic." *The Globe and Mail* (May 15, 2020). Accessed May 15, 2020. https://www.theglobeandmail.com/canada/article-montreal-converting-streets-temporarily-to-help-confined-residents-get-2/.

Ross, N., R. Wasfi, T. Herrmann, and W. Gleckner. 2018. *Canadian Active Living Environments Database (Can-ALE) User Manual.*

Salomon, Joshua A, Theo Vos, Daniel R Hogan, Michael Gagnon, Mohsen Naghavi, Ali Mokdad, Nazma Begum, Razibuzzaman Shah, Muhammad Karyana, Soewarta Kosen, et al. 2012. "Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010." *The Lancet* 380 (9859): 2129–2143.

Shareck, Martine, Katherine L Frohlich, and Yan Kestens. 2014. "Considering daily mobility for a more comprehensive understanding of contextual effects on social inequalities in health: A conceptual proposal." *Health & place* 29:154–160.

Statistics Canada, Canadian Mortgage and Housing Corporation. 2020. *Proximity Measures Database – Early release, April 17, 2020.*

Statistics Canada. 2023a. *Canadian Income Survey, 2021.* Accessed: 2023-06-12. https://www150.statcan.gc.ca/n1/daily-quotidien/230502/dq230502a-eng.htm.

———. 2016a. *Census Profile, 2016 Census.* Accessed: 2023-07-04. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

———. 2016b. *Data tables, 2016 Census.* Accessed: 2023-06-16. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=2&LANG=E.

———. 2021. *Household spending, Canada, regions and provinces.* Accessed: 2021-01-22. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110022201.

———. 2023b. *Population estimates, July 1, by census subdivision, 2016 boundaries.* Accessed: 2023-07-29. https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1710014201.

———. 2017. *Postal CodeOM Conversion File (PCCF).*

———. 2020. *Proximity Measures Database – Early release.* Accessed: 2023-04-27. https://www150.statcan.gc.ca/n1/pub/17-26-0002/172600022020001-eng.htm.

Tasker, John Paul. 2022. "Canada releases plan for a 40 per cent cut in carbon emissions by 2030." *CBC News* (March 29, 2022). Accessed March 29, 2022. https://www.cbc.ca/news/politics/canada-2030-emissions-reduction-plan-1.6401228.

Tehrani, Shadi O, Shuling J Wu, and Jennifer D Roberts. 2019. "The color of health: residential segregation, light rail transit developments, and gentrification in the United States." *International Journal of Environmental Research and Public Health* 16 (19): 3683.

The Centre for Active Transportation. 2020. *Active Transportation Actions in Response to Covid-19.* Accessed: 2023-06-26. https://www.tcat.ca/resources/active-transportation-actions-in-response-to-covid-19/.

Turrell, Gavin, Michele Haynes, Lee-Ann Wilson, and Billie Giles-Corti. 2013. "Can the built environment reduce health inequalities?" *Health & place* 19:89–98.

Wilkinson, Richard G. 1992. "Income distribution and life expectancy." *BMJ: British Medical Journal* 304 (6820): 165.

Appendix

Table 11: List of CSDs

Province	Census subdivisions
Alberta	Airdrie, Calgary, Cochrane, Edmonton, Fort Saskatchewan, Leduc, Leduc County, Morinville, Parkland County, Rocky View County, Spruce Grove, St. Albert, Stony Plain, Strathcona County
British Columbia	Burnaby, Coquitla, Delta, Langford, Langley, Maple Ridge, New Westminster, North Vancouver, Oak Bay, Pitt Meadows, Port Coquitlam, Port Moody, Richmond, Saanich, Sidney, Surrey, Vancouver, Victoria, West Vancouver, White Rock
Manitoba	Victoria, Winnipeg
New Brunswick	Dieppe, Fredericton, Moncton, Quispamsis, Richmond, Riverview, Saint John
Newfoundland	Burlington, Conception Bay South, Mount Pearl, St. John’s
Nova Scotia	Halifax
Ontario	Ajax, Aurora, Bradford West Gwillimbury, Brampton, Burlington, Cochrane, Grimsby, Halton Hills, Hamilton, London, Markham, Milton, Mississauga, New Tecumseth, Newmarket, Oakville, Ottawa, Pickering, Richmond Hill, Russell, St. Thomas, Strathroy–Caradoc, Toronto, Vaughan
Prince Edward Island	Sherbrooke, Victoria
Quebec	Beloeil, Blainville, Brossard, Côte-Saint-Luc, Châteauguay, Chambly, Dollard-des-Ormeaux, Dorval, Gatineau, La Prairie, L’Ancienne-Lorette, Lévis, Laval, Longueuil, Magog, Mascouche, Mirabel, Montréal, Ottawa (Gatineau), Québec, Repentigny, Richmond, Sainte-Thérèse, Saint-Eustache, Saint-Jérôme, Saint-Lambert, Sherbrooke, Terrebonne, Varennes, Vaudreuil-Dorion
Saskatchewan	Regina, Saskatoon