ELSEVIER

Evaluating methods for measuring daily walking to public transport: Balancing accuracy and data availability

Emily Grisé[a],*, Rania Wasfi[b], Nancy A. Ross[c], Ahmed El-Geneidy[d]

[a] School of Urban and Regional Planning, University of Alberta, 1-26 Earth Sciences Bldg, University of Alberta, T6G 2E3, Edmonton, Alberta, Canada
[b] Centre de Recherche Du CHUM, Université de Montréal, Department of Geography, McGill University, Canada
[c] Department of Geography, McGill University, Canada
[d] School of Urban Planning, McGill University, Canada

A B S T R A C T

Background: Quantifying physical activity accumulated through daily commuting is challenging due to the scarcity of detailed data, especially for public transport trips. Using Montreal, Canada as a case study, this paper measures and compares an individual's daily amount of walking to and from public transport in their regular commute to work using two datasets and two methods.
Methods: The first method used urban level detailed origin-destination microdata. Distances walked to and from public transport stations were measured using trip details provided from the survey. The second method used open data including commuting flows obtained from census data. Public transport trips for each flow were modeled using a fastest route algorithm applied to General Transit Feed Specification (GTFS) data obtained from public transport operators. Walking distances were then extracted from these trip paths. Multilevel mixed-effect regression modeling was used to identify the determinants of total walking when using both methods. A sensitivity analysis was then used to derive an adjustment table for those who wish to use open data to estimate walking.
Results: For commuter train users 471 m must be added to walking estimates obtained from the commuting flows data, while negative adjustments are required for subway users (122 m), city bus users (366 m), suburban bus users (516 m), and peripheral bus users (1186 m).
Conclusions: The methodology presented in this study provides researchers and professionals in other cities without access to detailed origin-destination survey data with a guide to use open data to accurately estimate total walking distances accumulated in a daily commute by public transport.

1. Introduction

Being active for at least 150 min a week can help reduce the risk of heart disease, type 2 diabetes, high blood pressure, and obesity (Canadian Society for Exercise Physiology, 2018). Only two in ten Canadian adults and one in ten children met the Canadian physical activity guidelines in 2013 (Public Health Agency of Canada, 2016). While the most commonly cited barrier to exercise is lack of time (Booth et al., 1997; Borodulin et al., 2016; Godin et al., 1994; Owen and Bauman, 1992; Yoshida et al., 1988), the good news is that individuals can achieve the recommended physical activity guidelines through small sessions of at least 10 min of activity throughout

the day (World Health Organization, 2017). Walking accumulated through daily travel is one way that individuals can meet physical activity targets. Compared with public health interventions that aim to promote active lifestyles and change the behaviour of individuals, policies that aim to increase the proportion of trips that are made on active modes of travel such as public transport, bicycling and walking, are likely to increase the activity levels of a given population in a way that can be sustained in the long term (Siegel et al., 1995).

While walking is regarded as one of the most accessible forms of physical activity for all ages, it is limited in its range as a mode of transport. The ability to walk to work is relatively uncommon and only for individuals who can afford to live in close proximity to their place of work. Increasingly, scholarly attention is being paid to walking to public transport, as almost all trips by this mode require some physical activity. Researchers have looked at how much physical activity is accumulated when commuting by public transport, and what characteristics of the built environment are associated with higher levels of walking (Besser and Dannenberg, 2005; Brown and Werner, 2007; MacDonald et al., 2010; Morency et al., 2011; Wasfi et al., 2013).

The goal of our study is to simulate a scenario where an urban planning or public health professional only has access to aggregate, publicly available commuting flow census data at the census tract (CT) level (Statistics Canada, 2011), which provides details on the number of residents that live and travel between different CTs. We propose that such census data can be used to measure walking to public transport when combined with General Transit Feed Specification (GTFS) data obtained from public transport operators. We undertake a multilevel regression analysis to explore how the open data method differed in the estimation of walking levels compared with a method that uses travel survey microdata (i.e., origin-destination survey). We then use a series of sensitivity analyses to compare both methods of calculating walking distances and generate an adjustment table that can be used by professionals who are estimating walking levels from widely available census commuting flows and open GTFS data. This study provides a methodological contribution to the literature by presenting a tool for urban planning or public health professionals, researchers, and policy-makers from municipalities of all sizes to measure the relationship between physical activity and public transport use in areas where access to detailed travel survey microdata is limited.

2. Literature review

In response to rising obesity levels observed in many Westernized countries, leading medical and public health researchers have advocated for the increased uptake of walking and bicycling as a means of achieving daily moderate-intensity physical activity goals for improved health (Dora, 1999; Koplan and Dietz, 1999). The specific level of physical activity achieved through daily public transport use is mediated by many factors. The most significant factor determining physical activity levels of public transport users is the mode being accessed. Distances walked to buses are generally shorter compared with commuter trains and subways (Alshalalfah and Shalaby, 2007; Wasfi et al., 2013), and thus catchment areas have traditionally been defined as 400 m for bus and 800 m for rail (for example Fielding et al., 1978; Murray and Wu, 2003; Zhao et al., 2003). However, several authors have observed walking distances to public transport beyond the 400 and 800 m threshold, thus presenting an opportunity for operators to increase stop spacing (Daniels and Mulley, 2013; El-Geneidy et al., 2014; van Soest et al., 2019; Weinstein et al., 2008). Variations have also been observed between downtown and suburban locations with suburban commuters tending to walk further to more comfortable and faster modes (El-Geneidy et al., 2014; O'Sullivan and Morrall, 1996). With respect to built environment characteristics, van Soest et al. (2019) observed through a systematic review of literature on the topic of walking to public transport that in higher density environments public transport services are often densely located thereby reducing distances people need to walk to access public transport services. Alternatively, density can be associated with high levels of local accessibility, which can be attractive to walkers, thereby either increasing the share of walking to access/egress transport or distance people will walk to public transport (Jiang et al., 2012), or by expanding the catchment area around stops or stations.

Using detailed origin-destination survey data in Montreal, Canada, Wasfi et al. (2013) modeled the underlying factors that contribute to achieving the recommended 30 min of daily physical activity while commuting by public transport. The authors concluded that approximately 11% of commuters achieved their recommended minutes of daily physical activity, and suburban train users walked the greatest number of minutes per day. Similar findings were observed by Besser and Dannenberg (2005), who used the National Household Travel Survey to assess walking times to public transport across the United States of America. The authors observed that 29% of public transport users achieved more than 30 min of daily physical activity from walking to and from stops, and that rail users were 1.67 times more likely than bus users to achieve their daily physical activity target. Using data from the English National Travel Survey, Patterson et al. (2019) observed that one in three public transport users achieved 30 min of daily physical activity: 21% bus users achieved this daily physical activity goal and 52% of train users met their physical activity target through taking public transport. Although these studies find that not all public transport riders regularly meet their daily physical activity guidelines solely through walking to public transport, it is important to note that individuals walking to public transport are reaching some level of physical activity that they would not have if commuting by a sedentary form of transport, such as driving.

More recently, researchers have been interested in understanding the exertion level of public transport related walking to ascertain whether these walkers are in fact meeting the required moderate-to-vigorous physical activity (MVPA) intensity level set within physical activity guidelines (World Health Organization, 2010). Using accelerometer data collected from participants residing within a half mile of a new light-rail stop in Salt Lake City, USA, Brown and Werner (2007) measured the length and intensity of walking trips to and from light rail stops, concluding that these walking trips are of moderate-intensity, and collectively can contribute to attaining the WHO-recommended physical activity level. MacDonald et al. (2010) studied the impact of a new light rail transit (LRT) line in Charlotte, USA on obesity and physical activity level. These authors found that commuting by LRT was associated with a reduction in BMI and a reduced odds of becoming obese over time, compared with a control group of individuals living in

similar neighbourhoods but did not commute by LRT. At an aggregate level, She et al. (2019) observed through a longitudinal analysis of county-level health and transport data that in the United State a 1% increase in the share of frequent public transport users in a county population is estimated to decrease the county population obesity rate by 0.47%. Similarly, She et al. (2017) found a 0.22% decrease in county population obesity prevalence through a cross sectional analysis of the United States 2009 Behavioral Risk Factor Surveillance System and the 2009 National Household Travel Survey.

Most studies to date have measured walking to public transport using either travel survey data (Besser and Dannenberg, 2005; Morency et al., 2011; Wasfi et al., 2013), or readings from biosensors (Brown and Werner, 2007). However, professionals often lack either the access to such detailed data or the resources to recruit participants and monitor physical activity levels at such a detailed level. Biosensor-assessed step counts are also burdensome for respondents and can induce non-routine behaviour. Veillette et al. (2018) recently calculated levels of walking to public transport using detailed and non-detailed travel survey data. While this study provided a better understanding of how modeled walking routes differ from reported routes, it did not present a methodology for professionals without access to travel survey data to measure walking to public transport. In this study we examine how the accuracy of the estimation of walking distances accumulated while taking public transport using publicly available commuting flow census data and public transport schedule data (GTFS data).

3. Data

3.1. Origin-destination microdata

A common method for cities to evaluate the travel behaviour of residents is through the collection of origin-destination (OD) surveys. OD microdata are used by cities in many ways including modeling travel demand and evaluating the modal split. The Montreal, Canada, OD survey has been typically conducted every five years by the Agence Métropolitaine de Transport (AMT) (Agence Métropolitaine de Transport, 2013). In our analysis we used the 2013 OD microdata. Five percent of households are sampled and asked to recall all trips made in the past day by each household member. This data includes coordinates of the location of the home and each location visited in that day, mode taken for each trip, and personal and household characteristics. For trips taken by public transport, each respondent was asked to provide route details including bus route(s) used, subway or train station of access and egress, and subway used. After selecting all home-based trips for the purpose of work or school that were taken by adults (over the age of 18) that began and ended in Greater Montreal, our sample consisted of 9588 trips that were made using bus, subway or commuter rail. We omitted trips with missing route information, as well as trips where individuals drove to a public transport station.

3.2. Commuting flows

Statistics Canada collects information on the commuting modes and patterns of all Canadians as part of the Census of Canada. Commuting modes and patterns are available at the level of a census tract (CT) – a small urban area that is generally socially homogeneous and made up of between 2500 to 8000 people (Fig. 1). Similar data exist in other countries around the world, making this approach appealing for international comparative research. We selected all commuting flow pairs within the Greater Montreal Region where there were at least 10 individuals who commute by public transport between CT pairs (smaller counts were suppressed for confidentiality). Our sample consisted of 2755 commuting flows. Such data can be present at finer details in some countries with large populations – the U.S., for example, provides this information at the block level, yet in the Canadian context such detail will be subjected to heavy data suppression so we opted to use the larger CT as the basic unit of analysis.

4. Methodology

4.1. Origin-destination microdata scenario

We modeled walking distances to and from public transport for each work-related trip in ArcGIS using two shapefiles. The first shapefile was a pedestrian street network where highways were removed from the road network. The second contained all stops serving the subway, commuter train and bus services for Montreal (see Fig. 2 for the transit system in the Montreal Metropolitan Region). The objective here was to calculate the walking distance from the individuals' home to the first public transport stop and the walking distance from the last public transport stop to the final destination. As we only knew the route(s) used in each trip, such as the bus route or subway line, we used the closest facility tool to generate lines from each home and work location to their respective nearest 200 public transport stops. We then selected only the lines connecting each home or work location to a nearby stop that matched the route that the respondent reported using. The two distances were summed to determine the total walking distance for each public transport trip.

Using the stop where each individual began their trip, we measured the headway (time between subsequent vehicles) of the first public transport route during the morning peak commuting time (7 and 9 a.m.). We also measured the distance traveled in-vehicle, to provide a complete picture of the individual's daily public transport commute. This was done using the previously identified first and last stop used in their trip, and by determining the length of the public transport trip between these points using a network built from GTFS data from all public transport operators in the Greater Montreal region. GTFS data produced in 2013 was used to represent the level of service that was present at the time of data collection. The number of transfers that occurred during this trip was calculated as the total number of distinct public transport routes used during this trip minus one. Note, we defined a transfer as a route or mode

Projection: NAD 1983 MTM 8
Data sources: Statistics Canada

Census tract boundaries

City of Montreal
Suburban regions
Peripheral regions

0   4   8   16
Km
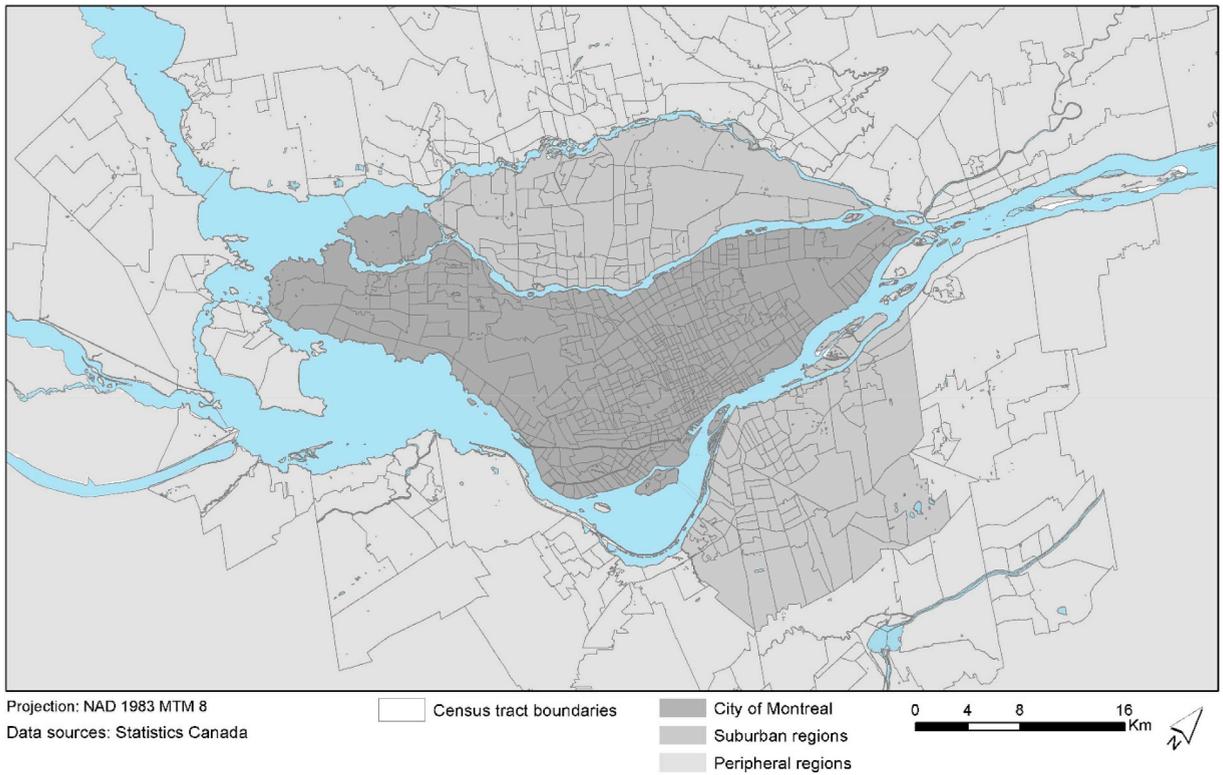
Fig. 1. Census tract boundaries within Montreal Metropolitan Region.

Projection: NAD 1983 MTM 8
Data sources: Montreal transport operators
& Statistics Canada

Rail network
Subway network
Bus network

City of Montreal
Suburban regions
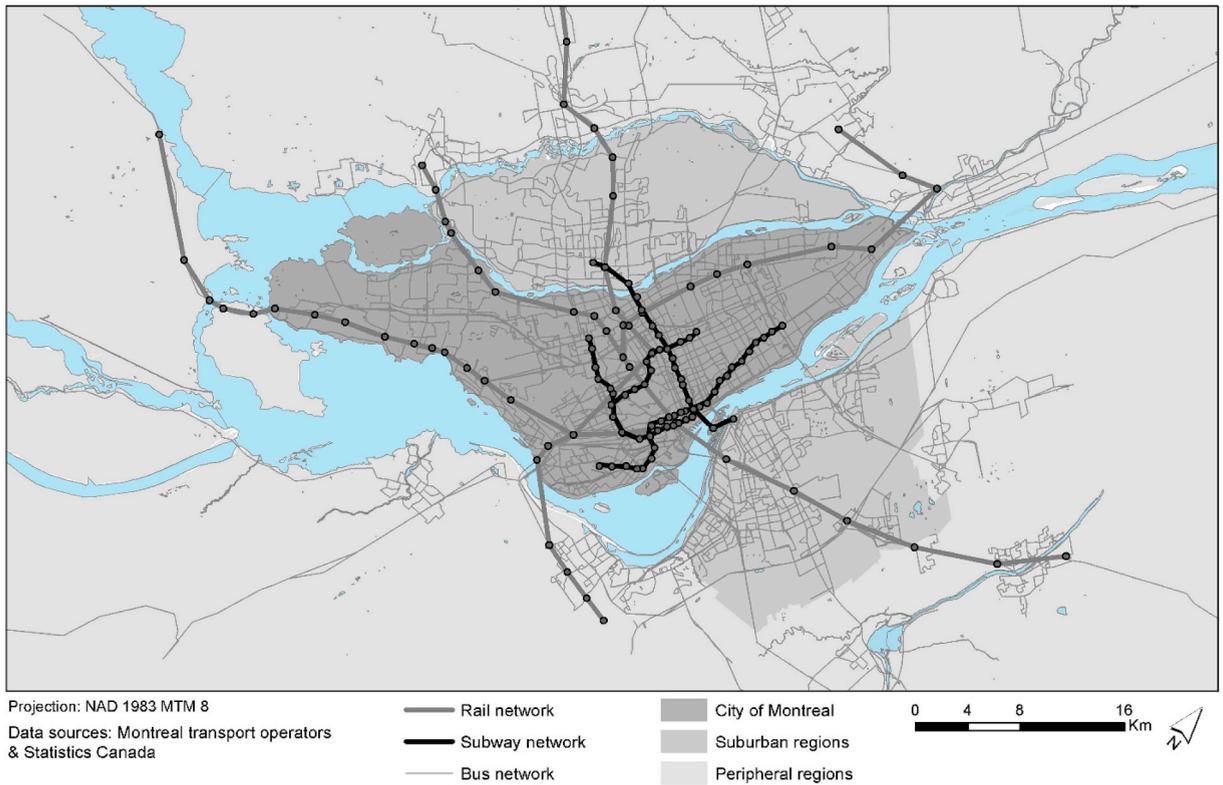Peripheral regions

0   4   8   16
Km

Fig. 2. Public transit network in Montreal Metropolitan Region.

switch that occurred during any trip. For example, a trip that began on a bus and then required two subway lines would be a two-transfer trip.

4.2. Commuting flows scenario

Using the CT centroids, we determined the public transport path of all flows according to the shortest travel time. GTFS data from all public transport operators in the Greater Montreal region were obtained and a public transport network was constructed in ArcGIS using the 'Add GTFS to a network dataset' toolbox available in ArcMap. The fastest travel time between each CT pairs was determined, which includes walking, waiting, transfer, and in-vehicle times.[1] We solved these paths using a departure time of 8 a.m. (the peak of morning travel) on a regular Tuesday in December of 2013. After modeling all trip pairs, a series of GIS computations gave us: walking distances for each trip, first mode used, distance traveled in-vehicle, and number of transfers.

The portion of each trip that required the individual to walk to the first and last public transport stop was calculated by intersecting each modeled trip flow with the street network, then we applied the "unsplit line tool" to merge coinciding walking segments. We selected only the walking segments that represented either walking from origin to first stop or walking from last stop to final destination. After selecting the two lines that were connected to the origin and destination point, we summed the two walking distances to represent total walking during a trip. See Fig. 3 for a schematic illustration of the measurement of walking distance obtained from both methods.

The number of transfers was obtained by subtracting the number of walking segments identified in step two above. These represent walking segments at the origin and destination. To represent transfers occurring between subways and trains we determined the number of subway and rail transfer stations each user passed through and adjusted the number of transfers to account for transferring to a subway or a train line. The first mode taken was determined by isolating the walking segment from each respondent's home to the first stop and identifying the type of public transport route that intersected with the first walking segment. Lastly, we determined whether a trip began on a city, suburban, or a peripheral bus by determining the agency boundary in which each trip originated.

4.3. Statistical analysis

Two multilevel mixed-effect regression models were constructed to uncover the key determinants of daily walking to public transport as well as to explore how the two approaches differed in their estimation of walking levels (under the assumption that the walking distances obtained from OD microdata are the gold standard as the OD has been extensively validated in the past (Chapleau et al., 1997; Trépanier et al., 2008)). The first model evaluated the determinants of walking to public transport according to OD microdata, to provide a baseline understanding of what factors influence walking levels. In this model, we nested each observation in its respective origin CT, to account for characteristics in the built environment unique to each neighbourhood that may influence walking behaviour. To account for unmeasured characteristics of each respondent's local built environment, the multilevel model approach adopted reduces the potential biases that may be imposed on the model due to not controlling for such characteristics.

In our second model, trip details obtained from the OD microdata and the commuting flows scenarios were merged and differentiated using a dummy variable. In this model we nested trips first according to their origin CT and second by their census flow. Nesting each observation in the census flow accounted for similarities that are likely to arise as a result of sharing the same origin and destination CTs, such as walkability, level of service and partially self-selection. For trips obtained from detailed OD microdata, the census flow was obtained by concatenating the origin and destination CT provided in the dataset. The estimates of this model were used to obtain estimates of daily walking distances according to first mode taken for both methods. Differences between estimates of walking distance obtained by the two scenarios were observed and presented as recommended adjustment values, capturing how much the commuting flows method either underestimated or overestimated walking distances relative to the OD microdata approach.

Both models used home to work trip walking distance (distance walked from home to first stop and from last public transport stop to work) as their dependent variable. The independent variables included in the two models vary slightly according to data availability. Trip details, including number of transfers, first mode taken, in-vehicle distance, and neighbourhood characteristics were included in both models (Table 1). Additional variables included in the first model (detailed OD microdata scenario) include personal characteristics such as income level, age and sex. In the second model (detailed OD microdata scenario and commuting flows scenario merged), interaction terms were included in the model to test the effect of first mode taken on total walking distance between the two methods of analysis.

5. Results

5.1. Comparing trip details by scenario

The average walking distance of a home to work commute as measured in the OD microdata scenario was 1007 m, compared with 1035 m estimated by the commuting flows method. On average, the commuting flows method overestimated daily walking distance

---

[1] For more details on the computation of shortest path using the 'Add GTFS to a network dataset' toolbox in ArcGIS please see the following link: http://esri.github.io/public-transit-tools/AddGTFStoaNetworkDataset.html.
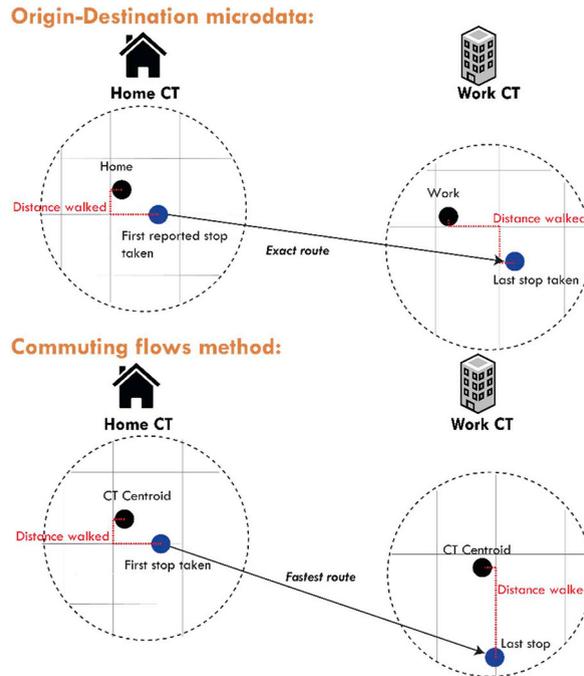
Fig. 3. Schematic diagram illustrating calculation of distance walked using both methods.

by 3%. Important differences in average walking distances between the two scenarios are apparent when comparing trip characteristics (Table 2). Looking at average walking distances by first mode taken, we observe that commuting flow trips that began on a peripheral bus overestimated walking by 44% when compared with detailed OD trips. Estimates of daily walking distances for trips beginning on a subway were most closely estimated by the commuting flows method (underestimated by 142 m). In line with the overall goal of increasing the proportion of the population meeting the recommended daily physical activity goals (30 min a day), we calculated the proportion of individuals meeting the daily physical activity goals through walking to and from public transport stations. Results of the commuting flows scenario show an underestimation of the proportion of train users that meet the recommended daily walking target, while this scenario overestimates the proportion of city bus, suburban and peripheral bus users who meet the recommended physical activity targets.

Some of the differences in the estimates between the two scenarios are likely a result of routing differences. In the OD microdata, 6% of the sample reported the commuter train as their first mode, compared with 1% when estimated using the commuting flows data. This difference may be attributed to measuring commuting flows from the centroids of the CTs, which might result in routing individuals on buses first before transferring them to a train. Alternatively, it could be a result of non-alignment with that train schedule, since all trips were routed at 8 a.m. on a weekday, making another public transport mode a faster option. Lastly, it could be a result of including waiting time for a public transport vehicle in the calculation of travel time, therefore for individuals who take the commuter train and adjust their departure time according to the train schedule, the fastest route algorithm may route them first on a bus if a shorter travel time to their destination is available. One solution to this issue that has been demonstrated in the literature, is to calculate travel time for multiple departures within a time period of interest, such as every minute (Farber et al., 2014), or randomly selected departure times during 5-min periods comprising a peak period (Karner, 2018), which can then be used to obtain the fastest route or the median travel time. Presently this approach is time consuming and cannot be easily adopted, and with the goal in mind of producing an adoptable method for planners and health professionals we opted not to use such approach in this paper.

The second noticeable difference in routing details between these methods are the number of transfers taken for each trip. Ten percent of trips estimated with the commuting flows method involved zero transfers compared with 41 percent of OD trips. Public transport users are often averse to transferring and in reality may choose to either walk further distances or choose longer alternative travel routes to avoid transferring. The shortest time travel algorithm used to model commuting flows trips does not include preferences or penalties for transfers. Future research can be conducted to assess how different routing preferences and/or penalties can be incorporated into routing algorithms with the aim of more closely modeling travel behaviour.

5.2. Results of the statistical models

Among the individual characteristics that were included in Model 1 (modeling total walking distance from OD microdata), age was the only variable associated with walking (Table 3). A one-year increase in age was associated with a decrease in distance walked by 2.5 m, while holding all other variables constant. Income was not a significant predictor of walking. All variables describing trip characteristics included in our model were found to be statistically significant, except for the headway of the first route taken.

Table 1
Description of variables and summary statistics.

Variable	Definition	OD microdata scenario		Commuting flows scenario	
		Mean/percent	Std. Dev	Mean	Std. Dev
Walking distance (m)	Individuals' total walking distance during home to work trip (includes distance to access first public transport stop and distance from final public transport stop to work)	1007.3	1003.47	1034.6	944.14
Individual Characteristics					
Sex	A dummy variable equaling one if the individual reported their sex as male	45%	-	N/A[a]	N/A
Age	Age of the individual in years	43.1	12.26	N/A	N/A
Middle income ($20K – 79K)	A dummy variable equaling one if the individual reported a household income between $20,000 - $79,000	46%	-	N/A	N/A
Low income (< $20K)	A dummy variable equaling one if the individual reported a household income under $20,000	11%	-	N/A	N/A
Trip characteristics					
In-vehicle distance (km) (((k(km(km)	Total distance spent in public transport vehicle(s)	10.0	7.02	10.1	7.04
Headway of first route used (min)	The headway (time between subsequent vehicles) of the first public transport route used at the beginning of the trip	16.6	33.89	N/A	N/A
First mode taken					
Subway	A dummy variable equal to one if the subway was the first mode used	30%	-	29%	-
Train	A dummy variable equal to one if the train was the first mode used	7%	-	5%	-
Peripheral bus	A dummy variable equal to one if a peripheral bus was the first mode used	5%	-	5%	-
Suburban bus	A dummy variable equal to one if a suburban bus was the first mode used	14%	-	14%	-
GTFS Dummy[a] First mode taken subway	An interaction term representing a trip modeled using GTFS data, where the first mode taken was a subway	N/A	N/A	-	-
GTFS Dummy[a] First mode taken train	An interaction term representing a trip modeled using GTFS data, where the first mode taken was a commuter train	N/A	N/A	-	-
GTFS Dummy[a] First mode taken suburban bus	An interactive term representing a trip modeled using GTFS data, where the first mode taken was a suburban bus	N/A	N/A	-	-
GTFS Dummy[a] First mode taken peripheral bus	An interactive term representing a trip modeled using GTFS data, where the first mode taken was a peripheral bus	N/A	N/A	-	-
Number of transfers					
1 transfer	A dummy variable equal to 1 if 1 transfer was undertaken during the trip	37%	-	34%	-
2 transfers	A dummy variable equal to 1 if 2 transfers were undertaken during the trip	19%	-	23%	-
3 or more transfers	A dummy variable equal to 1 if 3 or more transfers were undertaken during the trip	4%	-	10%	-
Characteristics of home neighbourhood					
Percent of the population with a university degree	Percent of people within the individuals' home CT with a university degree	27.9	13.66	27.8	13.61
Population density (km^2) of home CT	Population density of the individuals' home CT	7471.6	5090.53	7547.3	5117.05
Number of bus stops in home CT	Number of bus stops located in the CT of the individuals' home	79.4	32.46	79.7	32.26
GTFS Dummy	A dummy variable equal to one if the trip characteristics were modeled using GTFS data	N/A	N/A	0.2	0.42

[a] N/A represents a variable that was not included in the respective model.

7

Table 2

Comparing trip characteristics between detailed OD data and commuting flows scenario.

	OD microdata scenario	Commuting flows scenario
First mode taken (home-work trip) (%)		
Train	6	1
Subway	30	27
City bus	45	51
Suburban bus (STL and RTL)	14	15
Peripheral bus (CIT)	5	6
Number of transfers taken (home-work trip) (%)		
0 transfers	41	10
1 transfer	37	24
2 transfers	19	35
3 or more transfers	4	31
Average daily walking distance by first mode taken (m) [b]		
Train	4632	2932
Subway	2255	2113
City bus	1519	2010
Suburban bus (STL and RTL)	1681	2499
Peripheral bus (CIT)	2471	4410
Proportion of individuals meeting recommended daily walking target[a] (%)		
Train	71	58
Subway	24	18
City bus	10	17
Suburban bus (STL and RTL)	13	31
Peripheral bus (CIT)	32	68

[a] 30 min daily physical activity target.
[b] Daily walking distance was obtained by multiplying total walking distance during home to work commute by 2.

Previous research has observed a negative relationship between service frequency and walking distance, where service with larger waiting times reduces total walking distance (Wasfi et al., 2013). The relationship between service frequency and walking distance is likely better captured in this model by the first mode taken variable, as distances walked to access a public transport stop vary according to differences in service such as speed, comfort, and frequency (El-Geneidy et al., 2014; Fielding et al., 1978). Every additional kilometer traveled in vehicle, was associated with an additional 14 m of walking, while all other variables in the model were constant.

Train users walked on average 1.3 km farther than individuals who began their trip on a city bus, while subway users walked 581 m further than city bus users. On the other hand, suburban bus users walked 186 m less than city bus users, and similarly peripheral bus users walked 254 m less than city bus users. Using the mean value of all continuous variables, and assuming a no transfer situation, we found that individuals who began their daily commute on a train, walked on average 2.3 km on their home-work commute, which is the farthest walking distance of all modes and translates to a walking time of 25 min at a moderate pace (5.47 km/h or 3.4 miles/hour). This is followed by individuals who began their trip on a subway (1.6 km or 17 min walk), city bus users (977 m or 11 min walk), suburban bus users (790 m or 9 min walk), and lastly peripheral bus users (723 m or 8 min walk).

The number of transfers involved in completing a public transport trip was negatively associated with walking distance. When compared with trips that did not involve transferring modes or vehicles, a trip that included 1 transfer reduced the trip walking distance by 186 m. Similarly, a trip involving 2 transfers reduced walking distance by 326 m and three or more transfers by 436 m.

Lastly the model showed a negative association between walking distance and the percent of the population with a university degree and the number of bus stops within the origin CT. Every one percent increase in the population with a university degree or higher residing in the CT of trip origin was associated with a 3.5 m decrease in walking distance. At the individual level, Besser and Dannenberg (2005) observed individuals with low-levels of education walking significantly greater distances when compared with people with undergraduate or graduate-level education, which may be a result of lower personal automobile ownership rates, or that this population is more likely to reside in urban areas with better access to transit. Moreover, the results indicated that every additional bus stop present in the origin CT of a trip was linked with a decrease in walking distance by 5.7 m. The intra-class correlation coefficient showed that the origin census tract explains approximately 30% of the total variance in walking distance.

The dummy variable in Model 2 (Table 3), commuting flows scenario, was used to differentiate walking distance measured under the different scenarios. The coefficient for the commuting flows dummy variable showed that on average the commuting flows method overestimated walking distance by 366 m, when all other variables in the model were constant. However, the series of interactive terms, multiplying the commuting flows dummy variable by the first mode taken dummy, indicated that the commuting flows scenario underestimates walking for trips that began on a subway or train, whereas this method overestimated walking for trips which began on either a suburban or peripheral bus. The interactive terms must be interpreted in combination with the commuting flows and the first mode taken dummy variables. To illustrate the results, we conducted a scenario analysis to estimate walking distances achieved based on first public transport mode used. The mean value of all continuous variables was used in our calculations, and each scenario was calculated assuming no transfers occurred. Under the commuting flows scenario, trips beginning on a peripheral bus contributed to walking distances of approximately 2.1 km (23 min), which was the highest distance of all modes.

Table 3
Results of the multilevel mixed-effect regression modeling total walking distance.

Variable	Model 1				Model 2			
	Coeff.	Sig.	95% Conf. interval		Coeff.	Sig.	95% Conf. interval	
Male	10.6		−24	45.2				
Age	−2.5	c	−3.9	−1.1				
Medium income ($20K–79K)	−18.9		−80.1	42.3				
Low income (< $20K)	16.6		−21.2	54.3				
Headway of first route used (min)	0.1		−1	1.1				
In vehicle distance (km)	14.1	c	9.8	18.5	10.3	c	6.4	14.1
First mode taken (ref = city bus)								
Subway	580.6	c	524.7	636.6	580.6	c	531	630.1
Train	1307.5	c	1162.3	1452.7	1351.8	c	1254.7	1448.9
Suburban bus	−186.4	c	−295.8	−77	−109.7	a	−204.9	−14.5
Peripheral bus	−253.8	b	−428	−79.5	−82.0		−216.9	52.8
Commuting flows [a] First mode: subway					−243.5	c	−321.6	−165.5
Commuting flows [a] First mode: train					−836.4	c	−1198.3	−474.6
Commuting flows [a] First mode: suburban bus					150.3	b	56.5	244.2
Commuting flows [a] First mode: peripheral bus					820.4	c	683	957.9
One transfer	−186.4	c	−228.9	−143.2	−195.4	c	−233.3	−157.5
Two transfers	−325.7	c	−381.2	−270.2	−346.7	c	−394	−299.5
Three or more transfers	−435.9	c	−538.9	−333	−399.7	c	−467.9	−331.4
Percent of population with a university degree	−3.5	a	−6.9	−0.1	−3.1	a	−6	−0.2
Population density (1000 km²)	−10.9		−22.4	0.7	−14.5	b	−24	−5
Number of bus stops	−5.7	c	−7.6	−3.8	−4.4	c	−5.9	−3
Commuting flows Scenario					365.6	c	316	415.2
Constant	1573	c	1402.5	1743.6	1420.8	c	1288.6	1553.1
AIC	156,718				198,864			
BIC	156,854				199,012			
Observations	9549				12,220			
Log likelihood	−78,340				−99,412			

Groups	No. Groups	Intraclass correlation				Intraclass correlation		
CT origin	830	0.3			822	0.2		
CT flow					7377	0.5		

[a] 95% significance level |
[b] 99% significance level |
[c] 99.9% significance level.

Commuting by train was estimated in this scenario to contribute to 1.9 km (20 min), while subway trips contributed 1.7 km (18 min), city bus trips contributed 1.3 km (15 min), and suburban bus trips contributed 1.4 km (15 min). Potential explanations for the discrepancies in walking distances observed between the two models are discussed below.

5.3. Adjustment table

The commuting flows scenario both over and under-estimated walking distance depending on first mode taken. Differences in estimates between the two scenarios were calculated and are presented as adjustment values needed for practitioners and researchers who are estimating walking to and from public transport using the commuting flows data. OD microdata scenario estimates of walking distance obtained from Model 2 closely resembled the estimates of walking obtained from Model 1 (within the 95% confidence interval), and therefore adjustment values between the two scenarios were based on estimates from Model 2.

Table 4
Adjustment table.

First mode taken	OD microdata scenario walking distance (m)	Commuting flows scenario walking distance (m)	Adjustment (m) [a]
Commuter train	2327.83	1856.98	470.85 (20%)
Subway	1556.60	1678.64	−122.04 (−8%)
City bus	976.01	1341.57	−365.57 (−37%)
Suburban bus	866.30	1382.20	−515.91 (−60%)
Peripheral bus	893.96	2079.96	−1186.00 (−133%)

[a] These adjustment values are based on the home to work walking distance, and would thus need to be multiplied by 2 if using total daily walking distance.

The adjustment values (Table 4) indicate that the commuting flows scenario overestimated walking distance for trips that began on a bus in a peripheral region by 133%, and most closely estimated trips that began on a subway. There are two potential explanations for the large discrepancy in peripheral bus walking distances, first due to the use of CT centroids as an origin (see Fig. 3). In peripheral regions the area of a CT is much larger (average area is $15.2\,\mathrm{km}^2$) compared with a dense urban CT (average $1.2\,\mathrm{km}^2$), and the centroid of that CT might be quite distant from where individuals residing in that CT who do use public transport actually reside. Future research should explore alternative methods of defining the centroid of a CT, such as using population-weighted centroids.[2] The second potential explanation is related to the departure time of trips that were routed. We applied a rigid 8 a.m. departure time for all trips, and in peripheral regions where public transport service is infrequent this could likely result in the selection of alternative routes that involve longer walking distances than are observed in reality.

The commuting flows scenario overestimated walking distance of trips beginning on a suburban bus by 60%, a city bus by 37% and a subway by 8%. Interestingly, the overestimation of walking distance appears to decrease in regions of higher density. In dense CTs where subway stations are located (with an average area of $0.66\,\mathrm{km}^2$), the commuting flows method was able to estimate walking distance with a relatively high degree of accuracy. The commuting flows scenario underestimated average walking distances to commuter rail by 20%, although it is important to note that a very low number of trips (1%) were routed on the commuter train, whereas 6% of trips in the OD microdata began on commuter train. The willingness of commuters to walk farther distances to use a faster and more comfortable service, such as rail, is not captured in the fastest route algorithm. While the commuting flows method has some shortcomings when it comes to calculating and estimating overall walking to public transport in a region, such methods can be used to generate overall walking minutes accumulated by individuals commuting and help in proposing accurate public health policy interventions if an adjustment table similar to the one below is used, which is the main goal of this research.

6. Discussion

The study shows how open data (commuting flows scenario) can be used to estimate walking distances associated with commuting by public transport and the kind of adjustments needed when using such data to obtain reliable measures of walking to support future public health policies. Our results indicated that walking distances estimated using the commuting flows scenario closely resemble actual walking distances derived from origin-destination microdata. Specifically, we found that estimating walking to public transport using the commuting flows method underestimated average walking distances by only 3%. The accuracy of the commuting flows method varied depending on the first mode of public transport used for each trip. Using a scenario analysis, we compared walking estimates for each mode and presented an adjustment table. The adjustment values must be applied to estimates of walking distance obtained using the commuting flows method to realistically uncover how walking to public transport contributes to daily physical activity goals and to estimate the number of individuals meeting the recommended levels of physical activity through the use of public transport. Knowledge of how commuting by public transport contributes to physical activity levels of a given population is vital knowledge for understanding baseline physical activity levels of a population and for setting goals for future public health interventions.

Walking distances measured in this study resemble previous research contributions that have presented evidence of the levels of physical activity that can be attained on a daily basis when commuting by public transport. Consistent with previous literature, commuting by train is associated with the highest levels of walking relative to all other public transport modes (Besser and Dannenberg, 2005; Wasfi et al., 2013). For effective public health interventions, it is important for municipalities to have a base level understanding of how many residents are commuting by public transport, and the daily physical activity levels that those commuters are attaining from their daily commute. While, detailed travel survey data is a highly effective source of data to measure walking levels, the collection of a representative sample of the population is very costly. The commuting flows scenario presented in this paper is particularly beneficial for professionals and policy makers who wish to better understand regional levels of physical activity and to better understand local determinants of public transport use. One strength of this method lies in the usage of openly available census flows and GTFS data. GTFS data is present across major cities in Canada and other major cities around the world, whilst the commuting flows are present in many cities as well. The adjustment values presented from this comparative analysis will enable other municipalities to attain accurate estimates of daily walking to public transport or to generate similar ones using local context information when available to them.

6.1. Study limitations

When calculating walking distance for both scenarios, the distance walked while transferring was not measured. Using the example of a transfer from a bus to a subway, accessing the subway platform from street level can take several minutes. Similarly, the access time to a train or subway platform was not measured, certain station designs can contribute to additional minutes of physical activity, particularly when individuals has to use stairs.

A limitation associated with measuring walking to commuter rail stations is that we do not know which trips an individual drove to a station and parked there, as only main mode is reported in the database of the commuting flow. There are over 33,000 park-and-ride spaces across the Montreal commuter train network, and station access by park-and-ride is highly prevalent in suburban and peripheral regions of Montreal, while in higher density areas station access by other modes, walking, bicycle and public transport are

---

[2] A population-weighted centroid represents the centre of the population in the CT.

more common (Steer Davies Gleave, 2016). Walking distances of public transport trips that begin by park-and-ride are overestimated using this method.

Lastly, it is important to acknowledge that this method is only capturing walking associated with public transport commute trips, and these trips represent 10 percent of all trips that were observed in the Origin-Destination survey data applied in this study. With respect to the census flows data, this data omits walking information for informally employed individuals, and does not capture walking associated with trip chaining behaviour (i.e. dropping off children and grocery trips).

References

Agence Métropolitaine de Transport, 2013. Enquête Origine-Destination 2013, Montreal, Quebec.

Alshalalfah, B., Shalaby, A., 2007. Case study: relationship of walk access distance to transit with service, travel, and personal characteristics. J. Urban Plan. Dev. 133, 114–118.

Besser, L., Dannenberg, A., 2005. Walking to public transit: steps to help meet physical activity recommendations. Am. J. Prev. Med. 29, 273–280.

Booth, M., Bauman, A., Owen, N., Gore, C., 1997. Physical activity preferences, preferred sources of assistance, and perceived barriers to increased activity among physically inactive Australians. Prev. Med. 26, 131–137.

Borodulin, K., Sipilä, N., Rahkonen, O., Leino-Arjas, P., Kestilä, L., Jousilahti, P., Prättälä, R., 2016. Socio-demographic and behavioral variation in barriers to leisure-time physical activity. Scand. J. Public Health 44, 62–69.

Brown, B., Werner, C., 2007. A new rail stop: tracking moderate physical activity bouts and ridership. Am. J. Prev. Med. 33, 306–309.

Canadian Society for Exercise Physiology, 2018. Canadian 24-hour Guidelines: an Integration of Physical Activity, Sedentary Behaviour, and Sleep.

Chapleau, R., Trépanier, M., Allard, B., 1997. Origin-destination survey data dissemination in a metropolitan context: a multimedia experience. Transp. Res. Rec.: J. Transp. Res. Board 26–36.

Daniels, R., Mulley, C., 2013. Explaining walking distance to public transport: the dominance of public transport supply. J. Transp. Land Use 6, 5–20.

Davies Gleave, Steer, 2016. Transport Collectif Train de l'Ouest de Montreal, Montreal. https://www.cdpqinfra.com/sites/default/files/pdf/9_achalandage_mars_2016.pdf.

Dora, C., 1999. A different route to health: implications of transport policies. BMJ 318, 1686–1689.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., Surprenant-Legault, J., 2014. New evidence on walking distances to transit stops: identifying redundancies and gaps using variable service areas. Transportation 41, 193–210.

Farber, S., Morang, M., Widener, M., 2014. Temporal variability in transit-based accessibility to supermarkets. Appl. Geogr. 53, 149–159.

Fielding, G., Glauthier, R., Lave, C., 1978. Performance indicators for transit management. Transportation 7, 365–379.

Godin, G., Desharnais, R., Valois, P., Lepage, L., Jobin, J., Bradet, R., 1994. Differences in perceived barriers to exercise between high and low intenders: observations among different populations. Am. J. Health Promot. 8, 279–285.

Jiang, Y., Zegras, P., Mehndiratta, S., 2012. Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China. J. Transp. Geogr. 20, 1–14.

Karner, A., 2018. Assessing public transit service equity using route-level accessibility measures and public data. J. Transp. Geogr. 67, 24–32.

Koplan, J., Dietz, W., 1999. Caloric imbalance and public health policy. JAMA, J. Am. Med. Assoc. 282, 1579–1581.

MacDonald, J., Stokes, R., Cohen, D., Kofner, A., Ridgeway, G., 2010. The effect of light rail transit on body mass index and physical activity. Am. J. Prev. Med. 39, 105–112.

Morency, C., Trépanier, M., Demers, M., 2011. Walking to transit: an unexpected source of physical activity. Transp. Policy 18, 800–806.

Murray, A., Wu, X., 2003. Accessibility tradeoffs in public transit planning. J. Geogr. Syst. 5, 93–107.

O'Sullivan, S., Morrall, J., 1996. Walking distances to and from light-rail transit stations. Transp. Res. Rec.: J. Transp. Res. Board 19–26.

Owen, N., Bauman, A., 1992. The descriptive epidemiology of a sedentary lifestyle in adult Australians. Int. J. Epidemiol. 21, 305–310.

Patterson, R., Webb, E., Millett, C., Laverty, A., 2019. Physical activity accrued as part of public transport use in England. J. Public Health 41 (2), 222–230. https://doi.org/10.1093/pubmed/fdy099.

Public Health Agency of Canada, 2016. Health Status of Canadians 2016: Report of the Chief Public Health Officer. . https://www.canada.ca/en/public-health/corporate/publications/chief-public-health-officer-reports-state-public-health-canada/2016-health-status-canadians.html.

She, Z., King, D., Jacobson, S., 2017. Analyzing the impact of public transit usage on obesity. Prev. Med. 99, 264–268.

She, Z., King, D., Jacobson, S., 2019. Is promoting public transit an effective intervention for obesity? A longitudinal study of the relation between public transit usage and obesity. Transp. Res. A Policy Pract. 119, 162–169.

Siegel, P., Brackbill, R., Heath, G., 1995. The epidemiology of walking for exercise: implications for promoting activity among sedentary groups. Am. J. Public Health 85, 706–710.

Statistics Canada, 2011. Employed Labour Force 15 Years and over Having a Usual Place of Work by Income Groups in 2010 - Commuting Flow Census Metropolitan Areas, Tracted Census Agglomerations and Census Tracts.

Trépanier, M., Chapleau, R., Morency, C., 2008. Tools and methods for transportation household survey. URISA J. 20, 35–43.

van Soest, D., Tight, M., Rogers, C., 2019. Exploring the distances people walk to access public transport. Transp. Rev. 1–23.

Veillette, M., Deboosere, R., Wasfi, R., Ross, N., El-Geneidy, A., 2018. Small Steps, Big Differences: Assessing the Validity of Using Home and Work Locations to Estimate Walking Distances to Transit. Transportation Research Record: Journal of the Transportation Research Record in Press.

Wasfi, R., Ross, N., El-Geneidy, A., 2013. Achieving recommended daily physical activity levels through commuting by public transportation: unpacking individual and contextual influences. Health Place 23, 18–25.

Weinstein, A., Schlossberg, M., Irvin, K., 2008. How far, by which route and why? A spatial analysis of pedestrian preference. J. Urban Des. 13, 81–98.

World Health Organization, 2010. Global Reccomendations on Physical Activity for Health. World Health Organization, Switzerland.

World Health Organization, 2017. 10 Facts on Physical Activity.

Yoshida, K., Allison, K., Osborn, R., 1988. Social factors influencing perceived barriers to physical exercise among women. Can. J. Public Health 79, 104–108.

Zhao, F., Chow, L., Li, M., Ubaka, I., Gan, A., 2003. Forecasting transit walk accessibility: regression model alternative to buffer method. Transp. Res. Rec.: J. Transp. Res. Board 34–41.