

1                   **Evaluating methods for measuring daily walking to public**  
2                   **transport: Balancing accuracy and data availability**  
3  
4

5                   Grisé, E., Wasfi, R., Ross, N., & El-Geneidy, A.  
6

7 For Citation please use: Grisé, E., Wasfi, R., Ross, N., & El-Geneidy, A. (2019).  
8 Evaluating methods for measuring daily walking to public transport: Balancing  
9 accuracy and data availability. *Transport and Health*, 15, 2-11.

10 **ABSTRACT**

11 **Background**

12 Quantifying physical activity accumulated through daily commuting is challenging due to the  
13 scarcity of detailed data, especially for public transport trips. Using Montreal, Canada as a case  
14 study, this paper measures and compares an individual’s daily amount of walking to and from  
15 public transport in their regular commute to work using two datasets and two methods.

16 **Methods**

17 The first method used urban level detailed origin-destination microdata. Distances walked to and  
18 from public transport stations were measured using trip details provided from the survey. The  
19 second method used open data including commuting flows obtained from census data. Public  
20 transport trips for each flow were modeled using a fastest route algorithm applied to General  
21 Transit Feed Specification (GTFS) data obtained from public transport operators. Walking  
22 distances were then extracted from these trip paths. Multilevel mixed-effect regression modeling  
23 was used to identify the determinants of total walking when using both methods. A sensitivity  
24 analysis was then used to derive an adjustment table for those who wish to use open data to estimate  
25 walking.

26 **Results**

27 For commuter train users 471 metres must be added to walking estimates obtained from the  
28 commuting flows data, while negative adjustments are required for subway users (122 metres),  
29 city bus users (366 metres), suburban bus users (516 metres), and peripheral bus users (1186  
30 metres).

31 **Conclusions**

32 The methodology presented in this study provides researchers and professionals in other cities  
33 without access to detailed origin-destination survey data with a guide to use open data to accurately  
34 estimate total walking distances accumulated in a daily commute by public transport.  
35

36 **Keywords:** Public transport, walking, travel behaviour, daily physical activity  
37

1 **INTRODUCTION**

2 Being active for at least 150 minutes a week can help reduce the risk of heart disease, type 2  
3 diabetes, high blood pressure, and obesity (Canadian Society for Exercise Physiology, 2018). Only  
4 two in ten Canadian adults and one in ten children met the Canadian physical activity guidelines  
5 in 2013 (Public Health Agency of Canada, 2016). While the most commonly cited barrier to  
6 exercise is lack of time (Booth et al., 1997; Borodulin et al., 2016; Godin et al., 1994; Owen and  
7 Bauman, 1992; Yoshida et al., 1988), the good news is that individuals can achieve the  
8 recommended physical activity guidelines through small sessions of at least ten minutes of activity  
9 throughout the day (World Health Organization, 2017). Walking accumulated through daily travel  
10 is one way that individuals can meet physical activity targets. Compared with public health  
11 interventions that aim to promote active lifestyles and change the behaviour of individuals, policies  
12 that aim to increase the proportion of trips that are made on active modes of travel such as public  
13 transport, bicycling and walking, are likely to increase the activity levels of a given population in  
14 a way that can be sustained in the long term (Siegel et al., 1995).

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

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

38

39 **LITERATURE REVIEW**

40 In response to rising obesity levels observed in many Westernized countries, researchers from  
41 leading medical and public health have advocated for the increased uptake of walking and  
42 bicycling as a means of achieving daily moderate-intensity physical activity goals for improved  
43 health (Dora, 1999; Koplan and Dietz, 1999). The specific level of physical activity achieved  
44 through daily public transport use is mediated by many factors. The most significant factor

1 determining physical activity levels of public transport users is the mode being accessed. Distances  
2 walked to buses are generally shorter compared with commuter trains and subways (Alshalalfah  
3 and Shalaby, 2007; Wasfi et al., 2013), and thus catchment areas have traditionally been defined  
4 as 400 metres for bus and 800 metres for rail (for example Fielding et al., 1978; Murray and Wu,  
5 2003; Zhao et al., 2003). However, several authors have observed walking distances to public  
6 transport beyond the 400 and 800 metres threshold, thus presenting an opportunity for operators  
7 to increase stop spacing (Daniels and Mulley, 2013; El-Geneidy et al., 2014; van Soest et al., 2019;  
8 Weinstein et al., 2008). Variations have also been observed between downtown and suburban  
9 locations with suburban commuters tending to walk further to more comfortable and faster modes  
10 (El-Geneidy et al., 2014; O'Sullivan and Morrall, 1996). With respect to built environment  
11 characteristics, van Soest et al. (2019) observed through a systematic review of literature on the  
12 topic of walking to public transport that in higher density environments public transport services  
13 are often densely located thereby reducing distances people need to walk to access public transport  
14 services. Alternatively, density can be associated with high levels of local accessibility, which can  
15 be attractive to walkers, thereby either increasing the share of walking to access/egress transport  
16 or distance people will walk to public transport (Jiang et al., 2012), or by expanding the catchment  
17 area around stops or stations.

18 Using detailed origin-destination survey data in Montreal, Canada, Wasfi et al. (2013)  
19 modeled the underlying factors that contribute to achieving the recommended 30 minutes of daily  
20 physical activity while commuting by public transport. The authors concluded that approximately  
21 11% of commuters achieved their recommended minutes of daily physical activity, and suburban  
22 train users walked the greatest number of minutes per day. Similar findings were observed by  
23 Besser and Dannenberg (2005), who used the National Household Travel Survey to assess walking  
24 times to public transport across the United States of America. The authors observed that 29% of  
25 public transport users achieved more than 30 minutes of daily physical activity from walking to  
26 and from stops, and that rail users were 1.67 times more likely than bus users to achieve their daily  
27 physical activity target. Using data from the English National Travel Survey, Patterson et al. (2018)  
28 observed that one in three public transport users achieved 30 minutes of daily physical activity:  
29 21% bus users achieved this daily physical activity goal and 52% of train users met their physical  
30 activity target through taking public transport. Although these studies find that not all public  
31 transport riders regularly meet their daily physical activity guidelines solely through walking to  
32 public transport, it is important to note that individuals walking to public transport are reaching  
33 some level of physical activity that they would not have if commuting by a sedentary form of  
34 transport, such as driving.

35 More recently, researchers have been interested in understanding the exertion level of  
36 public transport related walking to ascertain whether these walkers are in fact meeting the required  
37 moderate-to-vigorous physical activity (MVPA) intensity level set within physical activity  
38 guidelines (World Health Organization, 2010). Using accelerometer data collected from  
39 participants residing within a half mile of a new light-rail stop in Salt Lake City, USA, Brown and  
40 Werner (2007) measured the length and intensity of walking trips to and from light rail stops,  
41 concluding that these walking trips are of moderate-intensity, and collectively can contribute to  
42 attaining the WHO-recommended physical activity level. MacDonald et al. (2010) studied the  
43 impact of a new light rail transit (LRT) line in Charlotte, USA on obesity and physical activity  
44 level. These authors found that commuting by LRT was associated with a reduction in BMI and a  
45 reduced odds of becoming obese over time, compared with a control group of individuals living  
46 in similar neighbourhoods but did not commute by LRT. At an aggregate level, She et al. (2019)

1 observed through a longitudinal analysis of county-level health and transport data that in the  
2 United State a 1% increase in the share of frequent public transport users in a county population is  
3 estimated to decrease the county population obesity rate by 0.47%. Similarly, She et al. (2017)  
4 found a 0.22% decrease in county population obesity prevalence through a cross sectional analysis  
5 of the United States 2009 Behavioral Risk Factor Surveillance System and the 2009 National  
6 Household Travel Survey.

7 Most studies to date have measured walking to public transport using either travel survey  
8 data (Besser and Dannenberg, 2005; Morency et al., 2011; Wasfi et al., 2013), or readings from  
9 biosensors (Brown and Werner, 2007). However, professionals often lack either the access to such  
10 detailed data or the resources to recruit participants and monitor physical activity levels at such a  
11 detailed level. Biosensor-assessed step counts are also burdensome for respondents and can induce  
12 non-routine behaviour. Veillette et al. (2018) recently calculated levels of walking to public  
13 transport using detailed and non-detailed travel survey data. While this study provided a better  
14 understanding of how modeled walking routes differ from reported routes, it did not present a  
15 methodology for professionals without access to travel survey data to measure walking to public  
16 transport. In this study we examine how the accuracy of the estimation of walking distances  
17 accumulated while taking public transport using publicly available commuting flow census data  
18 and public transport schedule data (GTFS data).  
19

## 20 DATA

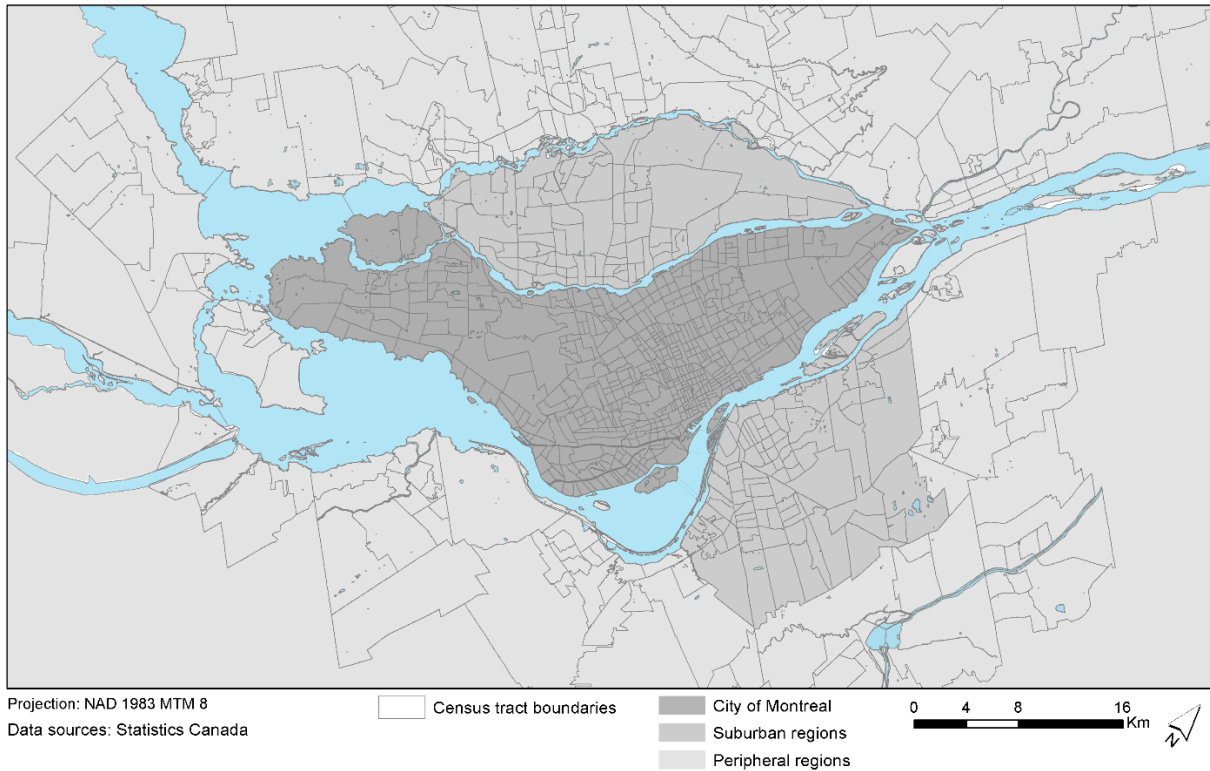
### 21 *Origin-Destination Microdata*

22 A common method for cities to evaluate the travel behaviour of residents is through the collection  
23 of origin-destination (OD) surveys. OD microdata are used by cities in many ways including  
24 modeling travel demand and evaluating the modal split. The Montreal, Canada, OD survey has  
25 been typically conducted every five years by the Agence Métropolitaine de Transport (AMT)  
26 (Agence Métropolitaine de Transport, 2013). In our analysis we used the 2013 OD microdata. Five  
27 percent of households are sampled and asked to recall all trips made in the past day by each  
28 household member. This data includes coordinates of the location of the home and each location  
29 visited in that day, mode taken for each trip, and personal and household characteristics. For trips  
30 taken by public transport, each respondent was asked to provide route details including bus route(s)  
31 used, subway or train station of access and egress, and subway used. After selecting all home-  
32 based trips for the purpose of work or school that were taken by adults (over the age of 18) that  
33 began and ended in Greater Montreal, our sample consisted of 9,588 trips that were made using  
34 bus, subway or commuter rail. We omitted trips with missing route information, as well as trips  
35 where individuals drove to a public transport station.  
36

### 37 *Commuting Flows*

38 Statistics Canada collects information on the commuting modes and patterns of all Canadians as  
39 part of the Census of Canada. Commuting modes and patterns are available at the level of a census  
40 tract (CT) – a small urban area that is generally socially homogeneous and made up of between  
41 2,500 to 8,000 people (Figure 1). Similar data exist in other countries around the world, making  
42 this approach appealing for international comparative research. We selected all commuting flow  
43 pairs within the Greater Montreal Region where there were at least 10 individuals who commute  
44 by public transport between CT pairs (smaller counts were suppressed for confidentiality). Our  
45 sample consisted of 2,755 commuting flows. Such data can be present at finer details in some

1 countries with large populations – the U.S., for example, provides this information at the block  
2 level, yet in the Canadian context such detail will be subjected to heavy data suppression so we  
3 opted to use the larger CT as the basic unit of analysis.



4  
5 Figure 1: Census tract boundaries within Montreal Metropolitan Region  
6

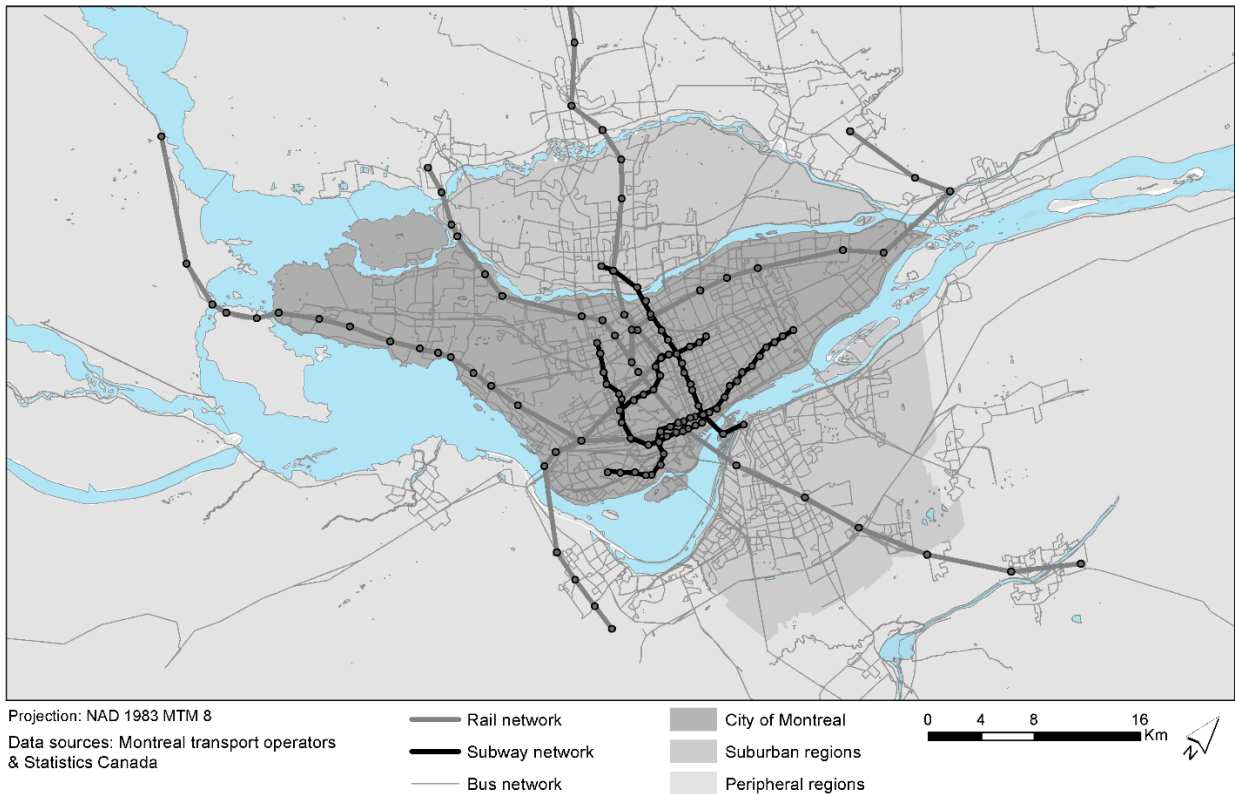
## 7 **METHODOLOGY**

### 8 *Origin-Destination Microdata Scenario*

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

21 Using the stop where each individual began their trip, we measured the headway (time  
22 between subsequent vehicles) of the first public transport route during the morning peak  
23 commuting time (7 and 9 AM). We also measured the distance travelled in-vehicle, to provide a  
24 complete picture of the individual's daily public transport commute. This was done using the

1 previously identified first and last stop used in their trip, and by determining the length of the  
 2 public transport trip between these points using a network built from GTFS data from all public  
 3 transport operators in the Greater Montreal region. GTFS data produced in 2013 was used to  
 4 represent the level of service that was present at the time of data collection. The number of transfers  
 5 that occurred during this trip was calculated as the total number of distinct public transport routes  
 6 used during this trip minus one. Note, we defined a transfer as a route or mode switch that occurred  
 7 during any trip. For example, a trip that began on a bus and then required two subway lines would  
 8 be a two-transfer trip.  
 9



10  
 11 **Figure 2: Public transit network in Montreal Metropolitan Region**

12  
 13 *Commuting Flows Scenario*

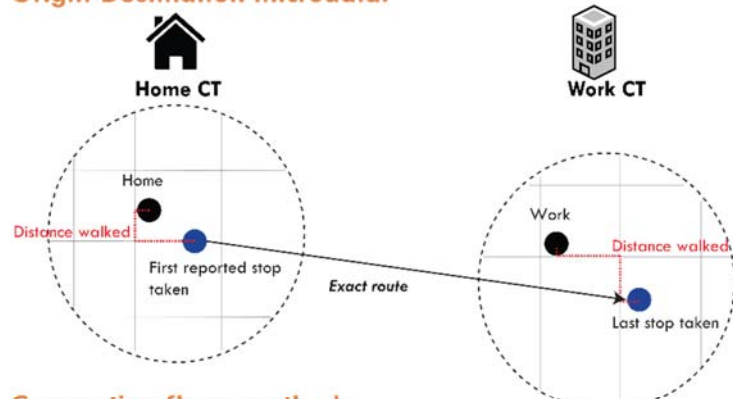
14 Using the CT centroids, we determined the public transport path of all flows according to the  
 15 shortest travel time. GTFS data from all public transport operators in the Greater Montreal region  
 16 were obtained and a public transport network was constructed in ArcGIS using the ‘Add GTFS to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

<sup>1</sup> 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/AddGTFSStoaNetworkDataset.html>

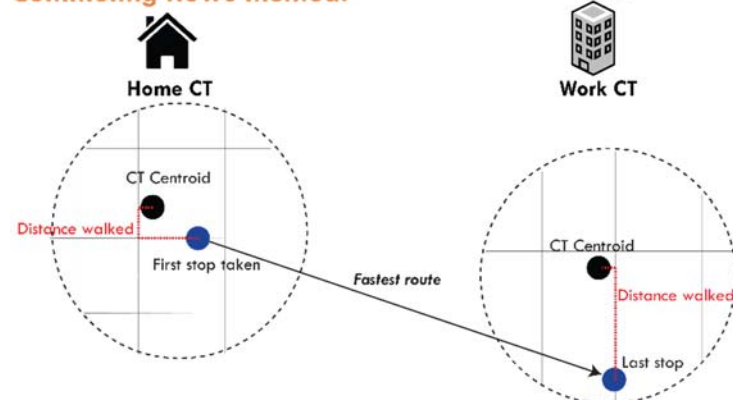
1 The portion of each trip that required the individual to walk to the first and last public  
 2 transport stop was calculated by intersecting each modeled trip flow with the street network, then  
 3 we applied the “unsplit line tool” to merge coinciding walking segments. We selected only the  
 4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

9 The number of transfers was obtained by subtracting the number of walking segments  
 10 identified in step two above. These represent walking segments at the origin and destination. To  
 11 represent transfers occurring between subways and trains we determined the number of subway  
 12 and rail transfer stations each user passed through and adjusted the number of transfers to account  
 13 for transferring to a subway or a train line. The first mode taken was determined by isolating the  
 14 walking segment from each respondent’s home to the first stop and identifying the type of public  
 15 transport route that intersected with the first walking segment. Lastly, we determined whether a  
 16 trip began on a city, suburban, or a peripheral bus by determining the agency boundary in which  
 17 each trip originated.  
 18

**Origin-Destination microdata:**



**Commuting flows method:**



19  
 20 Figure 3: Schematic diagram illustrating calculation of distance walked using both methods

21 **Statistical analysis**

22 Two multilevel mixed-effect regression models were constructed to uncover the key determinants  
 23 of daily walking to public transport as well as to explore how the two approaches differed in their

1 estimation of walking levels (under the assumption that the walking distances obtained from OD  
2 microdata are the gold standard as the OD has been extensively validated in the past (Chapleau et  
3 al., 1997; Trépanier et al., 2008)). The first model evaluated the determinants of walking to public  
4 transport according to OD microdata, to provide a baseline understanding of what factors influence  
5 walking levels. In this model, we nested each observation in its respective origin CT, to account  
6 for characteristics in the built environment unique to each neighbourhood that may influence  
7 walking behaviour. To account for unmeasured characteristics of each respondent's local built  
8 environment, the multilevel model approach adopted reduces the potential biases that may be  
9 imposed on the model due to not controlling for such characteristics.

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

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



**Table 1 Description of variables and summary statistics**

Variable	Definition	OD microdata scenario		Commuting flows scenario	
		Mean/percent	Std. Dev	Mean	Std. Dev
<b>Walking distance (m)</b>	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
<b>Individual Characteristics</b>					
<b>Sex</b>	A dummy variable equaling one if the individual reported their sex as male	45%	--	N/A*	N/A
<b>Age</b>	Age of the individual in years	43.1	12.26	N/A	N/A
<b>Middle income (\$20K - 79K)</b>	A dummy variable equaling one if the individual reported a household income between \$20,000 - \$79,000	46%	--	N/A	N/A
<b>Low income (&lt; \$20K)</b>	A dummy variable equaling one if the individual reported a household income under \$20,000	11%	--	N/A	N/A
<b>Trip characteristics</b>					
<b>In-vehicle distance (km)</b>	Total distance spent in public transport vehicle(s)	10.0	7.02	10.1	7.04
<b>Headway of first route used (min)</b>	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
<b>First mode taken</b>					
<b>Subway</b>	A dummy variable equal to one if the subway was the first mode used	30%	--	29%	--
<b>Train</b>	A dummy variable equal to one if the train was the first mode used	7%	--	5%	--
<b>Peripheral bus</b>	A dummy variable equal to one if a peripheral bus was the first mode used	5%	--	5%	--

<b>Suburban bus</b>	A dummy variable equal to one if a suburban bus was the first mode used	14%	--	14%	--
<b>GTFS Dummy * First mode taken subway</b>	An interaction term representing a trip modeled using GTFS data, where the first mode taken was a subway	N/A	N/A	--	--
<b>GTFS Dummy * First mode taken train</b>	An interactive term representing a trip modeled using GTFS data, where the first mode taken was a commuter train	N/A	N/A	--	--
<b>GTFS Dummy * First mode taken suburban bus</b>	An interactive term representing a trip modeled using GTFS data, where the first mode taken was a suburban bus	N/A	N/A	--	--
<b>GTFS Dummy * First mode taken peripheral bus</b>	An interactive term representing a trip modeled using GTFS data, where the first mode taken was a peripheral bus	N/A	N/A	--	--
<b>Number of transfers</b>					
<b>1 transfer</b>	A dummy variable equal to 1 if 1 transfer was undertaken during the trip	37%	--	34%	--
<b>2 transfers</b>	A dummy variable equal to 1 if 2 transfers were undertaken during the trip	19%	--	23%	--
<b>3 or more transfers</b>	A dummy variable equal to 1 if 3 or more transfers were undertaken during the trip	4%	--	10%	--
<b>Characteristics of home neighbourhood</b>					
<b>Percent of the population with a university degree</b>	Percent of people within the individuals' home CT with a university degree	27.9	13.66	27.8	13.61
<b>Population density (km<sup>2</sup>) of home CT</b>	Population density of the individuals' home CT	7471.6	5090.53	7547.3	5117.05
<b>Number of bus stops in home CT</b>	Number of bus stops located in the CT of the individuals' home	79.4	32.46	79.7	32.26
<b>GTFS Dummy</b>	A dummy variable equal to one if the trip characteristics were modeled using GTFS data	N/A	N/A	0.2	0.42

\* N/A represents a variable that was not included in the respective model

1 **RESULTS**

2 **Comparing trip details by scenario**

3 The average walking distance of a home to work commute as measured in the OD microdata  
4 scenario was 1,007 metres, compared with 1,035 metres estimated by the commuting flows  
5 method. On average, the commuting flows method overestimated daily walking distance by 3%.  
6 Important differences in average walking distances between the two scenarios are apparent when  
7 comparing trip characteristics (Table 2). Looking at average walking distances by first mode taken,  
8 we observe that commuting flow trips that began on a peripheral bus overestimated walking by  
9 44% when compared with detailed OD trips. Estimates of daily walking distances for trips  
10 beginning on a subway were most closely estimated by the commuting flows method  
11 (underestimated by 142 metres). In line with the overall goal of increasing the proportion of the  
12 population meeting the recommended daily physical activity goals (30 minutes a day), we  
13 calculated the proportion of individuals meeting the daily physical activity goals through walking  
14 to and from public transport stations. Results of the commuting flows scenario show an  
15 underestimation of the proportion of train users that meet the recommended daily walking target,  
16 while this scenario overestimates the proportion of city bus, suburban and peripheral bus users  
17 who meet the recommended physical activity targets.

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

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

43  
44  
45

1 **TABLE 2 Comparing trip characteristics between detailed OD data and commuting flows**  
 2 **scenario**

	OD microdata scenario	Commuting flows scenario
<b>First mode taken (home-work trip) (%)</b>		
Train	6	1
Subway	30	27
City bus	45	51
Suburban bus (STL and RTL)	14	15
Peripheral bus (CIT)	5	6
<b>Number of transfers taken (home-work trip) (%)</b>		
0 transfers	41	10
1 transfer	37	24
2 transfers	19	35
3 or more transfers	4	31
<b>Average daily walking distance by first mode taken (m) †</b>		
Train	4,632	2,932
Subway	2,255	2,113
City bus	1,519	2,010
Suburban bus (STL and RTL)	1,681	2,499
Peripheral bus (CIT)	2,471	4,410
<b>Proportion of individuals meeting recommended daily walking target* (%)</b>		
Train	71	58
Subway	24	18
City bus	10	17
Suburban bus (STL and RTL)	13	31
Peripheral bus (CIT)	32	68

\*30 minute daily physical activity target

† Daily walking distance was obtained by multiplying total walking distance during home to work commute by 2

3  
 4 **Results of the statistical models**

5 Among the individual characteristics that were included in Model 1 (modeling total walking  
 6 distance from OD microdata), age was the only variable associated with walking (Table 3). A one-  
 7 year increase in age was associated with a decrease in distance walked by 2.5 metres, while holding  
 8 all other variables constant. Income was not a significant predictor of walking. All variables  
 9 describing trip characteristics included in our model were found to be statistically significant,  
 10 except for the headway of the first route taken. Previous research has observed a negative  
 11 relationship between service frequency and walking distance, where service with larger waiting  
 12 times reduces total walking distance (Wasfi et al., 2013). The relationship between service  
 13 frequency and walking distance is likely better captured in this model by the first mode taken  
 14 variable, as distances walked to access a public transport stop vary according to differences in  
 15 service such as speed, comfort, and frequency (El-Geneidy et al., 2014; Fielding et al., 1978).  
 16 Every additional kilometer traveled in vehicle, was associated with an additional 14 metres of  
 17 walking, while all other variables in the model were constant.

18 Train users walked on average 1.3 kilometres farther than individuals who began their trip  
 19 on a city bus, while subway users walked 581 metres further than city bus users. On the other hand,  
 20 suburban bus users walked 186 metres less than city bus users, and similarly peripheral bus users

1 walked 254 metres less than city bus users. Using the mean value of all continuous variables, and  
2 assuming a no transfer situation, we found that individuals who began their daily commute on a  
3 train, walked on average 2.3 kilometres on their home-work commute, which is the farthest  
4 walking distance of all modes and translates to a walking time of 25 minutes at a moderate pace  
5 (5.47 km/h or 3.4 miles/hour). This is followed by individuals who began their trip on a subway  
6 (1.6 km or 17 minute walk), city bus users (977 metres or 11 minute walk), suburban bus users  
7 (790 metres or 9 minute walk), and lastly peripheral bus users (723 metres or 8 minute walk).

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

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

24

**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	***	-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	***	9.8	18.5	10.3	***	6.4	14.1
First mode taken ( <i>ref</i> = city bus)								
Subway	580.6	***	524.7	636.6	580.6	***	531	630.1
Train	1307.5	***	1162.3	1452.7	1351.8	***	1254.7	1448.9
Suburban bus	-186.4	***	-295.8	-77	-109.7	*	-204.9	-14.5
Peripheral bus	-253.8	**	-428	-79.5	-82.0		-216.9	52.8
Commuting flows * First mode: subway					-243.5	***	-321.6	-165.5
Commuting flows * First mode: train					-836.4	***	-1198.3	-474.6
Commuting flows * First mode: suburban bus					150.3	**	56.5	244.2
Commuting flows * First mode: peripheral bus					820.4	***	683	957.9
One transfer	-186.4	***	-228.9	-143.2	-195.4	***	-233.3	-157.5
Two transfers	-325.7	***	-381.2	-270.2	-346.7	***	-394	-299.5
Three or more transfers	-435.9	***	-538.9	-333	-399.7	***	-467.9	-331.4
Percent of population with a university degree	-3.5	*	-6.9	-0.1	-3.1	*	-6	-0.2
Population density (1000 km <sup>2</sup> )	-10.9		-22.4	0.7	-14.5	**	-24	-5
Number of bus stops	-5.7	***	-7.6	-3.8	-4.4	***	-5.9	-3
Commuting flows Scenario					365.6	***	316	415.2
Constant	1573	***	1402.5	1743.6	1420.8	***	1288.6	1553.1
AIC	156,718				198,864			
BIC	156,854				199,012			
Observations	9,549				12,220			
Log likelihood	-78,340				-99,412			
<b>Groups</b>	No. Groups	Intraclass correlation			Intraclass correlation			
CT origin	830	0.3			822	0.2		
CT flow					7,377	0.5		

\* 95% significance level | \*\* 99% significance level | \*\*\* 99.9% significance level

1 The dummy variable in Model 2 (Table 3), commuting flows scenario, was used to  
2 differentiate walking distance measured under the different scenarios. The coefficient for the  
3 commuting flows dummy variable showed that on average the commuting flows method  
4 overestimated walking distance by 366 metres, when all other variables in the model were  
5 constant. However, the series of interactive terms, multiplying the commuting flows dummy  
6 variable by the first mode taken dummy, indicated that the commuting flows scenario  
7 underestimates walking for trips that began on a subway or train, whereas this method  
8 overestimated walking for trips which began on either a suburban or peripheral bus. The interactive  
9 terms must be interpreted in combination with the commuting flows and the first mode taken  
10 dummy variables. To illustrate the results, we conducted a scenario analysis to estimate walking  
11 distances achieved based on first public transport mode used. The mean value of all continuous  
12 variables was used in our calculations, and each scenario was calculated assuming no transfers  
13 occurred. Under the commuting flows scenario, trips beginning on a peripheral bus contributed to  
14 walking distances of approximately 2.1 kilometres (23 minutes), which was the highest distance  
15 of all modes. Commuting by train was estimated in this scenario to contribute to 1.9 kilometres  
16 (20 minutes), while subway trips contributed 1.7 kilometres (18 minutes), city bus trips contributed  
17 1.3 kilometres (15 minutes), and suburban bus trips contributed 1.4 kilometres (15 minutes).  
18 Potential explanations for the discrepancies in walking distances observed between the two models  
19 are discussed below.

## 21 **Adjustment table**

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

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

41 The commuting flows scenario overestimated walking distance of trips beginning on a  
42 suburban bus by 60%, a city bus by 37% and a subway by 8%. Interestingly, the overestimation  
43 of walking distance appears to decrease in regions of higher density. In dense CTs where subway  
44 stations are located (with an average area of 0.66 km<sup>2</sup>), the commuting flows method was able to

---

<sup>2</sup> A population-weighted centroid represents the centre of the population in the CT.

1 estimate walking distance with a relatively high degree of accuracy. The commuting flows  
 2 scenario underestimated average walking distances to commuter rail by 20%, although it is  
 3 important to note that a very low number of trips (1%) were routed on the commuter train, whereas  
 4 6% of trips in the OD microdata began on commuter train. The willingness of commuters to walk  
 5 farther distances to use a faster and more comfortable service, such as rail, is not captured in the  
 6 fastest route algorithm. While the commuting flows method has some shortcomings when it comes  
 7 to calculating and estimating overall walking to public transport in a region, such methods can be  
 8 used to generate overall walking minutes accumulated by individuals commuting and help in  
 9 proposing accurate public health policy interventions if an adjustment table similar to the one  
 10 below is used, which is the main goal of this research.

11  
 12 **Table 4 Adjustment table**

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

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

15  
 16 **DISCUSSION**

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



1 understanding baseline physical activity levels of a population and for setting goals for future  
2 public health interventions.

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

## 21 **Study Limitations**

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

27 A limitation associated with measuring walking to commuter rail stations is that we do not  
28 know which trips an individual drove to a station and parked there, as only main mode is reported  
29 in the database of the commuting flow. There are over 33,000 park-and-ride spaces across the  
30 Montreal commuter train network, and station access by park-and-ride is highly prevalent in  
31 suburban and peripheral regions of Montreal, while in higher density areas station access by other  
32 modes, walking, bicycle and public transport are more common (Steer Davies Gleave, 2016).  
33 Walking distances of public transport trips that begin by park-and-ride are overestimated using  
34 this method.

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

## 1 REFERENCES

- 2 Agence Métropolitaine de Transport, 2013. Enquête origine-destination 2013, Montreal, Quebec.
- 3 Alshalalfah, B., Shalaby, A., 2007. Case study: Relationship of walk access distance to transit with  
4 service, travel, and personal characteristics. *Journal of Urban Planning and Development*  
5 133, 114-118.
- 6 Besser, L., Dannenberg, A., 2005. Walking to public transit: Steps to help meet physical activity  
7 recommendations. *American Journal of Preventive Medicine* 29, 273-280.
- 8 Booth, M., Bauman, A., Owen, N., Gore, C., 1997. Physical activity preferences, preferred sources  
9 of assistance, and perceived barriers to increased activity among physically inactive  
10 Australians. *Preventive Medicine* 26, 131-137.
- 11 Borodulin, K., Sipilä, N., Rahkonen, O., Leino-Arjas, P., Kestilä, L., Jousilahti, P., Prättälä, R.,  
12 2016. Socio-demographic and behavioral variation in barriers to leisure-time physical  
13 activity. *Scandinavian Journal of Public Health* 44, 62-69.
- 14 Brown, B., Werner, C., 2007. A new rail stop: Tracking moderate physical activity bouts and  
15 ridership. *American Journal of Preventive Medicine* 33, 306-309.
- 16 Canadian Society for Exercise Physiology, 2018. Canadian 24-hour guidelines: An integration of  
17 physical activity, sedentary behaviour, and sleep.
- 18 Chapleau, R., Trépanier, M., Lavigne, P., Allard, B., 1997. Origin-destination survey data  
19 dissemination in a metropolitan context: A multimedia experience. *Transportation*  
20 *Research Record: Journal of the Transportation Research Board*, 26-36.
- 21 Daniels, R., Mulley, C., 2013. Explaining walking distance to public transport: The dominance of  
22 public transport supply. *Journal of Transport and Land Use* 6, 5-20.
- 23 Dora, C., 1999. A different route to health: Implications of transport policies. *BMJ* 318, 1686-  
24 1689.
- 25 El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., Surprenant-Legault, J., 2014. New  
26 evidence on walking distances to transit stops: Identifying redundancies and gaps using  
27 variable service areas. *Transportation* 41, 193-210.
- 28 Farber, S., Morang, M., Widener, M., 2014. Temporal variability in transit-based accessibility to  
29 supermarkets. *Applied Geography* 53, 149-159.
- 30 Fielding, G., Glauthier, R., Lave, C., 1978. Performance indicators for transit management.  
31 *Transportation* 7, 365-379.
- 32 Godin, G., Desharnais, R., Valois, P., Lepage, L., Jobin, J., Bradet, R., 1994. Differences in  
33 perceived barriers to exercise between high and low intenders: Observations among  
34 different populations. *American Journal of Health Promotion* 8, 279-285.
- 35 Jiang, Y., Zegras, P., Mehndiratta, S., 2012. Walk the line: Station context, corridor type and bus  
36 rapid transit walk access in Jinan, China. *Journal of Transport Geography* 20, 1-14.
- 37 Karner, A., 2018. Assessing public transit service equity using route-level accessibility measures  
38 and public data. *Journal of Transport Geography* 67, 24-32.
- 39 Koplan, J., Dietz, W., 1999. Caloric imbalance and public health policy. *The Journal of the*  
40 *American Medical Association* 282, 1579-1581.
- 41 MacDonald, J., Stokes, R., Cohen, D., Kofner, A., Ridgeway, G., 2010. The effect of light rail  
42 transit on body mass index and physical activity. *American Journal of Preventive Medicine*  
43 39, 105-112.
- 44 Morency, C., Trépanier, M., Demers, M., 2011. Walking to transit: An unexpected source of  
45 physical activity. *Transport Policy* 18, 800-806.

1 Murray, A., Wu, X., 2003. Accessibility tradeoffs in public transit planning. *Journal of*  
2 *Geographical Systems* 5, 93-107.

3 O'Sullivan, S., Morrall, J., 1996. Walking distances to and from light-rail transit stations.  
4 *Transportation Research Record: Journal of the Transportation Research Board*, 19-26.

5 Owen, N., Bauman, A., 1992. The descriptive epidemiology of a sedentary lifestyle in adult  
6 Australians. *International Journal of Epidemiology* 21, 305-310.

7 Patterson, R., Webb, E., Millett, C., Lavery, A., 2018. Physical activity accrued as part of public  
8 transport use in England. *Journal of Public Health*.

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

13 She, Z., King, D., Jacobson, S., 2017. Analyzing the impact of public transit usage on obesity.  
14 *Preventive Medicine* 99, 264-268.

15 She, Z., King, D., Jacobson, S., 2019. Is promoting public transit an effective intervention for  
16 obesity? A longitudinal study of the relation between public transit usage and obesity.  
17 *Transportation Research Part A: Policy and Practice* 119, 162-169.

18 Siegel, P., Brackbill, R., Heath, G., 1995. The epidemiology of walking for exercise: implications  
19 for promoting activity among sedentary groups. *American Journal of Public Health* 85,  
20 706-710.

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

24 Steer Davies Gleave, 2016. Transport Collectif Train de l'Ouest de Montreal, Montreal:  
25 [https://www.cdpqinfra.com/sites/default/files/pdf/9\\_achalandage\\_mars\\_2016.pdf](https://www.cdpqinfra.com/sites/default/files/pdf/9_achalandage_mars_2016.pdf).

26 Trépanier, M., Chapleau, R., Morency, C., 2008. Tools and methods for transportation household  
27 survey. *URISA Journal* 20, 35-43.

28 van Soest, D., Tight, M., Rogers, C., 2019. Exploring the distances people walk to access public  
29 transport. *Transport Reviews*, 1-23.

30 Veillette, M., Deboosere, R., Wasfi, R., Ross, N., El-Geneidy, A., 2018. Small Steps, Big  
31 Differences: Assessing the Validity of using Home and Work Locations to Estimate  
32 Walking Distances to Transit. *Transportation research Record: Journal of the*  
33 *Transportation Research Record* In press.

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

37 Weinstein, A., Schlossberg, M., Irvin, K., 2008. How far, by which route and why? A spatial  
38 analysis of pedestrian preference. *Journal of Urban Design* 13, 81-98.

39 World Health Organization, 2010. Global recommendations on physical activity for health. World  
40 Health Organization, Switzerland.

41 World Health Organization, 2017. 10 facts on physical activity.

42 Yoshida, K., Allison, K., Osborn, R., 1988. Social factors influencing perceived barriers to  
43 physical exercise among women. *Canadian Journal of Public Health* 79, 104-108.

44 Zhao, F., Chow, L., Li, M., Ubaka, I., Gan, A., 2003. Forecasting transit walk accessibility:  
45 Regression model alternative to buffer method. *Transportation Research Record: Journal*  
46 *of the Transportation Research Board*, 34-41.