Small steps, big differences: Assessing the validity of using home and work locations to estimate walking distances to transit

Marie-Pier Veillette
School of Urban Planning, McGill University
E-mail: marie-pier.veillete@mail.mcgill.ca

Robbin Deboosere
School of Urban Planning, McGill University
E-mail: robbin.deboosere@mail.mcgill.ca

Rania Wasfi
Centre de Recherche du CHUM, Université de Montréal
Department of Geography, McGill University
E-mail: rania.wasfi@mail.mcgill.ca

Nancy Ross
Department of Geography, McGill University
E-mail: nancy.ross@mcgill.ca

Ahmed El-Geneidy
School of Urban Planning, McGill University
E-mail: ahmed.elgeneidy@mcgill.ca

July 2017

Word Count: 4764 words and 3 tables = 5114 words

For Citation please use: 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.* Paper to be presented at the 97th Annual Meeting of the Transportation Research Board, Washington D.C., USA.

ABSTRACT

Walking to and from public transport can form a seamless way to integrate physical activity into our daily lives, and thereby help us achieve the recommended minutes of physical activity. To measure the link between physical activity and public transport use, it is critical to determine how far individuals are walking, and are willing to walk, to different modes of transit. Many planners, however, do not have access to detailed information on the exact public transport lines used by individuals, and therefore need to estimate distances by making use of only home and work locations. This study therefore compares two methods of calculating walking distances: one method using widely available home and work locations and a fastest route algorithm leveraging GTFS data, and a second employing a detailed travel survey containing information on the real routes used by each respondent from Montreal, Canada, in order to generate more robust estimates of the distances individuals are walking to public transport stops. Results show that walking distances calculated from commonly available origin and destination information tend to underestimate real walking distances by 10% (1,816 compared to 2,021 meters of daily average walking). Multilevel mixed-effect regression models point out that these differing results can be mainly attributed to differences in travel behavior and mode choice: a fastest path algorithm between each origin and destination severely overestimates the number of transfers people are willing to make and underestimates the number of train users, and therefore results in biased estimates of daily walking distances. Findings from this study provide a better understanding of how modelled and real walking routes to public transport stops differ, which can be of interest to professionals and urban decision makers wishing to correctly model walking to transit in their region when only limited information is available.

Keywords: Public transport, walking, travel behavior, physical activity

1. INTRODUCTION

Obesity rates have skyrocketed in the last few decades, due to, among other reasons, an increasingly sedentary lifestyle, exacerbated by the widespread use of the private automobile as our main mode of transportation (*1-3*). In light of this epidemic, the World Health Organization stresses the importance of active transportation in determining physical health (*4*). Scholarly attention has particularly been paid to walking to public transport, and its ability to help individuals achieve the recommended minutes of weekly physical activity, and thereby contribute to a healthy lifestyle (*5-8*).

To measure the link between public transport and physical activity, it is critical to determine how far individuals are walking, and are willing to walk, to different modes of transit. While a large body of literature has examined this relationship by making use of detailed travel surveys, municipal planners are often lacking such information and only have access to home and work locations for individuals using public transport. Such a lack of detailed information might lead to errors in the calculation of walking distances, resulting in skewed and possibly incorrect findings, which then feed into public policy and municipal decision making processes.

The goal of this study is, therefore, to assess the levels of walking to public transport for different modes, and to compare how these levels differ between two methods of calculation. The first method uses information from a detailed travel survey which contains the exact route the commuter used, while the second method only uses origin and destination information, representing a common scenario in organizations where information is scarce. For this purpose, we employ the detailed 2013 origin-destination (OD) survey carried out by the Agence Métropolitaine de Transport (AMT) in Montreal, Canada (*9*), which was obtained under a special agreement with the agency and contains detailed information on each person's route, such as the different bus, train, or subway lines that were used in a trip, in addition to the number of transfers and the metro or train stations where each individual first boarded, transferred and left the system. Walking distances are subsequently derived from this detailed OD survey and compared to walking distances computed by using only origin and destination information and a shortest route algorithm based on data from the General Transit Feed Specification (GTFS). By making use of GTFS and a fastest route algorithm in ArcGIS for these calculations, we employ information that is abundantly available to planners and urban decision makers in a variety of settings. By measuring how the two estimates for walking distances differ, we can construct adjustment factors that practitioners can use to correct walking distances calculated from only origins and destinations. The finding from this study are expected to be of relevance to professionals and decision makers wishing to measure the health benefits resulting from public transport trips in their region when only limited information is available.

The rest of the paper is organized as follows. Section 2 presents literature discussing walking to transit, how walking distances are measured, and their relation to public health. In section 3, the data and two methods used to derive walking distances are discussed, while section 4 presents a multilevel mixed-effect regression model to disentangle the determinants of walking to transit.

Section 5 then concludes the paper and provides recommendations to policy makers and professionals.

2. WALKING TO TRANSIT

Only 15% of Canadian adults achieve the recommended levels of physical activity according to the 2007 to 2009 Canadian Health Measures Survey (*10*). While in the past the threshold for recommended physical activity was set to a minimum of 30 minutes per day, the rule of thumb was subsequently refined to 150 minutes per week of "moderate-to-vigorous physical activity" (*4; 11*). In this context, research suggests that walking to public transport, as an activity carried out daily by individuals, can form a seamless way to integrate physical activity into their daily routine. It is therefore not surprising that previous research has extensively documented positive health benefits associated with the use of public transport; transit users who walk to or from stations experience increased levels of physical activity and can thereby achieve the recommended levels of daily and weekly physical activity when using certain modes (*1; 5-8; 11-17*).

Walking distance to public transport varies worldwide across cities, according to trip purpose, the built-environment, household and individual trip characteristics, and factors related to the specific mode used. The following section briefly reviews recent findings on these factors, and present different methods used to measure walking distances.

Early research on the catchment area of public transport stops pinpoints the distance individuals are willing to walk to transit at 400 meters for bus stops, and 800 meters for heavy and light rail systems, indicating that the specific mode taken by individuals is a key determinant of walking distance (*18-20*). These catchment size areas have subsequently been refined to 524 meters for bus stops and 1,259 meters for commuter rail in a study of home-based trips in Montreal (*19*). Another Montreal study again demonstrates that commuter rail users walk longer distances to transit than subway and city bus users (*8*). In a non-western context, Wang, Chen and Xu (*21)* derive that 75 percent of survey respondents walk less than 623 meters to access a subway station in Beijing. Moreover, besides mode factors, increases in transit service frequency and the number of transfers made during each trip have been demonstrated to reduce the total distance people are walking to transit (*8; 19; 22*).

In addition to the factors related to the mode and route used by individuals, socio-economic characteristics have also been found to influence walking distance to transit (*19; 21; 23*). Men walk longer to reach transit stops than women (*8; 23; 24*), while the probability of reaching a transit stop using an active mode of transportation (such as walking) decreases considerably as people age (*8; 18; 23*). Furthermore, individuals living in low-income households are more likely to spend 30 minutes or more walking to transit on an average day than those living in medium and higher income households (*11*), and university graduates accumulate less total walking time than individuals with lower levels of education (*11*).

Neighbourhood characteristics, such as population density and the number of stops in close proximity to the home also affect walking distance to transit stops and stations (*21; 25*). In

California, the probability of reaching a transit stop with an active mode of transport decreases by 12% for every additional mile between the origin and the closest transit stop (*18*), while individuals living in more densely populated areas have an average daily walking time that is longer than those living in areas with a lower population density (*11*). In a similar vein, a study in China finds that walking distances to BRT stations increase by 75 meters for every additional kilometer away from the city center (*22*).

The results uncovered by the research presented above arguably depend on the type of data that was used to carry out the study. Most studies make use of travel surveys that report origins and destinations, and employ spatial analysis techniques in a geographic information system to calculate walking distances (*8; 11; 18; 19; 21-23*). Municipal planners and researchers, however, are often lacking detailed information on the actual transit lines used by respondents, and are therefore required to resort to different methods, such as shortest path algorithms between known origins and destinations, to estimate real walking distances. This study therefore compares walking distances generated by a detailed travel survey with estimates of walking distances that were derived from only origin and destination information.

3. METHODOLOGY

Our method consists of a two-step spatial and statistical analysis procedure. First, we calculated the levels of walking to five public transport modes (commuter train, subway, city bus, suburban bus, and peripheral bus) using ArcGIS for two different datasets: (1) the 2013 detailed OD Survey with full information on every respondent's route, and (2) only origin and destination information supplemented with GTFS data. The detailed origin-destination survey is conducted every 5 years and samples 5% of all households living in the Montreal Metropolitan Area, Canada (*9*). In addition to the origin and destination locations, it contains exclusive information on the respondents' socio-economic and transit trip characteristics, such as the number of different transit modes used, the specific bus, commuter train, or subway lines taken, and the number of transfers made during one trip. To compare walking distances generated from this detailed dataset with walking distances estimated from commonly available information, we artificially modelled a scenario where planners only have access to origin and destination information: we extracted only the origins and destinations from the detailed OD Survey, and supplemented these with GTFS data to represent the public transport network and schedule in Montreal. We will refer to the first scenario (with detailed information) as 'detailed OD', while the second scenario will be referred to as the 'GTFS scenario', 'GTFS calculations', or simply 'GTFS'. In a second step, the potential differences in the levels of walking to transit between the two datasets and methods were compared by conducting three multilevel mixed-effect regressions.

3.1 Sample

Our sample consists of 9,588 home-based trips realized by adults (over 18 years old) who are commuting to work and living in the Montreal Metropolitan Area. Starting from the full OD Survey (N=410,741 trips), we excluded all trips made with modes other than subway, commuter rail, and bus (city, peripheral and suburban busses). We also omitted trips associated with unreported and unknown route numbers and for purposes other than work. In addition, in order to retain only one trip per individual, we removed multiple trips made by the same respondents. Our sample represents 13.8% of all reported trips made for work purposes in the 2013 OD Survey.

3.2 Spatial analysis

The two methodologies – detailed OD and GTFS – allow us to measure the daily walking distance to transit, the number of transfers made by each individual, the total distance traveled in-vehicle and the frequency of public transport service at the first stop or station. As the information provided in both scenarios was different, two methods of calculation were employed in ArcGIS.

Detailed OD Survey

To calculate the total daily walking distance with the detailed OD Survey, we used the Network Analyst 'Closest Facility' tool. We first generated a ready-to-use pedestrian street network dataset and a file containing all stops served by the subway, train, city bus, peripheral bus and suburban bus lines. When several bus stops were used by more than one transit line, we duplicated the number of stops according to the number of routes serving it. As a result, each bus stop was assigned to one transit line. Pedestrian routes were then generated from each home and work location to their respective nearest 200 transit stops. Among the routes from home to stop generated with Network Analyst, we then identified the shortest pedestrian route linking the home location to the closest stop if that stop matched the route number reported in the detailed OD Survey. A similar operation was done to identify the shortest pedestrian route between the work locations and the closest stop serving the last line a respondent used. Note that, while for subway and train the exact first and last stops were reported, the method as explained above was also used for these two modes. The closest stops were compared with the real stops and no errors were noted. Both beginning and end-of-trip walking distances were subsequently summed to generate the total walking distance per trip, which was then multiplied by two to account for daily walking to and from public transport, assuming that the person commuting by public transport will be doing the same trip but in reverse order when going home.

To measure the distance travelled in-vehicle, we calculated the distance between the first and last transit stop, previously identified, with the Network Analyst 'New Route' tool and a transit network dataset. The number of transfers made during one trip was provided in the detailed OD survey and did not require any GIS computations.

Common scenario using GTFS

To calculate walking distances in the common scenario where detailed information is scarce, we first derived the fastest route between an individual's origin and destination using transit through the ArcGIS tool 'Add GTFS to a network dataset'. The fastest route algorithm using GTFS incorporates walking, waiting, boarding and alighting, and in-vehicle time (based on the transit schedule) to compute total travel time. This resulted in a single line for every person, representing their fastest route at 8 AM on a regular Tuesday in 2013. To compute each person's walking distance, his/her route was intersected with the street network, after which the walking parts were extracted. The walking segments closest to the origin or destination were then said to represent the distance walked from home to the first stop taken and from the last stop taken to the destination, respectively. These two distances were summed to represent the total walking during a single trip and then multiplied by two to represent the total walking to and from public transport in a day.

The number of transfers in this second scenario was derived by adding the number of walking segments (minus the home-stop and stop-destination segments), combined with the number of different subway or rail lines the person took (computed by intersecting the rail and subway network with the route taken, coded by line). As the first line a person took was not computed in the calculation with GTFS, the frequency of the first route taken was derived as follows: for every stop, the number of different arrivals between 7 and 9 AM (whether a subway, a train, or a bus) was divided by the time interval of 2 hours, resulting in the average time between two subsequent transit vehicles arriving at a stop during the morning peak. To ensure consistency, this method was also applied to calculate the frequency of the first stop an individual actually used as derived from the detailed OD survey.

Neighbourhood variables

Several neighbourhood variables were added to the dataset to supplement the trip and individual characteristics. Firstly, the percentage of people with a university degree in the origin census tract, and population density in this same area, was gathered from Statistics Canada. Secondly, the number of transit stops within 1,000 m of the origin was calculated using ArcGIS tools, this to reflect the density of the public transport network in close proximity to each respondent's home. In addition, street connectivity was measured through the number of 4-way street intersections within a 1,000-meter buffer from the home, although this variable was not used as it was highly correlated with the number of bus stops within 1,000 meters.

The data generated from the detailed OD survey and the GTFS calculations were then appended to generate a bigger dataset (N=9,588*2), and a dummy variable was included that equals one if the GTFS method was used and zero if the detailed OD method was used in the calculations. When appending the two datasets, the individual identification number was retained in order to be used in the multilevel analysis presented in the next section. The coefficient for the GTFS dummy variable will therefore show to what extent the total daily walking distance (as the dependent variable in the model presented below) is under- or overestimated by GTFS, while the coefficients of the other variables used in the model will offer explanations on the potential sources of the differences between the two methods.

3.3 Multilevel mixed effect regression

To disentangle the determinants of walking to transit, and to discern how the two methodologies of calculating walking distances differ, we conducted three multilevel mixed-effect regressions at the individual and census tract levels (CTs). CTs are geographical areas that share similar characteristics, house between 2,500 and 8,000 people, and are used as a proxy for neighbourhoods (*26; 27*). A multilevel mixed-effect regression has the ability to capture both neighbourhood and individual effects on the variation of the total walking distance (*28*). As we can expect individuals living in the same neighbourhood to share some unobserved similarities determining their walking distance, a multilevel model is thus more suitable than a regular ordinary least squares regression. Our two-level data structure allows nesting individuals within their neighbourhood, while also accounting for the two methods of calculation (detailed OD and GTFS) at the trip level (*29*).

4. ANALYSIS

Table 1 shows descriptive statistics of the variables used in the three regression models. The average total daily walking distance, accumulated solely by accessing and egressing transit, is 2,021 meters based on detailed OD survey data, whereas the shortest path algorithm using GTFS data estimates this number as 1,816 meters. The calculations through GTFS data thus underestimate the total daily walking distance by 205 meters (or 10%) compared to the real distances walked by the OD survey respondents. A t-test reveals that this difference in average daily walking distances is statistically significant. Other considerable differences between OD and GTFS calculations are also present: substantial dissimilarities in mode choice and in the number of transfers exist. More than twice as many people reported to have used a commuter train in the detailed OD survey compared to the modal split calculated through a shortest path algorithm using GFTS data. In addition, the GTFS calculations underestimate the modal split of the subway by 5%. In a similar vein, the number of transfers varies considerably between both methods of calculation; approximately three times more individuals realized three transfers or more according to the GTFS calculations. Note that the GTFS calculations assumed everyone left their home at 8 AM, which artificially increases travel time for individuals taking e.g. the train at 8.30 AM. This might cause a different route, for example using a bus, to be faster, partly explaining the large difference in train modal share. While individuals would most likely stay at home for some time instead of waiting at the station, the GTFS application in ArcGIS did not incorporate these behavioural calculations. Similarly, an individual that would normally catch a train at 7.59 AM would be rerouted to a different stop, thereby resulting in a different walking distance. Ideally, the GTFS routes should be calculated every minute between 7 and 9 AM, and the shortest route should be used, minimizing any potential errors related to transit service variability. Nevertheless, the differences between the two methodologies are considerable.

Table 2 breaks down daily walking distances by their mean, median, 75th percentile and 85th percentile for both methods of calculation, and presents the percent of respondents achieving the recommended minutes of physical activity. For this purpose, a walking speed of 5.47 km/h was used, cf. Wasfi, Ross and El-Geneidy (*8)*. According to the detailed OD Survey, around 71% of train users, 14% of subway riders and 10% of city bus users meet the recommended levels of physical activity solely by walking to transit. The calculations made with GTFS data severely underestimate the percentage of people achieving the recommended minutes of physical activity, by nearly 40% for train users and by almost 13% for subway users. Interestingly, both methods of calculation arrive at similar results for the percentage of city bus users meeting the recommended levels of physical activity. Note that the distances presented in the table represent total walking distance each day. The 85th percentile of walking distances from home to stop are 513m for city bus stops, and 1,020m for subway stations, which is consistent with findings from previous studies (*19*).

TABLE 1 Summary statistics of total daily walking distance to and from transit stops or stations in the Montreal Metropolitan Area, Canada using two methods of calculation

	Detailed OD		GTFS	
Continuous variables	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
Total walking distance per day (m)	2,020.84	2,013.43	1,816.49	1,229.22
Trip characteristics				
Distance travelled in vehicle (km)	10.05	70.09	11.13	7.59
Frequency of service at first stop (min)	16.61	33.97	14.53	34.47
Neighbourhood characteristics				
Percentage of people with university degree within origin census tracts	27.90	13.66	27.90	13.66
Population density within origin census tract (population/km^2)	7,473.37	5,091.67	7,473.18	5,091.44
Number of bus stops within a 1000m buffer of origin	79.66	32.41	79.66	32.40
Socio-economic characteristics				
Age	43.15	12.24	43.15	12.25
Dummy variables	**Proportion**		**Proportion**	
Socio-economic characteristics				
Gender - Male	0.45		0.45	
Low income (<$30k)	0.11		0.11	
Medium income ($30k-89k)	0.46		0.46	
Trip characteristics				
First mode taken				
Subway	0.30		0.25	
Commuter Train	0.07		0.03	
Peripheral Bus	0.05		0.08	
Suburban Bus	0.14		0.15	
City Bus	0.44		0.49	
Number of transfers				
No transfer	0.40		0.21	
One transfer	0.37		0.34	
Two transfers	0.19		0.30	
Three transfers or more	0.04		0.15	
Methodology				
GTFS	0 .00		1.00	

TABLE 2 Total daily walking distance by mode and method of calculation

Total daily walking distance (m)	Train	Subway	City Bus	Suburban Bus	Peripheral Bus
Detailed OD					
Mean	4,675	2,253	1,523	1,692	2,522
Median	3,822	1,986	1,197	1,346	2,025
75th percentile	5,477	2,706	1,832	2,068	3,172
85th percentile	6,610	3,163	2,316	2,606	4,195
Standard deviation	4,215	1,486	1,551	1,622	1,901
Percent achieving recommended minutes of physical activity	70.58	24.08	9.88	13.39	32.57
GTFS					
Mean	2,369	1,761	1,588	1,941	2,922
Median	2,088	1,664	1,456	1,730	2,376
75th percentile	2,939	2,205	2,027	2,451	3,817
85th percentile	3,310	2,587	2,397	2,970	4,547
Standard deviation	2,143	858	900	1,235	2,324
Percent achieving recommended minutes of physical activity	30.71	11.48	9.14	18.32	41.22

To determine the factors that drive the differences noted above, the results of the three regression models are presented (Table 3). The first model does not account for the mode taken and the number of transfers made by transit users. This model confirms the summary statistics shown above: the total daily walking distance is underestimated by 199 meters by the GTFS calculations, holding all other variables in the model constant.

When the first mode taken by a respondent is operationalized in the model, the coefficient for the dummy variable, representing a calculation based on GTFS data, decreases by 122 meters to 77 meters, see Model 2. Model 2 therefore suggests that the difference in the two walking distance calculations can be largely attributed to differing mode choice between the two modelling approaches. As OD Survey respondents are more likely to use the train than would be expected from the shortest route calculations with GTFS, they tend to walk more than predicted. Note that, as mentioned above, the difference in modal choice could by partly explained by the fixed start time (8 AM) for the GTFS calculations. Ideally, the GTFS routes should be calculated every minute during the morning peak and the start time resulting in the shortest route should be used. Nevertheless, the coefficients in Model 2 demonstrate that differing mode choice between GTFS and the detailed OD survey explains a large part in the variation between the two walking distance estimates. Planners could thus benefit from correctly modelling modal choice. This could be done, for example, by placing a priority on certain modes such as commuter trains, to take into account comfort and other factors not incorporated in our model. In practical terms, the GTFS calculations could artificially reduce the travel time for priority modes. More research is however needed to determine these priority modes and their weighting.

In Model 2, the coefficient for individuals living in low-income households is no longer statistically significant, which suggests that the shorter walking distances of low income groups in Model 1 could be explained by different mode choices: low income groups are more likely to use a bus than a subway or commuter rail and therefore walk less, which is consistent with previous research (*8*).

The third model, in addition to controlling for the first mode used by respondents, also controls for the number of transfers made during each trip. This model reveals that there are no longer statistically significant differences between both methods of calculation (the coefficient for the GTFS dummy is no longer significant). By comparing model 2 and 3, we can infer that the differing number of transfers accounts for around 77 meters of the difference in total walking distance between the two approaches. A shortest route algorithm based on GTFS data, in order to minimize total travel time, routes an individual to a stop close to home, even if that route does not directly lead to the destination, thus resulting in more transfers and shorter walking distances. However, as most people are averse to transferring between routes (*30*), they are happy to walk further to access a direct route, thereby partly explaining the differences between the rational GTFS approach and real-life (irrational) travel behavior. This shows that, in order to estimate robust walking distances to transit, planners could benefit from accurately modelling travel behaviour, in particular by limiting the number of transfers individuals are willing to make, and by correctly modelling modal choice. Making these two adjustments will allow for the calculation of accurate walking distances, even if only origin and destination information is available.

The other coefficients in model 3 present interesting findings regarding how far transit users are willing to walk according to the first transit mode they used. All else equal, trips made by a train are related to an increase in total daily walking distance of 1,750 meters compared to a city bus, while OD Survey respondents walking to subway stops walk, on average, 892 meters more per day than individuals using a city bus. Furthermore, a trip integrating two transfers is related to a reduction in the total walking distance by 485 meters, all else equal, and a trip with 3 transfers or more is associated with a decrease in the total walking distance of 656 meters compared to a trip with no transfer, ceteris paribus.

Finally, the total daily walking distance is also influenced by socio-economic, trip and neighbourhood characteristics. As expected, the coefficients for age, the number of university graduates in the origin census tract, population density, and the number of bus stops within 1,000 meters are statistically significant. Every year increase in age is related to a decrease in walking distance of 2.38 meters. For every percent increase in the number of university graduates, total walking distance is predicted to decrease by 11.03 meters. Furthermore, the total walking distance reduces by 0.05 meters for each percent increase in population density, while one additional bus stop within 1,000 meters of the origin is associated with a decrease in total walking distance of 8.08 meters. The intra-class coefficients for model 3 show that the origin census tract and the individual are explaining 41.4% and 46.2% of the variance in total walking distances, respectively.

TABLE 3 Multilevel regression models predicting total daily walking distance

	Model 1				Model 2				Model 3			
Variables	**Coef.**	**Sig.**	**Confidence int.†**		**Coef.**	**Sig.**	**Confidence int.†**		**Coef.**	**Sig.**	**Confidence int.†**	
Socio-economic characteristics												
Male	33.00		-11.73	77.72	20.77		-22.73	64.28	15.34		-28.36	59.03
Age	-2.65	***	-4.49	-0.81	-2.50	***	-4.29	-0.71	-2.48	**	-4.28	-0.68
Low income	-150.28	***	-229.39	-71.18	-75.41		-152.44	1.62	-52.77		-130.22	24.67
Medium income	-11.88		-60.70	36.94	12.26		-35.24	59.75	21.87		-25.84	69.58
Trip characteristics												
Distance travelled in vehicle	-0.001		-0.006	0.003	-0.01	***	-0.01	0.00	0.01	***	0.01	0.02
Frequency of service at first stop	4.08	***	3.16	5.00	0.72		-0.34	1.77	0.55		-0.50	1.60
First mode taken												
Subway					983.09	***	913.07	1053.12	891.58	***	820.90	962.27
Commuter Train					1935.14	***	1782.80	2087.47	1749.78	***	1596.60	1902.96
Peripheral Bus					-51.79		-245.46	141.88	-67.62		-258.93	123.70
Suburban Bus					-35.36		-192.97	122.25	-15.70		-170.67	139.27
Number of transfers												
One transfer									-319.55	***	-374.30	-264.81
Two transfers									-485.20	***	-551.79	-418.62
Three transfers or more									-656.21	***	-751.18	-561.25
Neighbourhood characteristics												
Percentage of people with university degree	-3.12		-9.81	3.57	-12.14	***	-18.88	-5.40	-11.03	***	-17.49	-4.57
Population density	-0.04	***	-0.06	-0.02	-0.06	***	-0.08	-0.04	-0.05	***	-0.07	-0.03
Number of bus stops	-5.68	***	-7.85	-3.50	-8.60	***	-10.75	-6.44	-8.08	***	-10.22	-5.95
Methodology												
GTFS	-198.68	***	-240.08	-157.28	-76.70	***	-116.49	-36.91	9.11		-31.66	49.88
Constant	3027.39	***	2780.20	3274.58	3376.51	***	3116.03	3636.99	3275.21	***	3019.73	3530.69
Random-effects parameters	**Est.**	**Sd. Error**	**Confidence int.†**		**Est.**	**Sd. Error**	**Confidence int.†**		**Est.**	**Sd. Error**	**Confidence int.†**	
Census Tract												
var(Constant)	1530977	121135	1311050	1787797	1539675	11985	1321805	1793457	1393737	112403	1189961	1632409
Trip ID												
var(Constant)	80707	2342	45631	142747	128403	21729	92157.97	178902.3	164204	21930	126388	213334
var(Residual)	2077314	30389	2018598	2137737	1859207	27225	1806606	1913339	1808063	26579	1756713	1860914

Dependent variable: Total daily walking distance (m)

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

† 95% confidence interval

5. CONCLUSION

This paper contributes to the literature examining walking to transit by employing a highly detailed OD survey containing the stops where each person first entered, transferred, and subsequently left the system. We compare robust estimates generated from this detailed survey to walking distances computed by making use of abundantly available origin and destination information and a fastest route algorithm based on GTFS data. We find that methods only making use of origin and destination information tend to underestimate the distance (and time) people are walking, or willing to walk, to transit. Our results show that walking distances calculated through GTFS data underestimate walking distances by 10% compared to real-life behavior (1,816 compared to 2,021 meters of daily average walking). Multilevel mixed-effect regression models point out that these differing results can be mainly attributed to differences in travel behavior and mode choice. Applying a shortest path algorithm to a transit network, while resulting in the fastest route, does therefore not accurately model travel behavior: while the routes calculated through GTFS data show over 15% of individuals taking three or more transfers, in reality only 3% of respondents transfer three times or more. We also find that more than two times more respondents commute by train than is expected from calculations based on a fastest route with GTFS data, resulting in considerable differences in estimated walking distances. In order to generate robust estimates of walking distances to transit, practitioners can therefore benefit from correctly modelling modal choice and limiting the number of transfers an individual can make.

The difference in walking distance estimates can impose a bias on the number of individuals assumed to not meet the recommended minutes of daily physical activity through public transport. Therefore, if practitioners solely rely on calculations from home and work locations, wrong interventions and policies might be proposed: practitioners might assume from their calculations that a large percentage of residents in a certain area are not meeting the recommended minutes of daily physical activity, while this might not be the case. This study therefore provides practitioners and researchers with a clear correction method when using GTFS data to calculate walking distances to public transport. Further research could improve on this study by using smartphone apps or wearable devices, which can provide even more detailed information about walking paths and walking speeds, although we do not expect large differences from our study in terms of total walking distances, except when an individual uses a short-cut through a private area or follows an irrational path to reach a public transport stop.

REFERENCES

[1] Jakicic, J., and K. Gallagher. Exercise considerations for the sedentary, overweight adult. *Exercise and Sport Sciences Reviews,* Vol. 31, No. 2, 2003, pp. 91-95.

[2] Frank, L., M. Andresen, and T. Schmid. Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine,* Vol. 27, No. 2, 2004, pp. 87-96.

[3] Katzmarzyk, P. Physical activity levels of Canadian children and youth: Current issues and recommendations. *Canadian Journal of Diabetes,* Vol. 28, No. 1, 2004, pp. 67-78.

[4] WHO. *Global recommendations on physical activity for health.* World Health Organization, Geneva. http://www.who.int/dietphysicalactivity/physical-activity-recommendations-18-64years.pdf?ua=1.

[5] Sallis, J., L. Frank, B. Saelens, M. Kraft, and E. Engelke. Active transportation and physical activity: Opportunities for collaboration on transportation and public health research. *Transportation Research Part A,* Vol. 38, No. 4, 2004, pp. 249–268.

[6] Brown, B., and C. Werner. A new rail stop: Tracking moderate physical activity bouts and ridership. *American Journal of Preventive Medicine,* Vol. 33, No. 4, 2007, pp. 306-309.

[7] MacDonald, J., R. Stokes, D. Cohen, A. Kofner, and G. Ridgeway. The effect of light rail transit on body mass index and physical activity. *American Journal of Preventive Medicine,* Vol. 39, No. 2, 2010, pp. 105-112.

[8] Wasfi, R., N. Ross, and A. El-Geneidy. Achieving recommended daily physical activity levels through commuting by public transportation: Unpacking individual and contextual influences. *Health and Place,* Vol. 23, 2013, pp. 18-25.

[9] Agence Métropolitaine de Transport. Enquête origine-destination 2013.In, Montréal, QC, 2013.

[10] Statistics Canada. Physical activity levels of Canadian adults, 2007 to 2009. 2015.

[11] Besser, L., and A. Dannenberg. Walking to public transit: Steps to help meet physical activity recommendations. *American Journal of Preventive Medecine,* Vol. 29, No. 4, 2005, pp. 273-280.

[12] Sallis, J., C. Spoon, N. Cavill, J. Engelberg, K. Gebel, M. Parker, C. Thornton, D. Lou, A. Wilson, C. Cutter, and D. Ding. Co-benefits of designing communities for active living: an exploration of literature. *International Journal of Behaviorial Nutritional and Physical Activity*, 2015, pp. 12-30.

[13] Brown, B., and C. Werner. Before and After a New Light Rail Stop: Resident Attitudes, Travel Behavior, and Obesity. *Journal of the American Planning Association,* Vol. 75, No. 1, 2008, pp. 5-12.

[14] Warburton, D., C. Nicol, and S. Bredin. Health benefits of physical activity: the evidence *Canadian Medical Association Journal,* Vol. 174, No. 6, 2006, pp. 801-809.

[15] Sener, I. N., J. R. Lee, and Z. Elgart. Potential health implications and health cost reductions of transit-induced physical activity. *Journal of Transport and Health,* Vol. 3, 2016, pp. 133-140.

[16] Bopp, M., V. V. Gayah, and M. E. Campbell. Examining the Link Between Public Transit Use and Active Commuting. *International Journal of Environnemental Research and Public Health,* Vol. 12, 2015, pp. 4256-4274.

[17] Lachapelle, U., and D. Gianini Pinto. Longer or more frequent walks: Examining the relationship between transit use and active transportation in Canada. *Journal of Transport and Health,* Vol. 3, 2016, pp. 173-180.

[18] Durand, P. C., X. Tang, P. K. Gabriel, I. N. Sener, A. O. Oluyomi, G. Knell, K. A. Porter, M. D. Hoelscher, and W. H. Kohl III. The association of trip distance with walking to reach

public transit: Daya from the California Household Travel Survey. *Journal ofTransport & Health,* Vol. 3, 2016, pp. 154-160.

[19] El-Geneidy, A., M. Grimsrud, R. Wasfi, P. Tétreault, and J. Surprenant-Legault. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation* Vol. 41, No. 1, 2014, pp. 193-210.

[20] Zhao, F., L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan. Forecasting transit walk accessibility: Regression model alternative to buffer. *Transportation Research Record,* Vol. 1835, 2003, pp. 34-41.

[21] Wang, Z., F. Chen, and T. Xu. Interchange between Metro and Other Modes: Access Distance and Catchment Area. *Journal of Urban Developement,* Vol. 142, No. 4, 2016.

[22] Jiang, Y. P., C. Zegras, and S. Mehndiratta. Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China. *Journal of Transport Geography,* Vol. 20, 2012, pp. 1-14.

[23] Morency, C., M. Trépaniera , and M. Demers. Walking to transit: An unexpected source of physical activity. *Transport Policy,* Vol. 18, No. 6, 2011, pp. 800-806.

[24] García-Palomares, J. C., J. Gutiérrez, and O. D. Cardozo. Walking accessibility to public transport: an analysis based on microdata and GIS. *Environment and Planning B: Planning and Design 2013,,* Vol. 40, 2013, pp. 1087-1102.

[25] Ewing, R., and R. Cervero. Travel and the built environment. *Journal of the American Planning Association,* Vol. 76, No. 3, 2010, pp. 265-294.

[26] Statistics Canada. Census tract: Detailed definition.In, 2015.

[27] Ross, N. A., S. Tremblay, and K. Graham. Neighbourhood influences on health in Montréal, Canada *Social Science & Medicine,* Vol. 59, No. 7, 2004, pp. 1485-1494

[28] Fuzhong, L., J. K. Fisher, R. C. Brownson, and M. Bosworth. Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiol Community Health,* Vol. 59, 2005, pp. 558-564.

[29] Sallis, J., B. Saelens, L. Frank, T. Conway, D. Slymen, K. Cain, J. Chapman, and J. Kerr. Neighborhood built environment and income: Examining multiple health outcomes. *Social Science & Medicine,* Vol. 68, 2009, pp. 1285-1293.

[30] Guo, Z., and N. Wilson. Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transportation Research Part A,* Vol. 45, 2011, pp. 91 - 104.