Catalogue no. 18-001-X
ISBN 978-0-660-72520-8

Reports on Special Business Projects

Active and Public Transportation Spatial Accessibility Measures: Methodology and Key Results

by Nick Newstead, Kaitlyn Hobbs, Cal Giunta, and Sheldon Birkett

Release date: November 20, 2024

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of contents

Table of tables

Table of figures

Acknowledgments

We would like to thank Mobilizing Justice for their help during this consultation process as well as the ongoing work they do to help alleviate transportation barriers across Canada.

Active and Public Transportation Spatial Accessibility Measures: Methodology and Key Results

by **Nick Newstead, Kaitlyn Hobbs, Cal Giunta,** and **Sheldon Birkett**

1 Executive Summary

This report provides methodological details and results of the spatial access measures produced by the Data Exploration and Integration Lab, Centre for Special Business Projects in partnership with Housing, Infrastructure and Communities Canada (formerly, Infrastructure Canada). A total of 28 measures were produced for seven destinations and three modes of transportation. Destinations include employment, healthcare facilities, public K-12 education facilities, public post-secondary facilities, grocery stores, sports and recreation facilities, and cultural and art facilities. For each destination, there are four transportation varieties: access via public transit during peak hours, access via public transit during off-peak hours, access via cycling and access via walking. This resulted in four databases, one for each transportation mode variant, comprised of 16 variables and recorded at the dissemination block level for all of Canada.

The main highlights on the methodology can be summarized as follows.

- Data was collected from a variety of open data sources as well as Statistics Canada data holdings.
- Representative points of dissemination blocks were determined using the density of buildings.
- The routing engines r5r and Valhalla were used to determine travel durations between dissemination block representative points via public transportation and active transportation, respectively.
- GTFS data was compiled by reviewing all civic and regional open data portals across all census subdivisions (CSDs). Around 72.4 percent of the Canadian population lives in a dissemination block with at least one transit stop, using the collected GTFS data.
- A total of 54 transit regions were generated for analysis of access in areas with openly available GTFS data. Each transit region is a compilation of GTFS sources that are grouped together based on the assumption that intersecting bounding boxes of the sources' transit stop locations meant commuters could connect between different transit providers.
- A gravity model, akin to the one used in the Proximity Measures Database, was used to quantify accessibility to all amenities except for grocery stores which utilized a dual measure.

The main highlights on the results for access by public transit during peak and off-peak hours are:

- Around 63% and 62% of Canadians have access to a grocery store during peak and off-peak hours, respectively.
- 76% of Canadians have access to places of employment.
- 65% of Canadians have access to cultural and art facilities (CAFs).
- 75% of Canadians have access to healthcare facilities (HFs). 69% of Canadians have access to educational facilities (EFs).
- 68% of Canadians have access to sports and recreational facilities (SRFs) by transit during peak and 69% during off-peak hours.

The main highlights on the results for access by cycling are:

- Approximately 80% of Canadians have access to a grocery store within 15 minutes.
- Approximately 96% of Canadians have access to places of employment.
- Approximately 85% of Canadians have access to CAFs.
- Approximately 93% of Canadians have access to HFs.
- Approximately 90% of Canadians have access to EFs.
- Approximately 89% of Canadians have access to SRFs.

The main highlights on the results for access by walking are:

- Around 71% of Canadians have access to a grocery store within 15 minutes.
- Around 95% of Canadians have access to places of employment.
- Around 90% of Canadians have access to HFs.
- Around 60% of Canadians have access to CAFs.
- Around 83% of Canadians have access to EFs.
- Around 80% of Canadians have access to SRFs.

2 Introduction

In 2021, the federal government announced $5.9 billion towards advancing public transit and active transportation for the next 5 years, to be followed by permanent annual envelop of $3 billion beginning 2026-27.[1] To better understand potential access to services and amenities across Canada using public and active transportation, this report details measures of access to various services and amenities by transit at the Census Dissemination Block (DB) geographic boundary level, via bus, train, subway, sea buses, light-rail trains, and streetcars (trams) services operating with a fixed schedule. These measures can support the creation of more sustainable and resilient public transportation infrastructure across Canada. For example, the proposed set of measures can enable policymakers to assess spatial patterns and socioeconomic disparities of potential access. Moreover, periodic updates of these measures will provide insights on whether access to a service or amenity has improved or worsened over a given timescale.

Previous research efforts have evaluated and compared transport poverty between various major cities in Canada. In one such study investigating neighbourhoods with low socio-economic status in Canada's eight largest cities, around 5% of the population was found to be living in low-income households which are also situated in areas that have limited access to employment by transit (Allen, et al. 2019). Expanding on these initiatives, this work aims to achieve a Canada-wide analysis and a national standardization of measures of access by active and public transportation.

This report advances the Statistics Canada's Proximity Measures Database (PMD) methodology through quantifying access to amenities and services via public transit and active transportation networks. A total of seven measures of access by transit were produced: (1) access to employment, (2) access to healthcare facilities, (3) access to public K-12 education facilities, (4) access to public post-secondary facilities, (5) access to grocery stores, (6) access to sports and recreation facilities, and (7) access to cultural and art facilities. Defining metrics on the DB scale permits the most micro-level Census boundary aggregation of access indicators, in turn enabling block and neighbourhood level comparisons across all of Canada. Further, having a nationwide standard metric for access enables accurate comparisons among different regions.

Measuring access to services in this way is not without precedent, with many similar projects being undertaken at smaller scales, such as a city or municipality. Several academic case studies and pilot projects have attempted to model access using active transportation and public transit. To create measures that are suitable for this type of analysis, as well as to ensure the best possible backwards comparability, a panel of experts at Mobilizing Justice[2] was consulted to gather feedback on best practices and common issues.

The remainder of this report details the background of the spatial accessibility measures, data sources and dependencies, methodology (from data acquisition to model computation), and provides highlights on key results. Specifically, Section 5 provides an overview of the modeling approach to measure access to amenities and services by public transportation; Section 3 explains the data dependencies required to produce the measures; Section 4 describes the methodology applied to acquire the appropriate data and calculate the spatial accessibility measures. Section 6 details the results and provides summary statistics of the measures, while Section 7 discusses the limitations given the data and assumptions applied for this study.

1. See: Infrastructure Canada - Active Transportation Fund - Applicant Guide.
2. See: Mobilizing Justice.

3 Data Sources and Dependencies

In summary, the following data are required to compute the measures of access by transit: (1) public transit routes, schedules, and stops formatted according to General Transit Feed Specification (GTFS) and (2) road networks for calculating durations; (3) origin and destination point locations (i.e., latitude and longitude); and (4) count of amenities, revenue, or number of employees of services linked to the destination locations. This section details the data leveraged or excluded to meet the project's scope.

3.1 Data Sources for Destination Masses

To measure the mass of amenities within each DB (or m_j), both open data and Statistics Canada administrative data were leveraged.

3.1.1 Business Register (BR)

Statistics Canada's Business Register (BR) was used to compute masses for the access to places of employment, healthcare facilities, and grocery stores. This is a continuously maintained central repository of businesses and institutions operating in Canada. For this project, the 2021 records from the BR were used. Although employment counts and other variables in the BR do not have the same level of accuracy and timeliness of specific labour survey programs, the major advantage is its comprehensive national coverage. Few other sources of business information have such a characteristic. Details on the data leveraged from the BR is provided in Section 4.5 Measuring presence and destination masses.

The specific North American Industry Classification System (NAICS) codes used are as follows:

- Healthcare facilities: 6211 (Offices of physicians), 6212 (Offices of dentists), 6213 (Offices of other health practitioners), 621494 (Community health centres), and 622 (Hospitals)
- Grocery stores: 445110 (Supermarkets and other grocery [except convenience] stores)

3.1.2 Linkable Open Data Environment (LODE)

Statistics Canada's Linkable Open Data Environment (LODE),[3] and specifically the Open Database of Educational Facilities (ODEF), Open Database of Cultural and Art Facilities (ODCAF), and Open Database of Healthcare Facilities (ODHF) were used for computing transit access to public K-12 education facilities, to public post-secondary facilities, and to cultural and art facilities. The LODE databases are compiled from open and public data sources, and in the case of these two databases, they are all georeferenced by the longitude and latitude point location of each facility. The ODCAF was collected in 2020 and contains nearly 8,000 records of art or cultural facilities.[4] Also, the ODEF version 2 was significantly adjusted to accommodate the breakdown of public K-12 and public post-secondary institutions for two separate measures of access. The ODEF was collected between 2019-2021 and contains over 18,500 records of educational facilities.[5] Further details on these two measures are provided in the Methodology, Section 4.5. The ODHF was utilized to supplement the BR healthcare data.

3.1.3 OpenStreetMap (OSM)

To ensure comprehensive national coverage that is not yet achieved by the Open Database of Sport and Recreational Facilities, OpenStreetMap (OSM) data was prepared for computing access to sports and recreational facilities. OSM provides crowdsourced georeferenced data. Similar to Google Maps, OSM is a platform that offers a map of the world; however, their data is sourced from an international user group and is freely accessible with an Open Database License (ODbL).[6] In order to define a subset of OSM data to be used, key values and associated tags were identified which represented sports and recreational facilities. Two OSM keys were used, 'leisure' and 'amenity', of which several tags were selected. Table (in appendix) provides the specific tags selected from each key. An extract of all OSM data within Canada with an 'amenity' or 'leisure' key was downloaded from GeoFabrik[5] on October 20, 2022. From this extract, all features with one of the sports and recreation facility tags were kept for computing the access measures.

3. See: Statistics Canada - Linkable Open Data Environment.
4. Art or cultural centre, artist, festival site, gallery, heritage or historic site, library or archive, miscellaneous, museum, theatre/performance and concert hall.
5. Early childhood education, kindergarten, elementary junior secondary, senior secondary, and post-secondary.
6. See: Accuracy Evaluation of the Canadian OpenStreetMap Road Networks.

3.2 Data Sources for Routing

3.2.1 Public Transit Data: General Transit Feed Specifications (GTFS)

Public transportation data structured in the GTFS format is required to compute the durations between origin and destinations (or d_{ij}) for transit access measures. The GTFS is a widely used transit standard that has made thousands of municipal and regional transit systems interoperable on platforms like Google Maps, allowing users to more easily navigate and plan transit commutes. While there are both real-time and static transit feeds, only static GTFS sources were utilized for this study given their simplicity and greater availability among data providers via open data portals. A static feed is comprised of a zipped folder containing machine-readable (.txt) files, including stops, agency, calendar, directions, and several other optional files.[7] Details on the coverage applied for this project are provided in the subsequent sections.

3.2.2 Road Network Data

Aside from GTFS sources, road network data is required for routing; this analysis makes use of the OSM road network. Existing literature has demonstrated the quality of the OSM road network within Canada is reliable compared to proprietary solutions, after evaluating the completeness, positional accuracy, attribute accuracy, semantic accuracy, and lineage of proprietary versus crowdsourced sources (Zhang and Malczewksi, 2018). Hence, the use of OSM's road network was deemed of sufficient quality for this project. An extract of all OSM data within Canada was downloaded from GeoFabrik[8] on August 27, 2022, for review of tag frequencies in Canada.

3.2.2.1 Road network data for active transportation

To assess Canadian cycling data coverage, cycling-related OSM highway, cycleway and bicycle tags specifically used by Valhalla[9] for routing were evaluated. Raw OSM data was extracted from Geofabrik and features were statistically summarized. The most prevalent cycling-related "highway" tags in the OSM network file which are leveraged by Valhalla were "service", "residential", "unclassified", "track", "path", "tertiary", "secondary", "primary", "cycleway", and "trunk".

Examining tags within the "cycleway" tag, the most prevalent features are "lane", "crossing", "shared_lane", "no", "shared", "track", "shoulder", "separate", "share_busway", and "yes". Tagging errors affiliated with crowd-sourcing and manual entry are evident with unconventional tags such as "I" and "1". However, these occur minimally and are excluded by Valhalla. In addition, tags are not ubiquitous across provinces.

Valhalla's generous inclusion of OSM road features results in cycling data having extensive national coverage with around 99.8% of Canada's population residing in DBs that intersect with an OSM network features.[10] Likewise, pedestrian related OSM tags used by Valhalla for routing (highway, foot, and pedestrian) were evaluated using protocol buffer binary OSM data extracted from Geofabrik. All road features that are denoted as cycling or pedestrian traversable – i.e., roads where cycling or walking are prohibited – were included in the cycling and walking Spatial Access Measures.

The most frequent tag observed in OSM features in Canada was "highway=service", which predominately occurred in Ontario and Québec. It is worth noting that shorter length features are expected to appear more frequently; for example, "highway=tertiary" represents roads connecting minor streets to major streets and is expected to be shorter than "highway=path", which represents a generic non-motorized feature. The shorter length may explain why tertiary tags are more frequent in Québec compared to path features.[11]

7. For further details of the specification see: Google - GTFS Reference - Stops.txt.
8. See: Geofabrik Canada.
9. See: Valhalla.
10. Due to the broad inclusivity of highway tags, there may be over-representation of unsafe or non-preferred cycling paths. For example, trunks ("important" roads that are not classified as motorways) and motorways are substantially present in Canada. Expectedly, routing with these paths would be mitigated through costing and current analyses on costing model sensitivity suggest that road types routed upon are able to be biased towards those of higher safety by adjusting the use_roads penalty to 0. That said, further investigations are underway to ascertain an efficacious costing threshold.
11. As mentioned in reference to cycling tags, generous inclusiveness of highway tags may over-represent unsafe or non-preferred pedestrian paths. On the other hand, the "pedestrian" tag, signifying a road that is mainly or entirely for pedestrians in which limited motorized vehicles may be authorized, is seldomly represented among provinces and territories and is missing in Nunavut and Prince Edward Island altogether. Moreover, as OSM is a crowd-sourced database, feature tagging may not consistently follow definitions set forth by OSM wiki page. Expectedly, routing preferences would be mitigating through costing; however, exclusion of unsafe or non-preferred road types for walking is undetermined for this project.

Similarly, cycling-related highway tags, Ontario and Alberta comprised the greatest proportion of OSM walking infrastructure in terms of length, followed by Québec. Once again, the territories comprised the smallest share of OSM walking features, each constituting less than 1% of total feature length. Both British Columbia and Saskatchewan constitute around 14% of Canada's total infrastructure length while maritime provinces comprise 0.5-3.5%.

3.2.3 Building Footprint Data

To create origin and destination points for the public transit, biking and walking access measures, a comprehensive dataset of building polygons from the best possible sources was created. To do this, three data sources were leveraged. These data sources were combined using the following hierarchy to create the best possible representations of buildings in Canada.

3.2.3.1 Open Database of Buildings (ODB)

The first source is the Open Database of Buildings; it comprised of data from municipal, regional, and provincial sources available to the general public through open government portals with open data licenses. The attributes from these sources vary from provider to provider and the common attributes are harmonized into the final product. Since these are administrative data generated by municipalities, the data quality is deemed to be generally high.

The second source is OpenStreetMap (OSM) which is a data set comprised of data from various sources all compiled using OSM formatting. The attributes vary widely as many of them are user generated. The polygonal quality varies quite a bit from region to region. This is a by-product of the user generated nature of the data.

The third source is Microsoft (MS). Using machine learning and computer vision, as well as the ODB as a training data set, Microsoft utilized its collection of satellite imagery to create a vast dataset of building footprints in all thirteen provinces and territories.[12] Due to the automated nature of the creation process, the data quality and presence of false positives and negatives are difficult to quantify.

A snapshot visualizing the footprints from the three data sources is shown in Figure 1. An approach analogous to the one specified on our GitHub was utilized.[13] That is, all stand-alone features (features that do not intersect with any other features) from each of the three sources are retained. In addition, ODB features that intersect with OSM or MS features are retained and OSM features that intersect with MS features are retained.

12. See: Microsoft - Canadian Building Footprints.
13. See: CSBP - Integrated Canadian Building Footprints.

Figure 1
Building footprint data from the Open Database of Buildings (green), OSM Buildings (blue), and Microsoft (red)

Source: Open Database of Buildings, OpenStreetMap, Microsoft Canadian Building Footprints.

4 Methodology: Data acquisition preparation

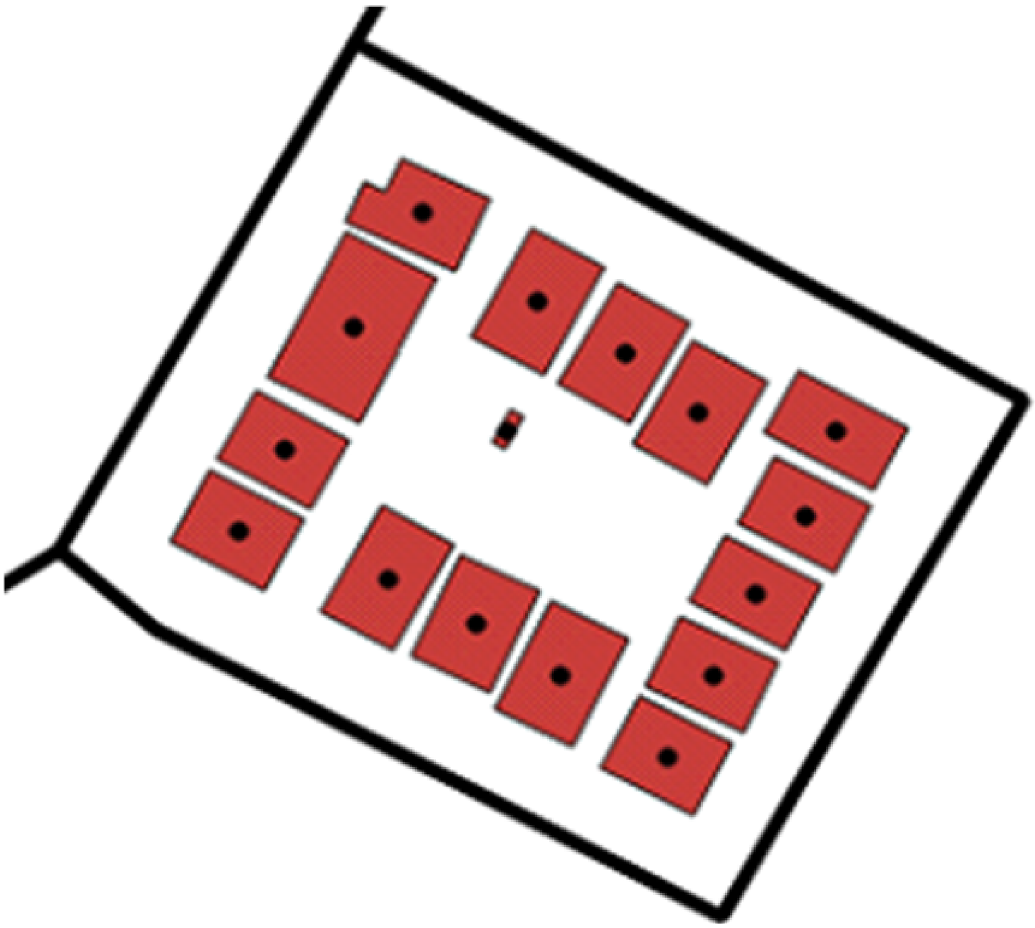
This section provides details on the following steps to compute the measures: (1) collecting and processing the General Transit Feed Specification sources; (2) calculating the travel time matrix from DBs to DBs with the compiled GTFS sources; and (3) computing the seven measures of access by transit-based equations (1) and (2) outlined in Section 5.

4.1 Origin and Destination Points

Origin and destination points were derived using 2021 Census Dissemination Blocks (DB) and building footprints. A DB is defined as an area bounded on all sides by roads and/or boundaries of standard geographic areas. DBs cover all the territory of Canada and are the smallest geographic area for which population and dwelling counts are disseminated (Statistics Canada 2017b).

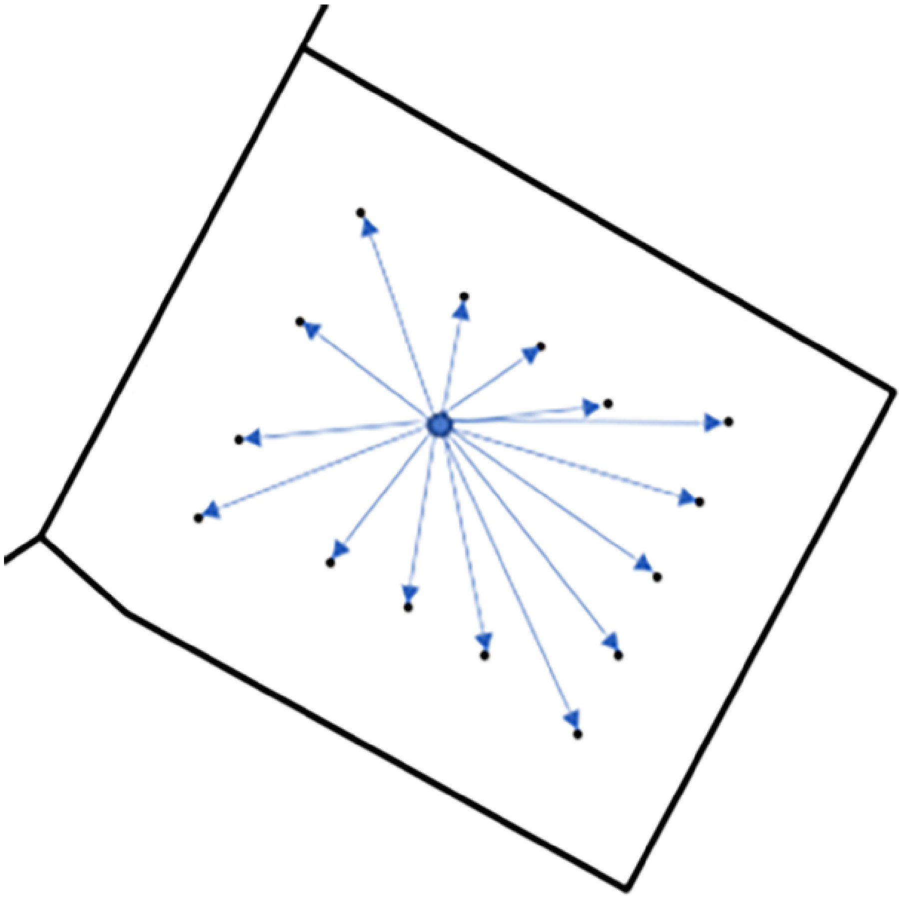
Representative points for each DB are determined by the following process. Firstly, the centroid points of each building in the DB are identified. Then the distance between every pair of building centroids is calculated. Finally, the building centroid with the lowest mean distance (i.e., it is the building located closest to all other buildings in the DB) is selected as the representative point for that DB (Figure 3).

Figure 2
Centroids of Building Footprints for a Specific DB



Source: Open Database of Buildings.

Figure 3
Building Footprint Centroid with Lowest Mean Distance from all Other Building Centroids



Source: Open Database of Buildings.

4.2 GTFS Data Collection

For this analysis, public transportation was defined as regularly scheduled, fixed route transportation systems available to the general public. This meant that on-demand services, typically for users with special accommodation requirements (e.g., para transportation, school buses) or in smaller communities, were excluded from the measurements. Moreover, larger inter-city services were excluded given their price and required advanced booking. Under these conditions, the following modes of transportation were considered public transportation: bus, commuter train, subway,[14] sea buses, light-rail trains, and streetcars (trams) services.

National coverage of GTFS data was compiled by surveying all openly available GTFS sources across the 5,161 Census Subdivisions (CSD) in Canada as of 2021. CSDs that did not have an easily accessible website or data portal to navigate and download the GTFS data were reviewed in Google Maps. If the CSD had public transit routes provided through Google's map interface, it was assumed that internally housed GTFS data may exist. As such, the data providers for those CSDs were contacted directly via email, requesting access to their static feeds, if available. From the 42 data providers contacted, 16 were able to provide GTFS data. To verify and supplement coverage obtained through the CSD search, non-government data portals, Transitland and OpenMobilityData[15] (formerly known as TransitFeeds), as well as an unofficial list of transit operators in Canada were referenced.[16]

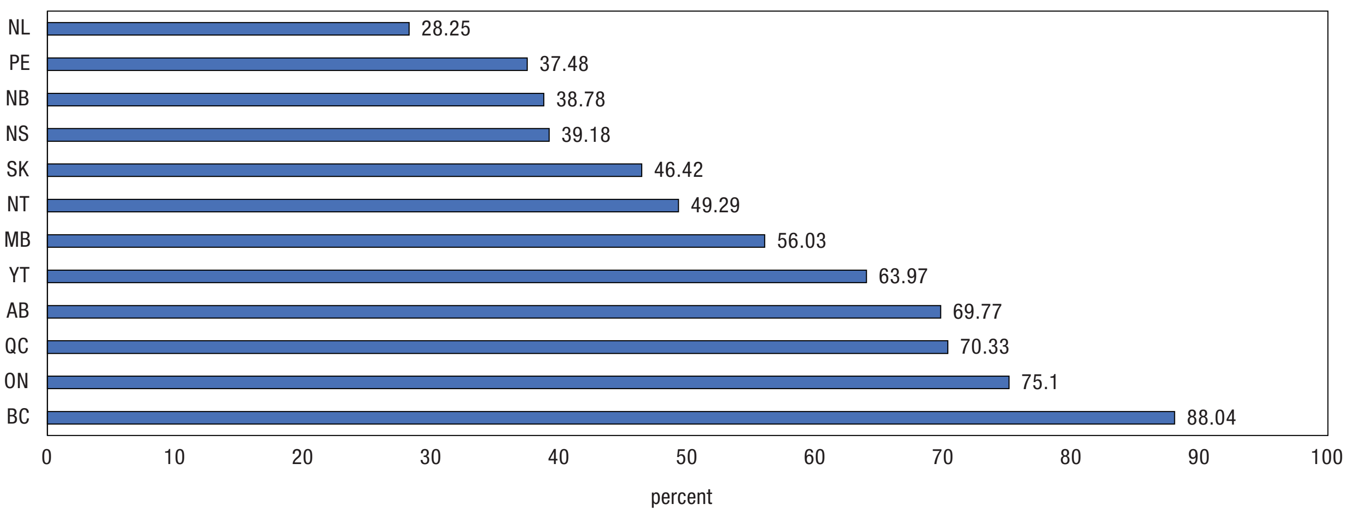
14. This includes SkyTrain.
15. See: Open Mobility Data.
16. See: Wikipedia - Public Transport in Canada.

The collection of GTFS data resulted in 106 sources with over 144,000 transit stops used for computing travel time matrices given availability to download via online open data portals and via email from the providers/municipalities that were contacted. Sources of GTFS data predominantly came from Ontario, Quebec, British Columbia, and Alberta.[17]

To assess rural and urban coverage, DBs that do not fall within Census Metropolitan Areas (CMAs) or Census Agglomerations (CAs) were considered rural while those that did were considered urban areas. British Columbia had not only the greatest provincial population coverage (86.88%) but also coverage of rural regions with 2.64% of the population living in DBs that are within 1-km of a transit stop. The province with the next greatest rural coverage was also Québec with less than 1% of the population outside of CMAs or CAs residing within 1-km of a stop. DBs with GTFS transit stops outside of CMAs or CAs only existed for British Columbia, Québec, Alberta, and Ontario.

Chart 1 presents the percentage of the urban population per province and territory living in DBs within 1-km from any transit stop. Population counts for each DB were derived from the 2021 Census.

Chart 1
Percentage of urban population living in Dissemination Blocks (DBs) within 1 km of a transit stop by province and territory



Note: Bars are coloured according to the percentage of provincial/territorial population in an urban area that resides within 1-km of a GTFS transit stop. Nunavut was excluded as no transit data is available within the territory.
Source: authors' computations.

4.3 Calculating Travel Times

As mentioned above, r5r was leveraged to compute many-to-many (or DB-to-DB) travel-time matrices within each transit region. A series of parameters were tested, including varying the travel-time window, maximum walking speed and distance, and start date/time. As a result of this testing, a commuting duration threshold of 90-minutes was applied with the default walking speed of 3.6 km/hr and a max walking distance from and to transit stops of 1000-meter were assumed. El-Geneidy et al (2014) identified that the 85th percentile walking distance from home to bus stops is 524 m and from home to commuter rail stops is roughly 1,259 m in Montréal, Canada. Considering this study does not distinguish between types of transit across both rural and urban regions, it was decided to use 1,000-meters instead.

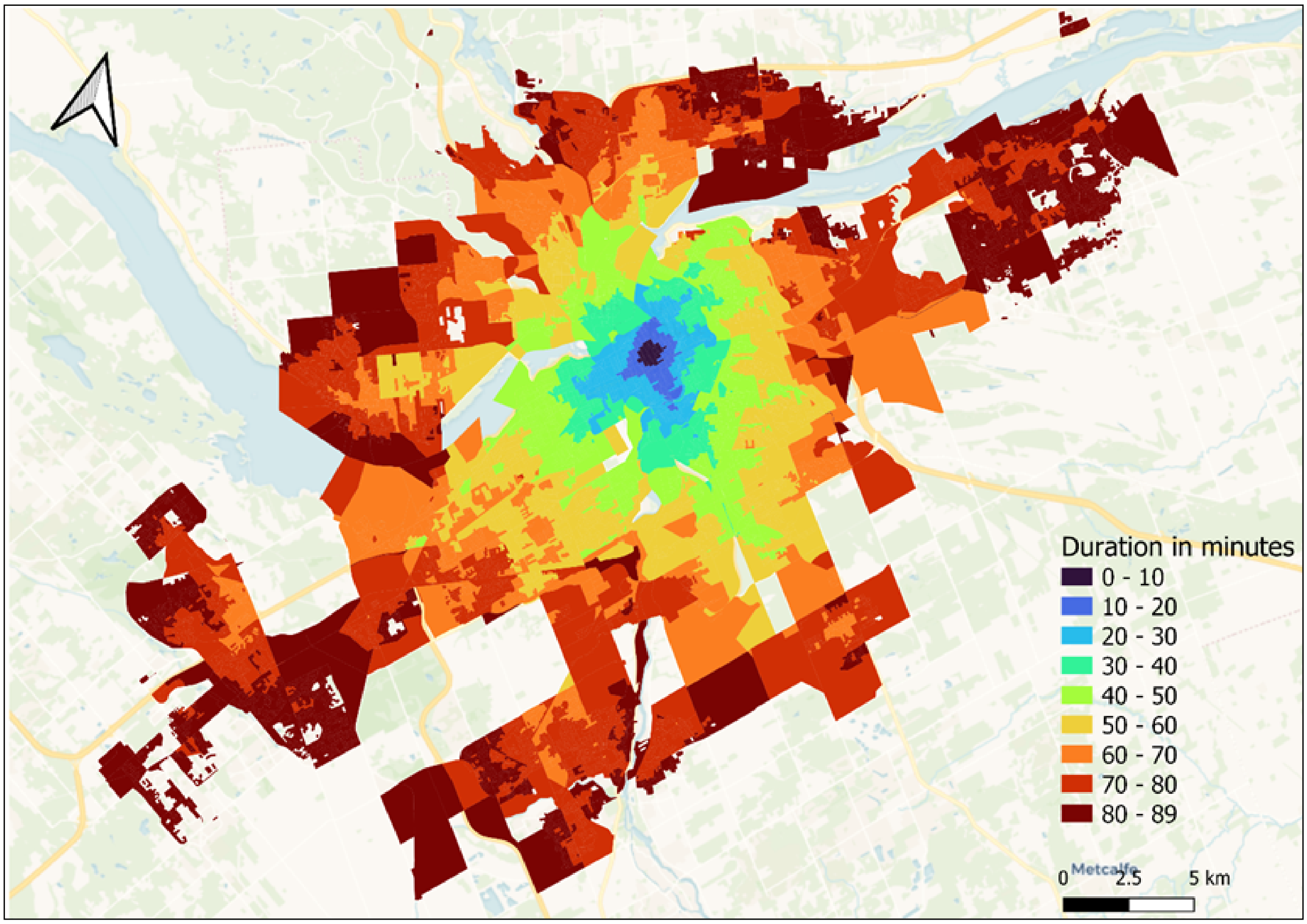
Transit travel times were ultimately computed for each minute from 7:00 AM to 9:00 AM (2:00 PM to 4:00 PM for off-peak hours) on several different Thursdays in the 2021 – 2022 calendar year to ensure peak hours were

17. Note that routes covered by on-demand services shared taxis (route type 1501) were removed from the route.txt file of the following municipalities: Chambly-Richelieu-Carnignan, L'Assomption, La Presque'Île, Laurentides, Le Richelain, Longueuil, Richelieu, Roussillon, Sainte-Julie, Sud-Ouest, and Terrebonne-Mascouche (all of which are part of the Exo – Réseau de transport métropolitain operator). Similarly, intercity travel such as BC Ferries and Via Rail were also excluded on the basis that tripes were typically longer than the defined duration threshold and fare prices were higher than typical public transportation services. More details about the GTFS data are available from the authors on request.

covered and the median duration represented final estimates. This meant that service areas with wide temporal coverage, especially high frequency in non-peak hours, were not favoured over service areas that only serviced peak hours.

For example, Figure 4 below illustrates the time between an origin DB centroid in Ottawa to all destination DBs that fall within a 90-minute duration threshold during peak transit hours. The map also showcases the use of both the Ottawa and Gatineau transit operators' systems within this transit region of Ottawa, Gatineau and La Pêche. Isochrones for public transit generally have "islands" where transit stops and hubs are more widely dispersed. This is a result of the stops being reached more quickly while riding the public transit system than by walking.

Figure 4
Example of travel times (in minutes) from an origin DB centroid (red point) to all DB centroids within a 90-minute travel time threshold, Ottawa-Gatineau region



Source: Statistics Canada, Centre for Special Business Projects & © OpenStreetMap contributors, © CARTO.

To validate the accuracy of the travel-time matrices, random DBs (representing origin points) were evaluated within each transit region. This was done by creating a subset of DBs that spatially overlapped both with a regional geography (CMA, CSD, or CD) and with a 1000-meter buffered area derived from all the transit stops. In addition, Google Maps was leveraged to provide some context to our results, offering the best approach to ground truthing the travel-time matrices produced.[18] Through this validation exercise, if an origin DB location's travel time to

18. Results from the r5r travel time computations generally indicated higher travel times, and thus smaller spatial distributions to destination DBs, compared to what was presented in Google Maps, but it was presumed this was due to Google's different routing algorithm and road network data, as well as the use of real-time GTFS and traffic data.

destination DB locations differed significantly from the results in a similar Google Maps query, that area was flagged for further inspection.[19]

After validating, a final output of 53 travel-time matrices, one for each travel region, with a total of nearly 320 million pairs of DB-to-DB durations were computed for the model.

4.3.1 Routing via Public Transportation

Adapted from PMD methodology (Alasia et al. 2021), the travel duration, d_{ij}, computations were adjusted for public transit routing. For the PMD, distance was calculated with the assumption that commuters were either walking or driving, enabling the use of the openrouteservice (ORS) open-source routing engine. This study measured transit access based on durations traveled on public transit systems, a mode of transportation that is not offered by ORS, Hence, an alternative approach was required to calculate travel times between DBs. After evaluating several routing engines, Conveyal's Rapid Realistic Routing on Real-world and Reimagined networks (R5)[20] was deemed the most appropriate routing engine for the purpose of the current analysis.[21]

4.3.1.1 Routing with Conveyal's R5

Conveyal base code is open source but the documentation for using R5 is limited. Therefore, the R5 R programming language package, r5r,[22] was used as a simple interface for operating the routing engine. It also has an active open community developing it, with substantial documentation. The routing engine, for R5 (and r5r), leverages the RoundbAsed Public Transit Optimized Router (RAPTOR), which is a unique routing algorithm designed specifically for public transit (tram, subway, bus, ferry, cable car, gondola, and funicular) because it leverages transit schedules to find the earliest arrival time and best route for each reachable stop to more efficiently compute aggregates of distance/time matrices (Delling, et al. 2015). Specifically, RAPTOR is efficient yet "realistic" because it computes durations every minute within a specified time window parameter to account for the varying nature of public transportation systems, notably transit routes and schedules change based on the time of day, day of week, and time of year. For example, if a 1-hour time window is assumed at 7:00AM, a duration from each origin to destination within a maximum duration (e.g., 90-minutes) is computed each minute between 7:00AM to 8:00AM, and then the median is assumed to be the best representative duration value for that origin-destination pair (Pereira et al 2021; Conway et al 2018; Conway et al 2017).

Besides specifying a time window, the origins and destinations must be provided (in this case DB centroids), and the mode of transportation used after egress from public transport must be specified. Additional parameters can be tweaked from defaults as well, including maximum walking distance and speed from origins to transit stops and from transit stops to destinations.

4.3.2 Routing via Active Transportation

To calculate cycling and walking durations between origin and destination pairs (i.e., Dissemination Block to Dissemination Block), the open-source routing engine, Valhalla, was applied because it leverages hierarchical indexed graphs as tile sets and an ensemble of algorithms (notably, Thor for routing and Sif for costing) that can account for variance in physical features (e.g., infrastructure and slopes). Similar to several popular open-source routing engines, Valhalla builds a graph derived from OpenStreetMap geometries and attributes (mainly for the road network), as well as elevation data from digital elevation models released on open data portals.

19. These problematic origins were then re-run following a more thorough inspection of the underlying data. Through this iterative process, issues related to how the data was stored and organized in GTFS were identified and resolved. For example, frequency based GTFS files (Montréal STM and Saguenay) led to errors in the routing algorithm; as a result, their frequencies had to be converted into stop times, which resolved the issue.
20. See: Conveyal - R5.
21. There are various open and paid services for calculating durations or distances between origin-destination pairs using public transit, including Google Maps, OpenTripPlanner, Mapbox's Valhalla, and Conveyal's R5. All rely on the standardized General Transit Feed Specification (GTFS); however, each has a unique approach for compiling data sources and computing networks, routes, and distance matrices. Transit schedules, routes, stop ids and locations, as well as calendar dates for which the data applies are organized in machine-readable, text-based formats following GTFS, which is further detailed in Section 3: Data Dependencies. Though Google Maps offers a reliable routing service using real-time GTFS and live traffic data, costs and limitations within the terms and conditions of this product deemed it out of scope. OpenTripPlanner is a legacy open-source tool for producing routes from GTFS and OpenStreetMap data. That said, it is computationally costly and intakes a single timestamp, ignoring schedule variability among transit operators and making it difficult to scale for nation-wide measures. Thereafter, Valhalla was considered because it flexibly offers costing parameters that penalize unfavourable routes (4.3.3 Routing via Active Transportation); but after issues building the Valhalla engine with pre-processed GTFS sources defined by another open-source project, Transitland, it was determined that the Conveyal routing engine, R5 demonstrated a more seamless implementation.
22. See: IPEA - r5r.

As mentioned above, Valhalla incorporates elevation as well as traffic grids (if available) in combination with weighted factors for dynamic costing. For example, the weights such as grade, surface material, etc. modify the cost factor which adds time penalties impacts the speed and thus travel time along an edge depending on the slope angle defined by the elevation grid. These weighted grades ultimately influence the resulting routes because lower graded edges will be prioritized over high graded ones. The weights are adjusted based on experience, comfort level, personal preferences, physical abilities and in the case of cycling, the type of bicycle itself. The specification of parameters used for dynamic route costing are reported in Table 3 of the appendix. Broadly, the specified values assume that road bikes are used and there is a preference for cycleways or roads with bicycle lanes that are separate from cars and roads, biasing the routes returned by Valhalla towards medium and high comfort pathways as classified in Can-BICS (Winters et al., 2020). Further, there is the assumption that less hilly paths are preferred.

The routing process is analogous for walking; however, a separate set of costing parameters are utilized. The specification used in this analysis is reported in Appendix Table 4. The specified values assume that the walker is more indifferent to changes in elevation relative to the cyclist but there are time penalties for traversing stairs.

4.3.2.1 Valhalla Costing Sensitivity Analysis

To better understand the effect of Valhalla's costing models on road types considered in routing, a sensitivity analysis was performed by modulating the use_roads parameter and reviewing overall cost, cost per minute, and cost per metre of each route.[23]

Although Valhalla's costing models are primarily comprised of parameters pertaining to rider preferences, sample analysis revealed both model sensitivity in road types routed upon and the possibility to bias against infrastructure that may be considered less comfortable by a rider. For example, arterial roads or roads within rural areas are suitably costed higher by Valhalla's costing models, with arterial roads having a cost per distance of 0.64 and roads within rural areas having a cost per distance of 0.82. It should be noted that additional factors beyond rider's preferences factor into the costing such as the cost of making turns. Further modelling efforts on these results may reveal a clearer threshold to parse out less safe infrastructure and allow for cycling metrics to be developed for riders of all ages and abilities. For this study, a cost per distance value of 0.7 was used to filter out less-desirable routes.

The decisions made for this measure were meant to exclude the most dangerous and unsafe routes for cycling, such as highways. It is not meant to represent access for all ages and abilities, as there may be routes included that may not be comfortable or safe for many users.

4.4 Defining Transit Regions

A total of 106 accessible GTFS sources of the 143 identified were accessible to download and were then processed into groupings of "transit regions". Transit regions were defined to compute the durations between DBs to DBs across different, yet linked, GTFS sources more easily. More specifically, each transit region is a compilation of GTFS sources that are grouped together based on the assumption that intersecting bounding boxes of the sources' transit stop locations meant commuters could connect between different transit providers. For example, Ottawa, Gatineau, and La Pêche have separate GTFS sources, but were grouped together as a single transit region because their GTFS stops overlapped. With this logic, a total of 54 regions were created with Figure 7 in the Appendix displaying the area and distribution of the transit regions.

Within each transit region, however, schedule dates of some GTFS sources presented some inconsistencies due to variance in the frequency at which transit providers updated their schedules or routes. This meant that each GTFS source was valid for a given period of time ranging from around 3 months to over a year. Identifying these temporal inconsistencies and adjusting the GTFS sources (e.g., downloading vintages) within each transit region to ensure the operators' schedules aligned was paramount to compute representative and accurate routes and durations.[24]

23. Results of the sensitivity analysis are not reported here but are available from the authors upon request.
24. For this analysis, schedule dates range from May 2017 to January 2022, with 66% of sources including transit data between 2020-2022.

4.5 Measuring presence and destination masses

The measure of mass is largely dependent on the nature of the service or amenity that the proximity measure is intended to capture. This section outlines some general considerations and the approaches used for the seven measures developed in this analysis.

As discussed in Alasia et al. (2021), there are two widely used approaches to estimate the mass of destinations: uniform and non-uniform weighting. Uniform weighting assigns a value of "1" to each DB that contains an amenity location. The result is that all DBs with a non-zero amount of service are assigned the same mass. It does not assess the potential scale of service provision. In this example, a DB with a corner store would receive the same weight as a DB with a major grocery chain location. Non-uniform weighting utilizes a mass that scales with the size of service or quantity of destinations, so for example a business's revenue or the number of employees may be used as a measure of mass.

All BR based masses are derived from a data set of all active businesses from 2021 that were extracted from the BR. Masses for BR entities were derived in the same fashion as in Alasia et al. (2021). That is, when necessary, enterprise-level reporting was allocated proportionally to the local-level using the BR allocation factors; entities with PO Boxes as addresses or unreliable geocoordinates were removed; and self-employed businesses with revenue under $30,000 were removed. The masses for places of employment and healthcare are derived from BR employment data while the mass for grocery store is simply a uniform mass. That is, the total mass of a DB is the sum of the masses of the grocery stores within in. Note that the spatial access measure for grocery stores follows a different specification which is detailed in Section 4.6.2.

Healthcare destinations were comprised of facilities servicing under the following NAICS codes in the Business Register (BR): 6211 (Offices of physicians), 6212 (Offices of dentists), 6213 (Offices of other health practitioners), 621494 (Community health centres), and 622 (Hospitals). This extract was supplemented by Ambulatory health care services, Hospitals, and Urgent care centres within the Open Database of Healthcare Facilities version 2 (ODHF). BR data on health care facilities was supplemented with facilities contained in the ODHF using a deduplication approach to remove facilities in common to both data sets. Furthermore, the ODHF records had their employment bin imputed using a random forest classifier.[25]

The measure of mass for amenities collected derived from open data sources is also a uniform mass. The total mass of a DB is the sum of the masses of the amenities within it except for sports and recreation facilities which were partitioned into 3 bins ([1-4), [4-6), 6+) for which the numerical categorical code, ranging from one to three, was used as a mass. The total mass

5 Modelling

The gravity model adopted for this analysis builds on the methods developed in Alasia et al (2021), who produced the first Proximity Measure Database for Canada (PMD). This section provides a summary of the changes in the methodology, namely the weighting and routing applied to produce the twenty-eight spatial access measures. For more details on the conceptualization and implementation of access measures using a gravity model approach, the reader should refer to Alasia et al. (2021).

5.1 Primal (Gravity) Model

Broadly, a measure of spatial access from an origin to a destination can be abstracted to the distance between them and the mass of the destination, a proxy of the quantity of service provisioned at the destination. In other words, destinations closer to an origin have stronger potential attraction, or they are associated with a greater willingness to travel there, compared to destinations further away. Moreover, the amenity mass influences the potential opportunity to receive services. For example, access to a small convenience store would not provide the same amenity mass as a large grocery store.

25. Random forests are an ensemble machine learning technique that operate by constructing a multitude of decision trees and outputting the class that is the mode of the classes for classification of the individual trees.

Mathematically, the attractiveness of a destination point j from an origin point i is proportional to the amenity mass (m) at j and to the willingness to travel (f) the distance or duration (d) between i and j. Given that, the access level (AL) of origin point i is the summation of the destination attraction of all destinations J within a designated range of i.

Equation (1) illustrates the formula and conditions that define the access level for a geographic unit i:

$$\mathrm{AL}_i = \sum_{j \in J} f\left(d_{ij}\right) \cdot m_j$$

where

$$f\left(d_{ij}\right) = 1 - \left(1 - e^{-\frac{d_{ij}}{\lambda}}\right)$$

Similar to the PMD, origin and destinations points are approximated with the representative points of the Dissemination Blocks. Furthermore, amenity mass is aggregated at the DB level.

However, different from the PMD, a decay or impedance function, which is expressed as $f\left(d_{ij}\right)$, is applied to account for the variation in willingness to travel to different types of amenities. It makes use of the complementary exponential cumulative distribution function, commonly used in transit literature (Palacios and El-Geneidy 2022; Bauer and Groneberg 2016; Luo et al. 2014; Wan et al. 2012; Kwan 1998). The parameter, λ, is chosen such that the median duration of travel to that amenity (see more detail below) corresponds to the midpoint of the interval $[0,1]$. The use of a time or distance decay function assumes a continuous distribution and weights destination DBs based on their relative distance to an origin DB. Essentially, the use of a decay function penalizes destination DBs that are located further away from the origin DB.

Median durations of travel per each type of service or amenity were determined from the General Social Survey (GSS), 2015.[26] This survey offers appropriate results to ensure a data-driven approach that better represents real commute times across Canada. Based on the twenty-eight access measures, the durations and weights of the following amenity and service types were extracted from the survey results (with the GSS type identifier): work (301), grocery (306), sports (308), and health (311) for public transit (316), walk (315) and bike (318) modes of transportation. Then, with the series of durations and their corresponding survey weights, the median duration of travel to each type of amenity was identified, as seen in Table 1. The decay curves with the aforementioned median duration of travel midpoints are displayed in Figure 5. The slope of the curve is influenced by the parameter, λ, in which a smaller value of lambda results in a steeper curve. Thus, a higher value of lambda indicates a willingness to travel further for a particular type of amenity. For example, one is typically more willing to travel further for employment ($\lambda = 29$) than for groceries ($\lambda = 22$).

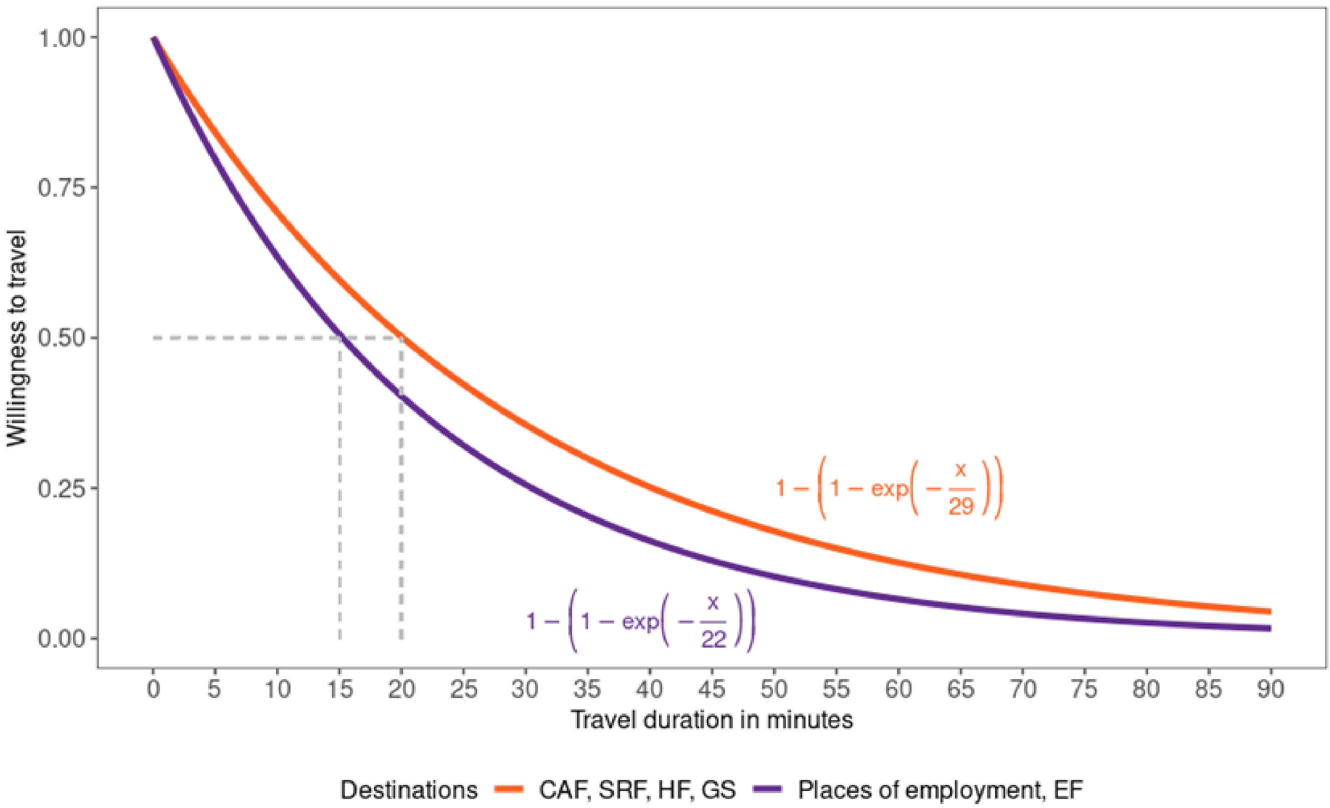
Table 1
Median trip durations by destination type in minutes

Type of destination facility	Median Duration Public Transportation	Median Duration Cycling	Median Duration Walking
	in minutes		
Cultural and Arts	15	15	10
K-12 Educational	20	20	10
Post-Secondary Educational	20	20	20
Recreational	15	15	10
Place of employment	20	20	15
Healthcare	15	15	10
Grocery Store	15	15	10

Source: GSS and authors' computations.

26. Statistics Canada - Cycle 29, 2015: Time Use, Episode File, specifically the Public Use Microdata files.

Figure 5
Decay curves by lambda parameters for cycling and public transportation



Source: authors' computations.

Lastly, following the methods applied for the PMD, all transit access measures presented in this study can then be expressed in relative terms, as rescaled indices. That is, the access level AL is transformed into an index, hereafter referred to as the Spatial Access Measure, SAM, by rescaling its values to a range between 0 and 1. As seen in equation (2), the following standard rescaling formula is applied:

$$\text{SAM}_i = \frac{\text{AL}_i - \min(\text{AL})}{\max(\text{AL}) - \min(\text{AL})}$$

The rescaling is done at the national level which is to say that $\min(\text{AL})$ corresponds to the lowest access level in Canada and $\max(\text{AL})$ corresponds to the highest. In this study, both the access levels and indices are provided.

5.2 Dual model

The dual or reciprocal of the accessibility model (Cui & Levinson, 2020), solves for the travel duration or cost to reach a minimum number of opportunities. The dual model can be expressed as:

$$\text{AL}_i = \max\left(Q_{ij} \cdot d_{ij}\right)$$

That satisfies

$$\min_Q \sum_{j \in J} m_j \cdot Q_{ij} \cdot d_{ij}$$

Subject to:

$$\sum_{j \in J} m_j \cdot Q_{ij} \geq n$$

$$Q_{ij} \in \{0,1\}$$

Where Q is the binary origin-destination (OD) matrix of the DB-to-DB network pairs, d_{ij} is the travel duration between origin i and destination j, n is some threshold of the count of amenities, and m_j is the mass of amenities in the destination DB j. This model is only used for the grocery stores measure and the mass used is the count of grocery stores in a DB.

The output of this model is the duration of travel in minutes to reach the n^{th} closest amenity.

This approach was chosen for two reasons, firstly, the distance to the nearest grocery stores is more relevant than the size of a grocery store grocery stores and secondly, being able to quantify whether there are a variety of grocery stores within a short duration of travel is important in identifying food deserts.

5.3 Assumptions

With any model, assumptions influence the output results. In the case of this project, the decay function (or willingness to travel) assumes a uniform commuter profile. Specifically, using a 1-km buffer around transit stops as an access constraint assumes commuters are willing and able to walk that distance to begin their trip. Cycling and walking parameters used in routing assume that commuters are equally abled and not significantly impeded by terrain or external activities, such as steep stairs or slopes. Though preference is given to avoid elevation and stairs if other routes are available. Another generalization over the population is found in the use of the General Social Survey as a representation of travel times. Although the generalization allows for standardized results, it operates on the expectation that commuters across Canada follow patterns and choices reported by the 2015 sample.

6 Key results: population with spatial access

The methods presented above resulted in a total of 28 measures of spatial access, computed for dissemination blocks (DB) located withing the spatial buffers used in this analysis, while DBs outside the spatial buffers are classified has not having spatial access for that service. These granular measures were released as Spatial Access Measure (SAM) database, which can be downloaded online.[27] This section summarizes these results by combining population counts data from the Census of population 2021 with the SEM database and focusing on the percentage of Canadians having access to the different amenities.

Highlights at presented at three geographic level, national level, provincial level, and an illustrative example of results at the metropolitan level. Local level analysis, where DB values can be mapped, is where the SAM database reveal its full potential for analysis and policy support. It should be stressed that what is presented here is just one and succinct analytical perspective that this database is enabling.

6.1 Highlights at the Canadian level

Aggregate results of SAM for Canada are reported in Table 2. At the Canadian level, SAM provide a broad framework to benchmark current spatial accessibility performances in aggregate terms. Despite inevitable methodological challenges, the current results could be used to monitor trends over time or, possibly, benchmark Canada to other jurisdictions.

Two main patterns emerge from the Canada level results. First, in aggregate terms, spatial access by transit is similar for peak hours and off-peak hours. Although the degree of access at local level may vary, when looking at the mere dichotomy “access/no-access”, the areas (dissemination block) served by public transit during peak areas are nearly identical during peak and off-peak hours.

27. See: StatCan - Spatial Access Measures.

Second, when looking at specific type of services or facilities and regardless of the means of transportation, the degree of spatial accessibility is greater for two essential services, health care facilities and primary education. Health care facilities are the most spatially accessible type of facilities; three quarters (74.5%) of Canadians have spatial access to these services by transit (peak and off-peak hours), while about nine out of ten Canadians have access to by cycling (93%) or by walking (90%). Primary education facilities are the second most spatially accessible services, with close to 70% of Canadians having access by public transit, 90% by cycling and nearly 83% by walking.

In contrast, post-secondary educational facilities are the least spatially accessible facilities among those considered in the analysis, with less than 60% of Canadians having access through public transit and about 40% and 12% having spatial access by cycling and walking, respectively. These results reflect a patten of spatial location observed for colleges and universities, which are frequently located on the outskirt of agglomerations. Hence, these services result markedly less accessible by cycling and walking, given the distance thresholds used in this analysis.

Recreational and sport facilities, on the one hand, and cultural and art facilities, on the other hand, rank relatively close in terms of aggregate spatial access. Between six and seven out of ten Canadians have access to these facility by public transit. Roughly between 8 and 9 Canadians out of ten have access to these facilities by cycling, while a more pronounced difference is noted for walking access, with nearly 80% having access to recreational and sport facilities and about 60% having access to culture and art facilities.

Finally, slightly over 60% of Canadians have access to a grocery store by transit, while around 87% have access to a grocery store within 15 minutes of cycling and around 71% have access to a grocery store within 15 minutes of walking.

Notably, places of employment have greater spatial accessibility among all amenities considered in this analysis. However, it should be recalled that the measure captures spatial access of people to any place of employment – not including self-employment, and not necessarily the employment held by each person. Keeping this in mind, results show that the vast majority of Canadian has some degree of spatial access to places of employment by cycling (97.2%) or walking (97.0%), while over three quarter have access by public transit (77.4%).

Table 2
Spatial access to selected amenities by active and public transportation in Canada

Type of service	Transit		Cycling	Walking
	Peak hours	Off-peak		
	percentage of people with some degree of access			
Health	74.5	74.5	93.4	90.3
Primary education	69	68.9	90	82.8
Postsecondary education	57.8	57	40.2	11.9
Grocery	62.9	61.9	87.1	71.3
Recreation and sport	68.4	68.9	89.4	79.8
Culture and art	65.8	65.6	85.1	60.3
Places of employment	77.4	77.4	97.2	97

Source: authors' computation.

6.2 Highlights for provinces and territories

When the analysis is disaggregated at the provincial and territorial level, a substantial degree of dispersion of spatial access measures is observed.

Overall, dispersion of SAMs at provincial and territorial level appears largely related to the different settlement patterns of each region of Canada, with more urbanized areas presenting, in general, higher degrees of spatial access, as compared to more rural provinces and territories.

Dispersion of provincial and territorial aggregate SAM values become particularly evident for access by public transit. Atlantic provinces report markedly lower percentage of population with spatial access by public transit, regardless of the type of service considered. This result aligns with shares of rural population of these provinces, roughly between 30% and 50%, which are among the highest in Canada.[28] In contrast, British Columbia, Ontario, and Quebec are consistently among the provinces with higher degree of access to services by public transit.

When considering any of the active transportation options available to Canadians, health care facilities and primary education facilities remains among the most spatially accessible type of facilities in all provinces and territories. For instance, except for Nunavut, in all other provinces and territories a share comprised between about 71% and 97% of the population could access healthcare facilities by cycling and between 63% and 95% by walking.

Overall, when looking at other active means of transportations, the main divide in spatial accessibility measures at the provincial level remains between Atlantic provinces and the three more urbanized provinces of Canada (Ontario, Quebec, and British Columbia), with prairies provinces standing on middle ground. Intuitively, this point to one of the main drivers of these results, the prevailing settlement patterns in these regions (high density urban vs. disperse and rural). This in turn call for a more localize level of analysis, which is illustrated with some examples next section.

6.3 Illustrative examples of analysis at the local level

Aggregate analysis provides a general understanding of broad patterns and issues at the national and regional level. However, the full analytical power of SAM data becomes evident when dissemination block level data are visualized. It is at local level that differences in access can be assessed between different type of settlements (high density, urbanized areas, vs. suburban areas, small town, or rural regions) or neighborhoods within metropolitan areas.

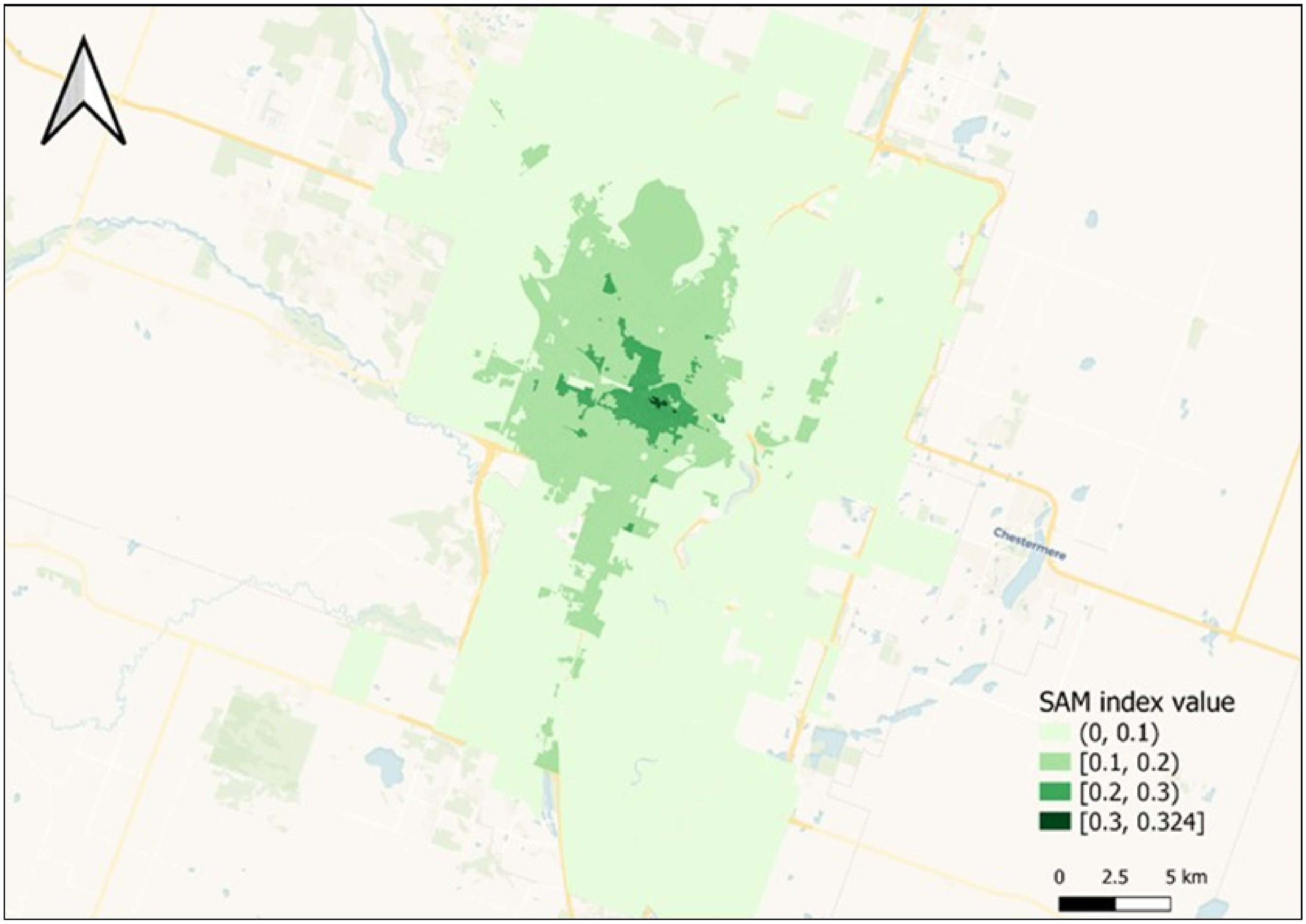
Given the size of the SAM database and the scope of this methodological paper, the results presented here are intended to be illustrative examples. Figure 6 shows the DB level value of spatial access measure to healthcare facilities in Calgary (Panel A) and primary educational facilities in Quebec City, using public transit in peak hours. In these maps, SAM values are grouped by deciles (see Figure's note). The results highlight large neighborhood differences in the degree of access to these facilities, driven by the location of the services and the layout of the public transit network.

As an example, 33.89% of the population of the census metropolitan area of Calgary live in an area with a high degree (top tercile) of spatial access to health care facilities by public transit in peak hours, while 11.51% live in neighborhood with low degree (bottom tercile, but some access) of spatial access, and 2.36% in neighborhoods with no spatial access (include outside transit region as well) by public transit. In an similar way, neighborhood level accessibility analysis of K-12 educational facilities indicates that 9.87% of the population of the census metropolitan area of Quebec City live in an area with a high degree (top tercile) of spatial access to K-12 educational facilities by public transit in peak hours, 14.48% live in neighborhood with low degree (bottom tercile, but some access) of spatial access, and 12.39% in neighborhoods with no spatial access to primary education by public transit.

These are merely two illustrative examples of SAM data use at the local level. Reader should keep in mind that this paper outlines a methodological framework providing some degree of flexibility in defining high and low level of access, while at the same time maintaining granular comparability across jurisdictions. SAM data at the DB level can be combine with other socio-demographic characteristics from other sources, for instance the census of population, to derive accessibility measures to services for different demographic cohorts, and sub-populations defined by gender, visible minorities, income level, employment status, etc.

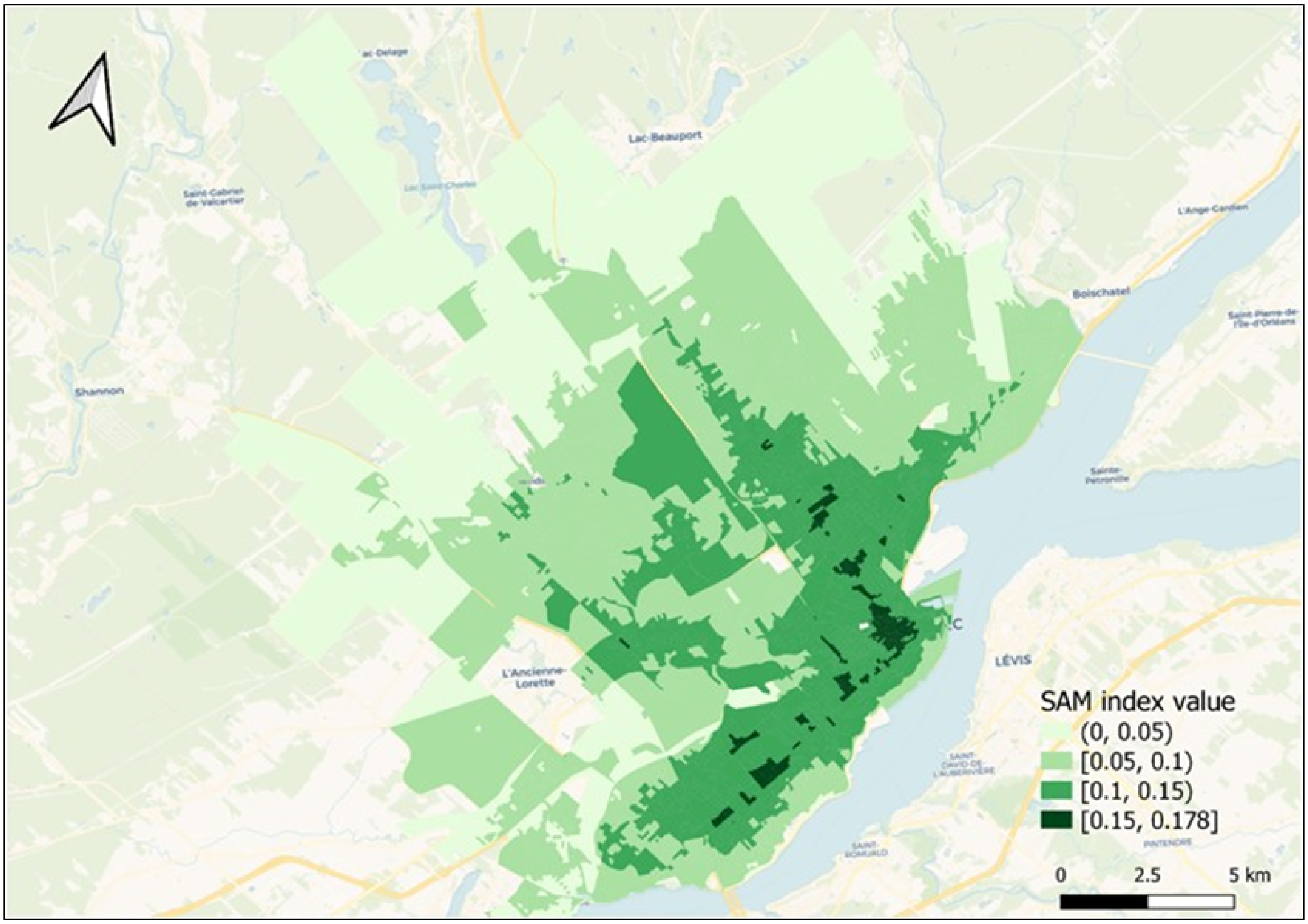
28. The share of population living outside Census Metropolitan Areas or Census Agglomeration (here referred as rural) was 47.2% in Newfoundland and Labrador, 37.1% in Prince Edward Island, 30.7% in Nova Scotia, and 36.8% in New Brunswick, compared to 18.2% in Quebec, 10.0% in Ontario, and 10.5% in British Columbia, the three most urbanized provinces of Canada (see: Statistics Canada - Population and dwelling counts by the Statistical Area Classification).

Figure 6-a
Access to healthcare facilities by peak transit in Calgary



Note: Symbology is based on intervals of 0.05 up to the maximum of 0.324 for the CSD of Calgary and 0.178 for the CSD of Québec City. Intervals consisting of only zeroes have been removed from the symbology. The index values are scaled between 0 and 1 where 0 is the minimum value for all of Canada while 1 is the maximum value for all of Canada. Since the index values are relative to the minimum and maximum values nationwide, most values appear quite close to 0.
Source: Statistics Canada, Centre for Special Business Projects & © OpenStreetMap contributors, © CARTO.

Figure 6-b
Access to educational facilities by peak transit in Québec City



Note: Symbology is based on intervals of 0.05 up to the maximum of 0.324 for the CSD of Calgary and 0.178 for the CSD of Québec City. Intervals consisting of only zeroes have been removed from the symbology. The index values are scaled between 0 and 1 where 0 is the minimum value for all of Canada while 1 is the maximum value for all of Canada. Since the index values are relative to the minimum and maximum values nationwide, most values appear quite close to 0.
Source: Statistics Canada, Centre for Special Business Projects & © OpenStreetMap contributors, © CARTO.

7 Discussion and Limitations

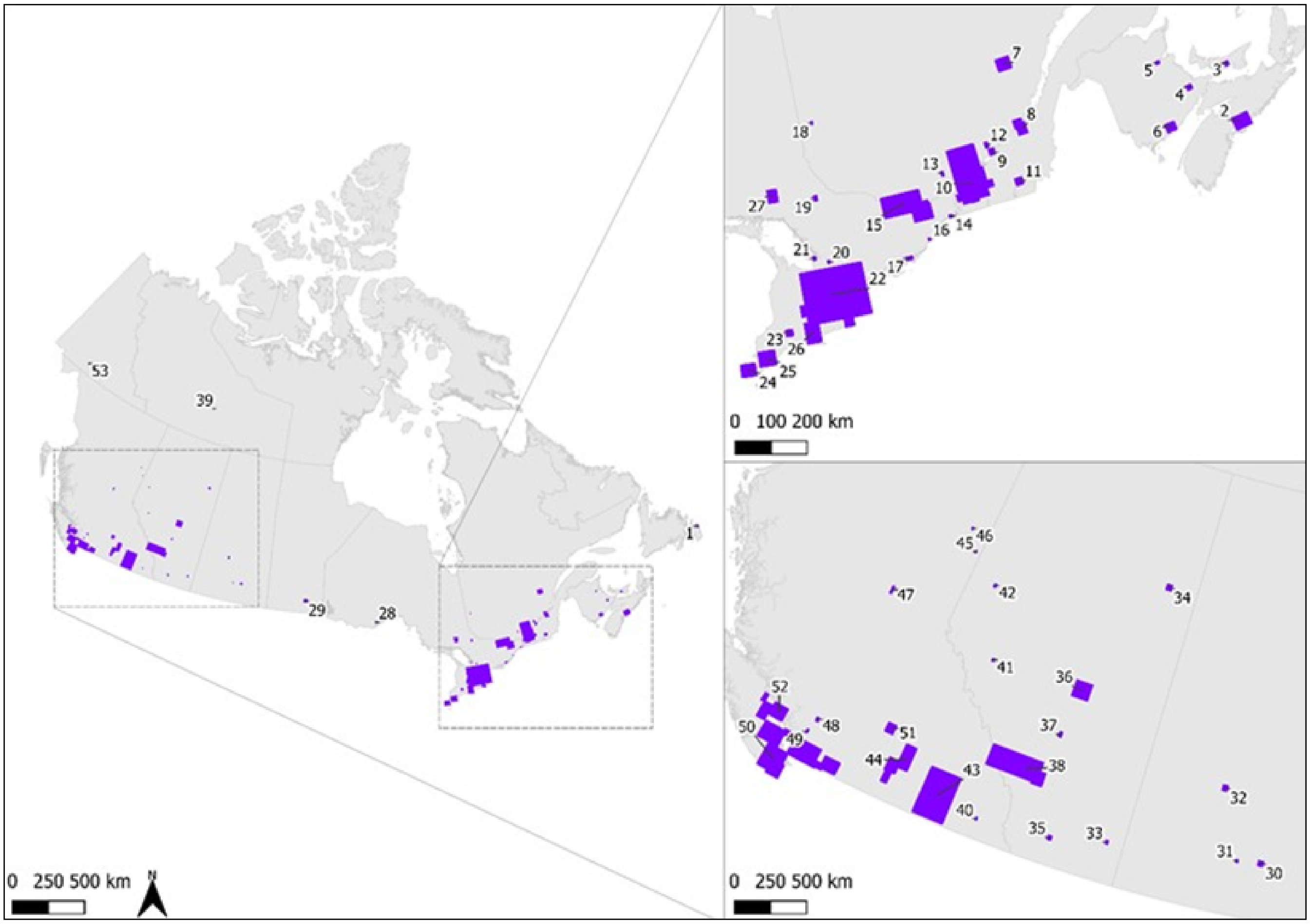
As noted throughout this report, there were several limitations to the data and methods. For this analysis, public transportation was defined as regularly scheduled, fixed route transportation systems available to the general public. This meant that on-demand services, typically for users with special accommodation requirements or in smaller communities, were excluded from measurements. As a result, more densely populated regions typically have more up-to-date and comprehensive coverage relative to rural areas. Also, inter-city transportation services that require pre-booking were not included, resulting in the exclusion of BC Ferries and VIA Rail from analysis.

Second, the static nature of GTFS data used in the model meant that only a snapshot of the transit system at a given date was represented. The newest available data was used to create estimates from each transit provider. For regions that had multiple transit operators, like Ottawa-Gatineau-La Pêche, static data with overlapping time periods were used to ensure accurate inter-operator trip estimates. This meant that our indicators did not reflect the exact reality of a given date, but rather a conceivable trip using the data available. Though variance in schedules within a day were accounted for with r5r's time window parameter (i.e., computing route every minute between 7:00-9:00 AM for peak hours), seasonal variances in schedules as well as circumstances such as COVID-19 impacting schedules were not considered for simplicity. In almost all jurisdictions, service providers changed routes and scheduling to accommodate the changing nature of public transportation during a global pandemic. These changes varied widely and often included limiting services to adjust for the decrease in overall public movement caused either by general lockdowns or specific event restrictions. While this impacted the data in this iteration, it did attempt to provide a more accurate snapshot of the realities faced by those looking to utilize public transportation during these times.

Thirdly, there are limitations to the base network itself. For example, misclassification of tags can occur, particularly in areas where there are fewer OSM contributors. These errors in tagging can compound when moving along to the routing process as it adds some error to the least-cost routing. Thus, it is possible that unsafe or even illegal (e.g., cycling on sidewalks) paths are considered routable; although, such cases are limited. Moreover, as geospatial mapping tools and data continue to improve and the number of contributors increases, the robustness OSM network and its tags also continues to improve.

There are inevitable limitations with measuring access to amenities and services through a model-based approach given the underlying assumptions that must be applied. Broadly, these assumptions include the following: (1) how the origins are defined; (2) how the destinations and their presence and size of service are represented; (3) the routing algorithm selected and associated required parameters (e.g., time window, walking speed, max distance, route preferences) are applied to compute the durations; and (4), the distance decay function leveraged to account for the willingness to travel. In turn, techniques such as applying standards outlined in existing literature and leveraging survey data (namely the GSS) to determine appropriate parameters (e.g., max walking speed, median travel duration to amenities) was implemented to ensure data-driven and fact-based assumptions. That said, further sensitivity testing could be explored to better evaluate and fine-tune the techniques applied to produce custom measures per each type of amenity or service.

Figure 7
Map of the transit region (i.e., intersecting bounding boxes of GTFS stops) across Canada



Source: authors' computations.

Appendix A

Table 3
The costing parameters used for bicycle routing from Valhalla

Bicycle Options	Description	Value
bicycle_type	Explains the type of bicycle used. The default is set as `Hybrid` - a cycle mostly used for city roads or casual roads/paths with good surfaces.	Hybrid
cycling_speed	Average travel speed on smooth and flat roads.	18 kph
use_roads	Cyclist's propensity to use roads alongside other vehicles. The value ranges from 0 to 1, where 0 indicates strongly prefer cycleways over roads and 1 indicates rider is comfortable using all roads.	0.0
use_hills	Cyclists desire to tackle hills and higher elevation paths. The value ranges from 0 to 1, where 0 indicates preference to avoid hills and steep grades and 1 indicates rider does not fear hills or steep roads (i.e., they are indifferent between hilly or flat paths of the same length).	0.0
use_ferry	Willingness to take ferries.	0.5
use_living_streets	Willingness to take living streets.[1] Living streets may be required to complete the route.	0.5
avoid_bad_surfaces	Represents desire to avoid roads with poor or rough surfaces relative to the type of bicycle used	0.25
shortest	Parameter to indicate the change in metric to quasi-shortest – purely distance-based costing.	FALSE

1. See: OpenStreetMap Wiki - Living Street.
Source: table taken from Valhalla documentation – costing models.

Table 4
The costing parameters used for pedestrian routing from Valhalla

Walking Options	Description	Value
walking_speed	Walking speed in kilometres per hour.	3.6 kph
walkway_factor	A factor that multiplies the cost when traversing designated footpaths or sidewalks along residential roads. A value of 1 indicates a preference for utilizing walkways.	1.0
sidewalk_factor	A factor that multiplies the cost when traversing roads with designated sidewalks. A value of 1 indicates a preference for utilizing sidewalks.	1.0
alley_factor	A factor that multiplies the cost when traversing alleys or narrow service roads.	2.0
driveway_factor	A factor that multiplies the cost when traversing driveways, which are often private, service roads.	5.0
step_penalty	A penalty in seconds added to each transition onto a path with steps or stairs.	30 s
use_hills	A pedestrian's desire to tackle hills and higher elevation paths. The value ranges from 0 to 1, where 0 indicates preference to avoid hills and steep grades and 1 indicates pedestrian does not fear hills or steep roads (i.e., they are indifferent between hilly or flat paths of the same length).	0.5
use_ferry	A pedestrian's willingness to take ferries to complete the route. The value ranges between 0 and 1. Values close to 0 will avoid routes involving ferries and values close to 1 will favor ferries.	0.5
use_living_streets	A pedestrian's willingness to take living streets.[1] Living streets may be required to complete the route. The value ranges from 0 to 1 where a greater value indicates that pedestrians prefer more scenic or less busy routes over shorter or faster routes. Not that this tag is not commonly used in Canada.	0.6
shortest	Parameter to indicate the change in metric to quasi-shortest – purely distance-based costing.	FALSE

1. See: OpenStreetMap Wiki - Living Street.
Source: table taken from Valhalla documentation – costing models.

References

Alasia, Alessandro, Frédéric Bédard, Julie Bélanger, Eric Guimond, and Christopher Penney. 2017. "Measuring remoteness and accessibility: A set of indices for Canadian communities." Reports on Special Business Projects, Statistics Canada.

Alasia, Alessandro, Nick Newstead, Joseph Kuchar, and Marian Radulescu. 2021. "Measuring proximity to services and amenities: An experimental set of indicators for neighbourhoods and localities." Reports on Special Business Projects, Statistics Canada.

Allen, Jeff, and Farber, Stephen. 2019. "Sizing up transport poverty: A national scale accounting of low-income households suffering from inaccessibility in Canada, and what to do about it." Transport Policy, 74, 214-223. doi: 10.1016/j.tranpol.2018.11.018.

Bauer J, Groneberg DA.2016. "Measuring Spatial Accessibility of Health Care Providers - Introduction of a Variable Distance Decay Function within the Floating Catchment Area (FCA) Method." PLoS ONE 11(7): e0159148. doi:10.1371/journal.pone.0159148.

Conway, Matthew Wigginton, Andrew Byrd, and Michael van Eggermond. 2018. "Accounting for Uncertainty and Variation in Accessibility Metrics for Public Transport Sketch Planning." Journal of Transport and Land Use .11 (1). https://doi.org/10.5198/jtlu.2018.1074.

Conway, Matthew Wigginton, Andrew Byrd, and Marco van der Linden. 2017. "Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks." Transportation Research Record: Journal of the Transportation Research Board 2653 (1): 45–53. https://doi.org/10.3141/2653-06.

Cui, M. and Levinson, D. (2020), Primal and Dual Access. Geogr Anal, 52: 452-474. https://doi.org/10.1111/gean.12220.

Delling Daniel, Pajor, Thomas, and Werneck, Renato F. 2014. "Round-Based Public Transit Routing." SIAM. https://doi.org/10.1137/1.9781611972924.13.

El-Geneidy, Ahmed, Grimsrud, Michael, Wasfi, Rania, Tétreault, Paul, and Suprenant-Legault, Julien. "New evidence on walking distances to transit stops: identifying redundancies and gaps using variable service areas." Transportation 41, 193–210 (2014). https://doi.org/10.1007/s11116-013-9508-z.

Farber, S., Winters, M., & Palm M. (2022). Cycling and Walking Access Metrics: Findings and Recommendations. Mobilizing Justice.

Kwan, Mei-Po .1998. "Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework." Geographical Analysis 30(3).

Palacios, Manuel Santana and El-Geneidy, Ahmed. 2022. "Cumulative versus gravity-based accessibility measures: Which one to use?" Transport Findings, February. doi: https://doi.org/10.32866/001c.32444.

Palm, M. & Farber, S. (2022). National Accessibility Metrics Workshop: Findings and Recommendations. Mobilizing Justice.

Pereira, Rafael H. M., Saraiva, Marcus, Herszenhut, Daniel and Braga, Carlos Kaue. 2021. "R5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R." Findings, March. https://doi.org/10.32866/001c.21262.

Zhang, Hongyu and Malczewski, Jacek. 2018. "Accuracy Evaluation of the Canadian OpenStreetMap Road Networks*,"* International Journal of Geospatial and Environmental Research*:* Vol. 5: No. 2, Article 1. https://dc.uwm.edu/ijger/vol5/iss2/1