URBAN INSTITUTE · ELEVATE · THE · DEBATE

Access to Opportunity through Equitable Transportation

Technical Appendix

Alena Stern, Christina Stacy, Kristin Blagg, Yipeng Su, Eleanor Noble, Macy Rainer, and Richard Ezike

October 2020

This appendix documents the technical steps supporting the *Access to Opportunity through Equitable Transportation* report and the "The Unequal Commute: Examining Inequities in Four Metro Areas' Transportation Systems" feature. In the report, we examine transportation equity and inclusion in different types of metropolitan regions and explore how these regions might track and improve transportation equity over time. We draw on case studies of four Metropolitan Statistical Areas (MSAs) facing different barriers to providing equitable transportation: the Seattle, Washington, MSA, a West Coast region that faces exponential growth and housing affordability issues; the Lansing, Michigan, MSA, a smaller Midwestern metro with a state capital and university; the Baltimore, Maryland, MSA, an East Coast metro with fiscal challenges and a declining population; and the Nashville, Tennessee, MSA, a sprawling southern metro with population growth but low population density in many areas. In the feature, we display results from a quantitative analysis that examines access to opportunity through transportation in these four case study cities. In this document, we detail the data sources and methodology used in our analyses.

Data Sources

Transportation Data

We use the open-source routing software OpenTripPlanner (OTP) to calculate travel time between pairs of start and end points. OTP requires two main inputs: transit data in the General Transit Feed Specification (GTFS) format and road grid data from OpenStreetMap (OSM). OTP uses these two data inputs to create a model of the transportation system (called a graph) that it uses to identify the fastest route between start and end points. (We discuss this routing process in greater detail in step 4 on page 7.) One limitation of both OSM data and OTP routing is that neither accounts for traffic conditions when calculating drive time. To address this limitation, we apply a traffic adjustment for car trips taken at peak times, explained in the "traffic data" section at the bottom of this page.

Road grid data: We obtain data for the road grid from OpenStreetMap, a free, open-source global map that includes detailed information about roadways and walkways. We first download the OpenStreetMap file for the United States from the GeoFabrik Download Server, which hosts OSM extracts that are updated daily. For each of our MSAs, we then clip the US file to a bounding box that fully contains the geographic area under consideration as described in step 1 on page 5.

Transit data: Data on public transit are sourced from the Transitland feed registry, which aggregates feeds from transit agencies in more than 50 countries. We identify the relevant transit agencies for our analysis by first identifying the states in which our start and end points fall (see step 1). We exclude transit agencies from our analysis that meet either of the following conditions:

- The transit agency's services are not open access, such as shuttles for a specific workplace or university that are only available to employees or students of that institution.
- The agency's data feed is incomplete or corrupted in a way that cannot be corrected during pre-processing (described in step 2 on page 5) and breaks OTP's graph build process when included.

We include data from 166 transit agencies and exclude data from 16 agencies (a full list is available upon request). To be comprehensive, we include all relevant agencies in each state covered by our analysis—even agencies with service areas that fall entirely outside our chosen focal areas. We then pull data for all the identified transit agencies using the Transit Land API. To capture transit system data before services were reduced in response to the COVID-19 pandemic, we use January 13–17, 2020, as our period of analysis. Transit agencies publish updates to their feeds at varying intervals, and each update includes the associated start and end dates. For each agency, we use the first update that covers our entire analysis period. Several—typically smaller—transit agencies have no data available on Transit Land for our period of analysis because they either stopped updating their GTFS feed before January 2020 or began reporting data after that time. These agencies are excluded from analysis.

Traffic data: We used INRIX's 2019 Global Traffic Scorecard to incorporate traffic times and congestion data into our analysis of trips by car. INRIX used anonymized observed trip data to identify the busiest commuting routes and areas in metro regions across the world. INRIX then creates multiple metrics related to congestion using travel times to and from these identified commuting areas, including peak speed (average miles per hour [MPH] during the most congested portion of morning and afternoon commutes), off-peak speed (average MPH during the least congested point of the commute), and free flow (fastest travel time over 24 hours).

Demographic and Geographic Data

Demographic characteristics: Data for demographic information are taken from 2014–18 ACS five-year estimates at the census block group level.[1] Racial and ethnic groups are defined as Asian (both Hispanic and non-Hispanic); Black or African American (both Hispanic and non-Hispanic); Hispanic or Latino; non-Hispanic white; and other race (both Hispanic and non-Hispanic). We obtain this information using the R tidycensus package. To account for car ownership and commute method, we use the share of workers ages 16 and older who travel to work by car, truck, or van from census block group–level 2014–18 ACS five-year estimates, obtained from the National Historical Geographic Information System. When this variable is missing, we replace it with the share of households that own at least one vehicle. When that variable is also missing, we use the MSA average car commute percentage for the block group.

Geographic boundaries: We use census-defined boundaries for each of our four MSAs. Within each MSA, our unit of analysis is the census block group. We obtain the geographic boundaries for the block groups from the Census TIGER/Line Shapefiles using the R tigris package. We also use the census place boundaries to approximate the city boundaries as well as the census urbanized area boundary definition.

Population-weighted centroids: We use the population-weighted centroid for each block group as the start/end point for trips from or to that block group. We use the populated-weighted centroid so our start/end point better reflects where population is concentrated in the block group. We obtain the populated-weighted centroid data from the Missouri Census Data Center Geocorr 2014: Geographic Correspondence Engine. The data center uses 2010 Census population estimates at the census block group level to identify each population-weighted centroid.

Opportunity Data

We gather data on several types of opportunities that city and community leaders identified as most important to measuring equitable access to opportunity via transportation. This set of opportunities is certainly not exhaustive, and we hope to include other key businesses and services (e.g., grocery stores) and amenities (e.g., parks) in future analyses. Ultimately, we focus on access to jobs to validate an approach that could be applied to analyzing access to other types of opportunity.

Jobs: We use census block group–level total jobs by job location and worker location aggregated from 2017 census block–level LEHD Origin-Destination Employment Statistics (LODES) residence area characteristics and workplace area characteristics files. For both files, we calculate the total jobs in each block group excluding federal jobs, disaggregated by income level at the following cut points: monthly earnings of $1,250 or less, $1,251–$3,333, and $3,334 or more. We define low-wage jobs as those with monthly earnings of $3,333 or less, a combination of the bottom two categories.

Hospitals: We use open-source point-level data from an automated inventory of all US hospitals provided by ESRI and the US Geological Survey as of February 2020. We then aggregate hospitals by census block group. This dataset excludes clinics and does not take hospital quality or affordability into account.

Libraries: We use fiscal year 2018 data from the Institute of Museum and Library Services' Public Libraries Survey. We use the location of currently operating main and branch campuses of public library systems and exclude bookmobile locations.

Higher education: We use data on college campuses, as gathered from the US Department of Education's Postsecondary Education Participants System by the Center for American Progress. We include all colleges currently operating, and merge in contextual and performance data from the Department of Education's Integrated Postsecondary Education Data System and the College Scorecard.

Methodology

Analytical Approach

Our analysis aims to measure access to opportunity via transportation in each of our four MSAs. We consider four social and economic opportunities: jobs, higher education, libraries, and hospitals. We start by calculating the travel time between pairs of start and end points in each MSA via different modes of transit and times of day. We then use these travel times to calculate an adjusted duration measure that accounts for wait time at the end of a trip and traffic for car trips during peak hours. We use the adjusted duration to calculate the number of each type of opportunity that residents in each block group can access within 30 minutes. For jobs, we calculate access using a gravity model that accounts for competition for jobs. Finally, we use these access metrics to look at disparities in access to jobs for third-shift workers and peak-shift workers as well as disparities in access by race and ethnicity.

Analytical Process

STEP 1: DEFINE GEOGRAPHIC SCOPE

We restrict our examination of access to opportunity to the Lansing, Michigan; Seattle, Washington; Baltimore, Maryland; and Nashville, Tennessee MSAs. But we realize that people do not limit themselves by MSA boundaries when making decisions about accessing jobs, hospitals, libraries, or higher education institutions. Accordingly, we draw a buffer around each MSA border of 0.75 degrees (~45 miles) to conservatively capture opportunities that people living in the MSA might access outside the MSA. As we consider the number of opportunities accessible within a 30-minute trip for our access to opportunity metrics, this buffer contains the potential set of opportunities. For analyses considering larger distances, a larger buffer may be needed.

Because our access to jobs metric (described in step 7 on page 10) accounts for both the jobs accessible from the MSA and the competition for those jobs, we expand our buffer a further 0.75 degrees (~45 miles) to capture all workers that could access jobs located in the first buffer area. We then intersect the buffer areas with the population-weighted block group centroids from Missouri Census Data Center (see page 3) to identify the block groups that we include in our analysis for each MSA. Table 1 describes the resulting geographic areas.

TABLE 1

Geographic Areas Included in Analysis

MSA	Number of block groups in MSA	Number of block groups in buffer	States included in buffer
Baltimore	1,949	14,805	MD, DC, DE, VA, WV, NJ, PA
Lansing	371	8,911	MI, OH, IN
Nashville	1,022	4,846	TN, KY, MS, AL, GA, IL, IN
Seattle	2,484	5,075	WA, OR

STEP 2: GATHER ROAD AND TRANSIT DATA AND BUILD GRAPH

Next, we gather the relevant transit and road grid data required by OpenTripPlanner as described on pages 1–2. We use the dimensions of the geographic area identified in step 1 to clip the OpenStreetMap data for the United States to the relevant area for each MSA.

We then obtain the GTFS data for all transit agencies in the states included in the geographic area, using the process outlined in the transit data section (see page 2). The GTFS data are split into several component files that together provide the relevant information on the agencies, routes, stops, trips, and stop times that compose a transit system. In rare cases, these files have missing, inconsistent, or duplicate information that requires cleaning to prevent interference with the graph build or routing analysis. We try to preserve as much usable data for routing as possible by performing three data cleaning steps before building the graph:

- ***Impute missing data:*** We check for missing agency data in the agency file and replace missing fields with a set of default values. We also check for cases where the final stop time in a given route is missing from the stop-times file. In these cases, we fill the missing data with an interpolated guess for the arrival time and departure time of the final stop based on the time between the preceding stops on the same route. In cases where a route referenced in the stop-times file does not appear in the routes file, we add it to the routes file so OTP can use the stops.
- ***Remove invalid data:*** We remove values from the stop-times file that are associated with stop IDs that do not appear in the stops file and are thus invalid. As the stops file includes the location for each stop, a stop time associated with a stop ID that does not appear in that dataset cannot be linked to a geographic location and therefore cannot be used in routing.
- ***Remove duplicate trips and routes:*** We drop duplicate values in the trips file and routes file based on the unique identifiers for trips and routes.

Each of the above fixes occurs very rarely in the transit data we used for this analysis but is necessary to ensure that the graph builds properly.

We then use the processed road grid and transit data to build the representation of the transportation network (called a "graph") that OTP uses for routing. OTP allows users to configure the graph-building program by setting parameters; for our analysis, we set only the maximum transfer distance allowed between stops. Research shows that people are generally willing to walk a half-mile to a transit stop, or, more precisely, a half-mile for rail stops and a quarter-mile for bus stops.[2] Because OTP only allows a single maximum transfer distance for all transit modes, we set the maximum transfer distance for the graph build to a half-mile (804.672 meters).

STEP 3: IDENTIFY START AND END POINTS FOR ROUTING

We then identify the relevant pairs of start and end points for routing. The set of all possible pairs can be found by matching each population-weighted centroid in the set of block groups identified for each MSA in step 1 with all other centroids in that set. To reduce the complexity of our analysis, we impose rules to determine the subset of pairs that we will use for routing:

- ***Car routes:*** We take the full set of centroids as the possible start points and the set of centroids within the first buffer drawn around the MSA boundary (~45 miles around the MSA) as the end points. We do this because we only use the points in the larger buffer to calculate competition for jobs in the smaller buffer based on how many workers can access those jobs (see step 7 for more detail). Therefore, the points in the larger buffer but not the smaller buffer are only needed as start points, not end points. In addition, we only consider pairs of start and end points that are within 0.75 degrees (~45 miles) of each other.

- ***Transit routes:*** We apply the two rules for car trips above. Then, we only consider pairs of start and end points where both points are within 0.04 degrees (~2.5 miles) of any transit stop included in the analysis. As described further in step 4 below, we set the maximum walk distance for a single leg in a route to a half-mile; this restriction is very conservative to avoid excluding any potentially valid trips.

Based on these rules, we identify the following transit and car route pairs for each of our four MSAs:

TABLE 2

Start and End Pairs for Routing by MSA

MSA	Number of transit pairs	Number of car pairs
Baltimore	21,357,441	30,721,275
Lansing	648,667	5,975,717
Nashville	449,541	1,597,004
Seattle	6,427,746	7,509,536

STEP 4: CALCULATE TRAVEL TIMES

We use the OTP routing API to calculate the route, travel time, and travel duration between each start-end pair. OTP offers numerous parameters that can be used to configure routing requests. We set the following values in our analysis:

- ***Number of itineraries:*** This parameter controls the number of itineraries returned by OTP. By default, OTP optimizes the route to minimize the travel duration of the trip. We set this parameter to 1, which means we obtain only the itinerary that represents the fastest trip between the start and end points.

- ***Maximum transfer walk distance:*** This parameter controls the maximum allowed walk distance when transferring between transit legs. We set this parameter to a half-mile (804.672 meters).

- ***Maximum walk distance:*** This parameter controls the maximum allowed walk distance from the start point to the first transit stop and from the last transit stop to the end point. We set this parameter to a half-mile (804.672 meters).

- ***Transfer slack:*** This parameter sets a global minimum time that must pass between exiting one vehicle and boarding another in a transfer. Using this parameter prevents trips with very close transfers; a rider would likely choose not to attempt these trips because of the risk of missing the transfer. We set this parameter to five minutes (300 seconds).

- ***Arrive by:*** This parameter determines whether the trip should depart at the provided time or arrive by the provided time. We set this parameter to "true" so trips arrive by the time provided.

OTP also offers a parameter which, when set to true, requires the trip be wheelchair accessible. When true, OTP only uses the routes and stops indicated as wheelchair accessible in the GTFS data. Unfortunately, the wheelchair accessibility fields are optional in the GTFS data standard and are missing for many transit agencies in our analysis. We are therefore unable to assess disparities in access to opportunity by this parameter. With access to complete data on wheelchair accessibility, this would be an important area for future analysis.

In addition to the parameters discussed above, all routing requests require that the user define the eligible set of transportation modes, the trip date, and the trip time. For our analysis, we model both trips where the traveler travels by car (available modes = walk and car) and trips where the traveler travels by transit (available modes = walk and transit). We also model trips where the traveler needs to arrive at a "peak" time as well as a time reflective of the start time of third-shift jobs. For our analysis, we consider peak travel to cover arrivals between 7:30 a.m. and 8:30 a.m. and third shift to cover arrivals between 10:30 p.m. and 11:30 p.m.

We restrict our analysis to January 13–17, 2020, to model weekday travel under the transportation system before COVID-19 service reductions. For each city, we calculate one set of car routes (as car travel time is not affected by arrival date/time) and multiple sets of transit routes for peak and third-shift travel, taking the average to calculate our travel time measures. We randomly select three days and times (used as days offset from the first possible date and minutes offset from the first possible time) for our transit routes. The full set of routes calculated for each metro is provided in table 3.

TABLE 3

Routes Calculated

Trip type	Modes	Arrival date and time
Car	Walk, car	8:04 a.m., 2020-01-16
Peak transit	Walk, transit	8:04 a.m., 2020-01-16
		7:52 a.m., 2020-01-17
		8:12 a.m., 2020-01-13
Third-shift transit	Walk, transit	11:04 p.m., 2020-01-16
		10:52 p.m., 2020-01-17
		11:12 p.m., 2020-01-13

In many cases when using transit, in order to arrive by the specified arrival time a traveler has to arrive *before* that time—in some cases, considerably before. We calculate a measure of adjusted duration that accounts for this "wait time" at the destination. We calculate wait time by finding the number of minutes before the arrival time that the passenger arrives, multiplying that number by 0.5, and adding that quantity to the trip duration to calculate adjusted duration.[3]

For our analysis, we set an adjusted duration of 0 to access opportunities within the starting point block group.

STEP 5: IMPUTE MISSING ROUTES

In rare cases, OTP is unable to identify a driving route between two points. This may be because no road exists (such as Vashon-Maury Island in the Seattle MSA) or because the population-weighted centroid cannot be snapped to the graph for routing (e.g., if the centroid happens to fall in a body of water or a forest). For our analysis, we decide to impute all missing car routes by taking the average of the adjusted durations between all block groups bordering the start/end block group and the end/start block group for which the route cannot be calculated.

For transit routes, in many cases the route cannot be calculated because it is not possible to travel between the start and end point given our maximum walk distance limits. For transit, we only impute routes where OTP can calculate no driving routes for a given start or end point, on the logic that the same issues preventing the centroid point from being snapped to the graph would prevent any transit routes from being calculated. We use the same approach of taking the average of the adjusted durations between all block groups bordering the start/end block group and the end/start block group for which the transit route cannot be calculated.

In a small number of car routes and most transit routes where imputation is attempted, the data cannot be imputed because the route can also not be found for all bordering block groups. In Seattle, percentage of car routes where imputation is successful is lower owing to the presence of islands in the MSA. The table below summarizes statistics on imputation.

TABLE 4

Imputation Summary

MSA	# routes–car	# imputation attempted–car	% imputation successful–car	# routes–transit	# imputation attempted–transit	% imputation successful–transit
Baltimore	30,721,275	585,942	99.2	21,357,441	277,141	46.0
Lansing	5,975,717	95,490	99.6	648,667	18,986	15.1
Nashville	1,597,004	18,859	99.7	449,541	4,290	24.6
Seattle	7,509,536	232,918	74.8	6,427,746	85,049	50.1

In future analysis, we may explore approaches that more intelligently select the start and end points to avoid issues with snapping the point to the graph. This could include setting the start/end point in each block group as the point closest to the population-weighted centroid that falls on a road.

STEP 6: APPLY TRAFFIC ADJUSTMENT FOR PEAK DRIVING TIMES

We select block groups within the busiest commuter destinations identified in a metro region by INRIX as a proxy for areas in a metro region that are most likely to experience disproportionate congestion. Next, we create a traffic multiplier using a ratio of the off-peak traffic speed to the peak traffic speed. This multiplier is applied as a congestion weight for all trips via car taken during high-commute times.

STEP 7: CALCULATE ACCESS TO JOBS USING GRAVITY MODEL

To calculate access to jobs and spatial mismatch (the mismatch between where low-income households reside and where job opportunities exist), we follow Shen (2001)[4] but modify the methods to calculate job access for all jobs and the low wage labor force rather than just job openings and job seekers, and then convert this measure into a measure of spatial mismatch for low wage workers.

To do so, we first calculate the following equations:

$$A_i^{auto} = \sum_j \frac{J_j \times f(C_{ij}^{auto})}{\sum_j [\alpha_k P_k \times f\left(C_{kj}^{auto}\right) + (1-\alpha_k) P_k \times f(C_{kj}^{tran})]}$$

$$A_i^{tran} = \sum_j \frac{J_j \times f(C_{ij}^{tran})}{\sum_j [\alpha_k P_k \times f\left(C_{kj}^{auto}\right) + (1-\alpha_k) P_k \times f(C_{kj}^{tran})]}$$

where A_i^{auto} and A_i^{tran} are accessibility scores for low wage job seekers who are automobile drivers and public transit riders, respectively, living in block group *i*; J_j is the number of low-wage jobs in block group *j*; $f(C_{ij}^{auto})$ and $f(C_{ij}^{tran})$ are impedance functions for low-wage automobile drivers and public transit riders traveling between *i* and *j*, where the impedance function is a travel time threshold function that equals 1 when the adjusted car travel duration and the adjusted transit travel duration, respectively, are less than 30 minutes; α_k is the percentage of workers within the block group who travel to work via car (or, when missing, the percentage of households in the block group who own at least one car) times a multiplier to reduce this number so it better represents car commuting for low-wage workers; P_k is the low-wage labor force (the number of employed low-wage workers plus the number of unemployed residents) living in block group *k*; and $f\left(C_{kj}^{auto}\right)$ and $f\left(C_{kj}^{tran}\right)$ are impedance functions for automobile drivers and public transit riders traveling between block groups *k* and *j*. In essence, these equations allow us to sum up all the jobs accessible to a low-wage worker in a block group via both driving and public transit (weighted by access to these different transit modes), divided by the competition for those jobs, or the number of workers who live within a reasonable commuting time of each job.

We then combine these equations to calculate an overall job accessibility score for low-wage workers in each block group:

$$A_i^G = \alpha_i A_i^{auto} + (1 - \alpha_i) A_i^{tran}$$

where A_i^G is the general accessibility score for people living in residential zone *i*.

We then standardize A_i^G so the maximum is 1 and the minimum is 0 and each value within it is proportional to the position relative to the maximum and minimum. We do the same for the low-wage labor force in each block group. We then subtract the standardized accessibility score from 1, so higher values denote lower accessibility and lower values denote higher accessibility, and multiply it by the standardized low-wage labor force (low-wage workers plus unemployed workers) to get an overall spatial mismatch number. This final number tells us which block groups have concentrations of low-wage workers with low job accessibility.

The multiplier for car commuting of 0.898 is calculated by dividing the national car-commuting rate for people with incomes below the poverty level by the national car-commuting rate for *all people*. This multiplier helps us better approximate car-commuting percentages for low-wage workers since the car-commuting measures for the block group from the census are for all residents, not just low-wage workers.

We measure the low-wage labor force by combining low-wage workers with unemployed residents since we want to account for both job seekers and currently employed low-wage workers. While it would be ideal to include only unemployed low-wage workers in this equation, we do not that have that information available to us. Therefore, we work under the assumption that all unemployed workers are in the pool of job seekers for low-wage jobs. This likely overstates the number of job seekers, but not in a way that is correlated with job locations and therefore unlikely to bias our estimates.

Unlike Shen, we include all jobs, not only job openings, in our analysis. We want to understand all job options for low-wage workers since many are likely to have accepted jobs that are farther from their homes than what is ideal; looking only at job openings would overlook this challenge. Additionally, since we are interested in informing transportation policy, we would like to know the location of all low-wage workers, not only the ones that are unemployed.

STEP 8: ANALYZE DISPARITIES IN ACCESS TO JOBS FOR PEAK AND THIRD-SHIFT TRANSIT

We compare the transit accessibility scores (A_i^{tran}) calculated for each block group in step 7 using the adjusted travel times for peak trips and third-shift trips to assess the disparity in access to jobs for third-shift workers versus peak-hours workers. We restrict our focus to transit accessibility for this comparison because access to jobs via car will not vary for third-shift and peak workers. One limitation to this approach is that the LODES data do not enable us to differentiate between peak-hour jobs and third-shift jobs; therefore, we assume that the spatial distribution of third-shift jobs follows the overall spatial distribution of jobs. Similarly, we are unable to differentiate third-shift workers from all workers in the ACS data. We think this measure will help planners and advocates identify the areas where a given third-shift worker would face the greatest reduction in accessibility relative to peak workers. In future analysis, we would try to identify the number of third-shift workers in each block group to better target additional transit investments.

STEP 9: CALCULATE RACIAL DISPARITY IN ACCESS TO JOBS

We analyze racial and ethnic disparities in access to jobs by comparing the racial and ethnic distribution of residents in neighborhoods with the highest rates of spatial mismatch in the urbanized area to the distribution of residents by race and ethnicity in the urbanized area as a whole. The calculation can be broken down into three steps:

1. Calculate the neighborhoods that have the highest rates of spatial mismatch, using the equations above.
2. Describe the distribution of residents in these neighborhoods by race and ethnicity. The racial and ethnic groups are Asian (both Hispanic and non-Hispanic); Black or African American (both Hispanic and non-Hispanic); Hispanic or Latino; non-Hispanic white; and other race (both Hispanic and non-Hispanic).
3. Compare this distribution to the distribution of residents by race and ethnicity for the urbanized area as a whole.

Conclusion

This analysis aims to respond to demand for better definitions and data to measure transportation equity that we heard from city and community leaders throughout the country. We hope community organizations, city officials, transit agencies, and other local leaders can use our analysis to inform conversations about how to improve access to opportunity through transportation. Our analysis demonstrates that using this approach highlights neighborhood disparities and can inform where to target investments. In future analyses, we hope to expand this approach to explore access to opportunity more holistically, by incorporating other opportunities, such as grocery stores and parks, and other measures of transportation quality, such as safety and reliability. We also hope to measure access for people with differing levels of ability, such as people in wheelchairs, for which the data in the GTFS are mostly missing. A further discussion of the barriers to equitable transportation in each of our four case-study MSAs can be found in the report along with some of the analytical results, which can be explored in further detail in the feature (https://www.urban.org/features/unequal-commute).

Notes

1 A census block group is a subdivision of a census tract that generally contains between 600 and 3,000 people. For more information, see the Census Bureau's online glossary, https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_4.

2 See Eric Jaffe, "What Does Living 'Close' to Transit Really Mean?," Bloomberg CityLab, January 12, 2015, https://www.bloomberg.com/news/articles/2015-01-12/what-does-living-close-to-transit-really-mean.

3 Because OTP does not include the wait time between the trip end time and the requested arrival time as part of the trip duration, in some cases OTP's behavior to optimize for the trip with the shortest duration will yield a trip with a longer wait time and adjusted duration than another potential trip with a longer duration that arrives closer to the requested arrival time. Unfortunately, there is no effective way to configure OTP to account for wait time. The best possible option is to set the maxHours parameter in combination with useRequestedDateTimeInMaxHours, which establishes a maximum allowed trip time and includes the wait time. In our tests of this approach, we found that the impact was fairly evenly split between increasing and decreasing adjusted duration. Based on these tests, we decided not to use these parameters, but we acknowledge this limitation.

4 See Qing Shen, "A Spatial Analysis of Job Openings and Access in a U.S. Metropolitan Area," *Journal of the American Planning Association* 67, no. 1 (2001): 53–68, https://doi.org/10.1080/01944360108976355.

Acknowledgments

This research was funded in collaboration with the Mastercard Center for Inclusive Growth. Thank you in particular to Daniel Barker and Ali Schmidt-Fellner who served as thought partners throughout this process. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

Thank you to the community and city leaders that provided input into this research through interviews, a convening, and review of the report.

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.