**Introduction**

Cities support dense social, economic, physical and intellectual networks, which facilitate the frequent and complex interactions needed to drive high levels of productivity, efficiency and innovation (Bettencourt & West, 2010). However, inequality also tends to be greatest in cities, as people embedded in different network locations experience different levels of connectivity to opportunities and resources, which in turn leads to differentiated outcomes among individuals. This can be tangibly seen in how workers who live in different neighbourhoods experience unequal levels of accessibility to jobs (defined as the ability to reach suitable job opportunities within a reasonable travel time or cost via the transport network) (Hansen, 1959), which constrains the job options available to them and shapes their eventual employment outcomes. In some locations, the situation may be more complicated as workers experience both lower accessibility and higher social disadvantage (such as low education, poverty or physical disabilities) (El-Geneidy, et al., 2016).

While the unevenness of job accessibility in cities has been widely studied in the literature, most extant analyses under-explore an aspect of urban transportation that is particularly salient in urban dwellers’ actual experience: the day-to-day variability in the time needed to carry out their usual trips (Katz & Quealy, 2019), which may result from congestion, varying service schedules and failures of parts of the network. However, recent improvements in the availability of real-time trip planning data have made it possible to analyse how reliability in trip times may also be unevenly distributed.

In this study, I will analyse the distribution of accessibility to jobs and variability in travel times via public transport in the Greater London Authority area, with the aim of identifying areas that are burdened with high social disadvantage, poor accessibility to jobs and/or low travel time reliability. In addition to demonstrating how accessibility analyses can productively incorporate data on travel time variability, the findings from this study can also be used to target improvements in accessibility to jobs and the reliability of travel times via public transport to most benefit socially disadvantaged areas.

**Literature Review**

Many studies in urban development and transport geography examine the distribution of accessibility to opportunities in cities. Deboosere & El-Geneidy (2018) identify two main types: those that analyse the “horizontal” *inequality* of accessibility levels across different areas of a city (Levine, 1998; Ford, et al., 2015; Lucas, et al., 2015), and those that analyse the “vertical” *inequity* of the distribution of accessibility levels among population segments (Shen, 1998; Sanchez, 1999; El-Geneidy, et al., 2016; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Deboosere & El-Geneidy, 2018). This corpus clearly shows that it is normal for vast differences in accessibility levels to exist within a city. Also, sometimes disadvantaged residents are more likely to live in areas with poorer accessibility; this is more likely in poorer cities with less extensive public transport systems.

In the context of London, both the UK government (Social Exclusion Unit, 2003) and the Greater London Authority (GLA) (2018) recognise that the spatial distribution of accessibility is uneven. In particular, low accessibility may significantly hinder disadvantaged residents’ access to jobs, healthcare, services, education and leisure, thereby constraining their options, increasing travel burdens, lowering quality of life and entrenching social disadvantage. To make the situation fairer, authorities recommend targeting accessibility improvements to better connect areas with more disadvantaged residents to the opportunities they need.

Researchers have also developed many metrics to measure accessibility, from the perspective of the links in the transport network, the individual traveller, the choices facing travellers or the different areas of the city (Geurs & Wee, 2004). This last “location-based” type is widely used for city-level analyses (Shen, 1998; Sanchez, 1999; Kotavaara, et al., 2012; El-Geneidy, et al., 2016; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Deboosere & El-Geneidy, 2018). One simple location-based cumulative potential measure (Geurs & Wee, 2004, p. 133) expresses accessibility as the total number of opportunities (in the case of this study, jobs) that can potentially be reached from each area:

where is the accessibility to jobs for area i, is the number of jobs in area j, is the cost (distance, time and/or fiscal) of travelling from area i to area j, and is an impedance function which decreases the extent to which contributes to as increases.

While measures accessibility as a function of the distributions of jobs and travel costs and is simple to calculate and interpret, a significant drawback is that it does not account for the distribution of the workers who seek jobs (Shen, 1998, p. 348; Hu & Downs, 2019, p. 279); two neighbourhoods that can both access the jobs in the same areas would register the same value, but intuitively the neighbourhood that has more workers should have a lower accessibility level as the available job supply is shared among a larger demand. To address this shortcoming, Shen (1998) proposed an adapted measure that accounts for the number of workers in all areas k (including area i) that can access j and distributes the available jobs from area j to the areas k proportionally by their populations, thus accounting for “competition effects” between workers living in different areas for the same jobs:

where is the number of workers in area k and is the travel cost between areas k and j. Thus, accessibility is a function of 3 components: the distribution of workers (demand), jobs (supply) and travel costs.

To assess the extent to which socially disadvantaged populations are more likely to experience poorer accessibility, many researchers use location-based measures of social deprivation, such as neighbourhood average income (Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018) or an index combining multiple indicators of deprivation (El-Geneidy, et al., 2016; Deboosere & El-Geneidy, 2018). In the UK, the Department for Housing, Communities and Local Government have developed the Indices of Multiple Deprivation (IMD) to reflect the extent to which small areas (down to the Lower-level Super Output Area) are burdened by deprivation across a wide range of domains, including income, employment, health, education, crime, housing and the living environment (Department for Housing, Communities and Local Government, 2015). In this study, I use the IMD data to capture the relative deprivation of different Middle-layer Super Output Areas (MSOAs).

Researchers have also increasingly recognised that the spatial distributions of the supply and demand for opportunities and travel costs can all vary significantly over time (Tenkanen, et al., 2016; Järv, et al., 2018; Hu & Downs, 2019). Using geolocated social media or mobile phone data, workplace opening times and public transport schedules, researchers have built time series of these 3 variables over a typical day and calculated accessibility measurements for each area at each time point. Undoubtedly, the increasing availability of such data will drive further methodological innovations in this direction. However, the extant studies calculate accessibility based on “typical” (mean or ideal) values of these variables at each time point, without accounting for the stochastic distribution of these variables around their typical values. Thus, extant analyses do not account for the uncertainty that actually exists around accessibility levels.

**Research Questions**

My analysis will concentrate on accessibility to jobs for working-age residents. While not all trips involve commuting to work, work commutes are a primary driver of travel demand, particularly during peak periods. In addition, most of the opportunities that people travel to seek (such as education, healthcare or shopping) also provide jobs, so accessibility to jobs will capture accessibility to opportunities in general to some extent (Deboosere & El-Geneidy, 2018, p. 56).

This study will address two questions regarding the distribution of accessibility to jobs via non-private transport in London:

1. To what extent are accessibility to jobs and the variability of this accessibility *unequally* distributedacross London’s working population?
2. To what extent are poor accessibility to jobs and higher variability of this accessibility disproportionately likely to impact disadvantaged segments of the working population in an *inequitable* way?

**Data Collection Strategies**

Location-based cumulative potential accessibility measures require the city to be divided into discrete areas. Some studies divide the city in question into a grid of equally-sized squares, in order to minimise distortions in the travel cost calculations due to differences in size and shape between areas (Kotavaara, et al., 2012; Tenkanen, et al., 2016; Järv, et al., 2018; Hu & Downs, 2019). However, aggregating spatial data to these squares requires very fine-grained data (which is unlikely to be publicly available) or additional modelling and assumptions to reproject available data to the squares, at the risk of distorting the data. To avoid this, I follow the majority of extant studies (Shen, 1998; Sanchez, 1999; El-Geneidy, et al., 2016; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Deboosere & El-Geneidy, 2018) in dividing up my study area into pre-existing spatial units, specifically the 983 Middle-Layer Super Output Areas (MSOAs) within the GLA boundary. These are spatial data reporting units created by the UK’s Office of National Statistics, and thus provide the ability to seamlessly incorporate a wealth of spatial demographic and economic data from open government databases into my analysis.

A key dataset required for this study is the inter-area travel cost matrix. Most extant studies obtain this by compiling a multimodal transport network dataset, incorporating network link speeds, transport service schedules or assumed service frequencies, implementing a routing algorithm, and then finding the shortest route or routes between every pair of areas and calculating the associated costs. See Ford et al (2015, pp. 129-130) for a step-by-step guide. A significant benefit of this approach is that researchers have full knowledge of the underlying network data and routing assumptions used to generate the travel costs. However, the approach requires an extensive data compilation and cleaning effort, is sensitive to the researchers’ choice of routing algorithm, and does not incorporate real-time data on transport service performance.

In contrast, I will obtain inter-MSOA travel data from Transport for London’s (TfL’s) Journey Planner (2019). This is a data service that provides the public with guidance on how to travel between any two points in London by non-private transport (i.e. by walking, cycling or motorised public transport like buses or trains, but not by car or taxi), including information on the fare, time, distance and route for up to 6 travel options. This allows me to obtain travel cost data without having to build my own transport network dataset for London, and the results reflect the network and routing assumptions actually used by TfL to assist real travellers in London. Also, crucially, the Journey Planner accounts for real-time disruptions on the transport network; for instance, if a certain rail line is shut down due to a fault, the Journey Planner will not provide options that require travel on that rail line. Thus, this data source is uniquely capable of providing the information on travel time variability due to real-time fluctuations in transport service performance that underpins the methodological innovation in this study.

TfL maintains an application processing interface (API) for the Journey Planner (Transport for London, 2019), which allows for the automated bulk download of travel option data. However, to ensure that the service remains available to all users, each user can only submit a maximum of 300 queries per minute to the API. I will use the API to compile the following databases:

1. The travel options for trips between every pair of MSOAs ending at 0830 or 1330 on 28 Jun 2019. This will constitute the “baseline” travel option data for all trips for the weekday AM peak and off-peak periods, assuming all services run as planned. Due to the API request limit, data collection for all 983 x 983 trips will occur over 2 months.
2. The travel options for trips beginning between 0600 and 0800 or 1100 and 1300 every weekday for 50 days. This will constitute “real-time” travel option data during the weekday AM peak and off-peak periods, reflecting actual day-to-day variability in transport service performance. Because of the API request limit, I will only be able to gather data on trips starting from all 983 MSOAs but ending at 7 MSOAs that correspond to 7 key transport nodes evenly spaced across London.

The methodological benefits of using the Journey Planner data are compelling enough to justify its use in this study, but this does impose some constraints on this study’s design. This is because the Journey Planner provides information only for trips that do not involve cars or taxis, and is most reliable for trips within the GLA boundary. As such, my analysis will concentrate only on areas within the GLA boundary and on non-private transport services. TfL data (2018) also show that only 37% of trips within the GLA in 2017 used cars or taxis, while 80% of trips within the GLA boundary were by GLA residents. Nonetheless, I recognise that my results will be distorted by the fact that they omit accessibility by car and taxi, and do not consider areas outside the GLA boundary that can be accessed from London.

**Data Analysis Strategies**

The first task of data analysis is to estimate the variability of travel times for each trip due to fluctuations in transport service performance. For the 983 x 7 trips where real-time travel data is collected, I will simply calculate the standard deviation of travel time for the fastest option over the 50 observed days. However, for the 983 x (983 – 7) trips where only baseline travel option data is available, I will estimate their standard deviations of travel times for their fastest options based on their baseline data. This requires establishing how the observed standard deviation of travel time correlates with the baseline data for the 983 x 7 trips, probably through regression-based techniques like locally-weighted regression or generalised additive models.

Next, I will calculate and for each MSOA. This involves a choice of impedance function . Many studies use a power or exponential function, such that decreases smoothly as increases, resulting in a “gravity”-based accessibility measure where increases as the “mass” increases and the “distance” decreases (Shen, 1998; Sanchez, 1999; Kotavaara, et al., 2012; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018; Hu & Downs, 2019). This more accurately reflects the tendency for people to prefer nearer locations with more opportunities to further locations with fewer opportunities, but the resulting raw accessibility values are scores that have no intrinsic meaning and can only be interpreted relative to other scores (El-Geneidy, et al., 2016, p. 304). In contrast, I will follow El-Geneidy et al (2016), Tenkanen et al (2016), Deboosere et al (2018) and Järv et al (2018) to use what Geurs and van Wee (2004) call a “contour” measure, in which:

where X is some travel cost cut-off. This is undoubtedly a distortion of reality, as it treats all locations where as if they were equally attractive and all locations where as if they were totally unattractive. However, the distortion is relatively straightforward and does not require further assumptions and modelling, and it has the benefit of maintaining the raw accessibility values as simple counts and ratios that can be interpreted directly (Deboosere & El-Geneidy, 2018, p. 55). At any rate, El-Geneidy et al (2016, p. 303) found that contour and gravity-based measures are highly correlated and can be interchangeably used. For the purpose of this study, I will set X to 45min, following TfL’s (2018, p. 224) own analyses.

In order to assess the extent to which accessibility is unequally distributed across London, I will calculate for each MSOA under the following conditions:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Measure | Workers | Jobs | Time Period | Incorporates Reliability? |
| A1 | All | All | AM Peak | No |
| A2 | All | All | Off Peak | No |
| A3 | All | All | AM Peak | Yes |
| A4 | All | All | Off Peak | Yes |

Measures A1 and A2 will indicate the number of jobs that are accessible within 45min travel from each MSOA assuming all services run as planned, during the weekday AM peak and off-peak respectively. The ratio of A2/A1 for each MSOA indicates the amount of accessibility lost due to less-frequent scheduled services during the off-peak, and is similar to the measurements in recent studies of temporal variability in typical levels of accessibility (Tenkanen, et al., 2016; Järv, et al., 2018; Hu & Downs, 2019).

Measures A3 and A4 will indicate the number of jobs that are accessible within 45min travel from each MSOA at least 95% of the time. That is, for a job to be considered accessible from an MSOA, not only must the expected travel time to the job be less than 45min, but one also must be able to make the trip to the job within 45min at least 19 out of 20 times. Assuming that the travel times for the fastest option for each trip are Normally distributed, calculating Measures A3 and A4 will involve first adding 1.6 standard deviations of travel time to the expected travel time for each trip. Then, the ratio A3/A1 will indicate the amount of accessibility lost due to unreliability in travel times, and the ratio A4/A1 will indicate the amount of accessibility lost due to both less-frequent off-peak services and travel time unreliability. This will also indicate the extent to which accessibility measurements that do not adjust for travel time unreliability may over-estimate accessibility levels.

To assess the extent to which the unequal distribution of accessibility among different areas is equitable for the working population of London, I will calculate for each MSOA for all workers, low-income workers and non-low-income workers matched only to their respective suitable jobs. One important feature of is that because it accounts for competition effects in allocating accessible jobs to workers, each job (or part thereof) is allocated exactly once to a worker (Shen, 1998, pp. 349, 363-364). Thus, unlike in , there is no double-counting of jobs or workers, and jobs are treated as rival goods to be distributed among workers. This then allows us to proceed to assess the equity of the distribution using well-established tools for measuring distributional inequality between groups, such as Lorenz curves (Lucas, et al., 2015).

will be calculated under the following conditions:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Measure | Workers | Jobs | Time Period | Incorporates Reliability? |
| R1 | All | All | AM Peak | No |
| R2 | Non-Low-Income | Non-Low-Wage | AM Peak | No |
| R3 | Low-Income | Low-Wage | AM Peak | No |
| R4 | All | All | AM Peak | Yes |
| R5 | Non-Low-Income | Non-Low-Wage | AM Peak | Yes |
| R6 | Low-Income | Low-Wage | AM Peak | Yes |

Measures R1, R2 and R3 indicate the number of accessible jobs allocated to each worker after accounting for competition effects and assuming all AM peak services run as planned. For each MSOA, we can then multiply this ratio by the relevant number of workers to obtain the number of jobs allocated to all workers within it. After sorting all MSOAs by the number of jobs allocated to their workers, we can then plot Lorenz curves for the cumulative distribution of jobs to workers. The further a Lorenz curve is from the diagonal, the more unequal the distribution is. If the Lorenz curve for low-income workers is further from the diagonal than the curve for non-low-income workers, then we can say that the distribution of accessibility among non-low-income workers is less unequal than among low-income workers, and thus there is an inequitable distribution of accessibility favouring non-low-income workers over low-income workers. A similar analysis can be carried out for Measures R4, R5 and R6, which will additionally adjust for travel time fluctuations and thus will reflect the redistributive effect of travel time uncertainty on how jobs are allocated to workers. If the Lorenz curves become further from the diagonal after accounting for travel time fluctuations, then this indicates that travel time fluctuations have increase the inequality in the distribution of jobs among workers.

**Anticipated Outcomes**

This study will measure the extent to which accessibility to jobs by non-private transport varies across London, in the same vein as extant accessibility studies. This is likely to be of interest to planners and policymakers in London. For instance, urban planners can consider targeting housing growth to locations where there is more than 1 job available per worker, while transport planners can target transport infrastructure investments to better connect areas where there is fewer than 1 low-wage job per low-income worker to areas with a surplus of such jobs.

This study is also expected to provide insights into accessibility from a novel angle, in terms of how day-to-day fluctuations in transport service reliability can affect accessibility to jobs. The results could identify locations where accessibility for low-income workers is disproportionately affected by service unreliability, and transport planners can consider targeting operational improvements to the existing services there. Also, for academics, the results could provide an indication of the extent to which analyses that do not account for unreliability in transport services may over-estimate accessibility levels. The methods used in this study to harness real-time travel information to obtain a distribution of travel times for each trip can be the basis for other researchers to propose more innovative ways to investigate how accessibility to opportunities and actual travel behaviours are impacted by uncertainty in travel times.

**Limitations and Further Research**

One of the most significant limitations of this study is that the underlying network dataset and routing algorithm used to produce the travel costs for accessibility calculations remain unknown to me. As such, I must use TfL’s data at face value and am not able to judge whether TfL’s data and routing assumptions reflect actual behaviour in the transport system. In addition, researchers that have the underlying network data are able to assess the impacts of hypothetical scenarios (such as the construction of new transport infrastructure or changes to fares) on accessibility, by adjusting the network and recomputing the travel costs (Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018), but this is something that I am unable to do.

Another significant limitation is that this study does not incorporate accessibility by car or taxi, as there is no data source that can provide real-time travel data for these modes in London for free, automated bulk download. However, these modes undoubtedly provide significant accessibility, particularly for people who live in outer London, who do not travel through central London and who live in areas that are not well served by public transport (Transport for London, 2018, pp. 41-43). Thus, the accessibility levels estimated in this study will underestimate the actual levels of accessibility available to residents who have the option to use cars or taxis. If a researcher could secure real-time car routing data (for example, from apps like Waze and Gett), then a follow-up study could carry out similar analyses incorporating uncertainty in car and taxi travel times.

As mentioned above, accessibility is a product of the distributions of the supply and demand for opportunities and travel costs. In this study, temporal variability in non-private transport services are considered, but not temporal variability in the distribution of workers and jobs. A follow-up study could address this by using the methodology in Tenkanen et al (2016) and Järv et al (2018) to estimate the expected number of residents and the opportunities that they seek in different locations and at different time points. However, it will also be advisable to account for the day-to-day variability in these distributions, to ensure accessibility measurements holistically capture the extent of uncertainty around the distributions of residents, opportunities and travel times.

Policymakers who wish to use findings from accessibility studies like this one to guide decisions about urban development and transport improvements should also be aware that such studies are not designed to measure the causal effects of accessibility to opportunities on resident outcomes. While they can indicate the correlation between lower accessibility and poorer outcomes, they only measure one factor out of a constellation affecting the opportunities available to people and how they choose among the available options to maximise their individual objectives. And while taking such results into account when making policy decisions arguably makes the decisions more professionally defensible and ethically sound, one should not expect that improvements in accessibility will always lead to improvements in outcomes, particularly in the short term and particularly in the presence of a complex web of economic, human capital, environmental and cultural factors that shape the distributions of residents and opportunities and how they connect to each other in reality.

# **References**

Bettencourt, L. & West, G., 2010. A unified theory of urban living. *Nature,* 10, Volume 467, pp. 912-913.

Deboosere, R. & El-Geneidy, A., 2018. Evaluating equity and accessibility to jobs by public transport across Canada. *Journal of Transport Geography,* 12, Volume 73, pp. 54-63.

El-Geneidy, A. et al., 2016. The cost of equity: assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice,* 9, Volume 91, pp. 302-316.

Ford, A., Barr, S., Dawson, R. & James, P., 2015. Transport accessibility analysis using GIS: assessing sustainable transport in London. *ISPRS International Journal of Geo-Information,* 1, Volume 4, pp. 124-149.

Geurs, K. T. & Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography,* 6, Volume 12, pp. 127-140.

Greater London Authority, 2018. *Mayor's transport strategy: March 2018.* London: Greater London Authority.

Guzman, L. A. & Oviedo, D., 2018. Accessibility, affordability and equity: assessing `pro-poor' public transport subsidies in Bogotá. *Transport Policy,* 9, Volume 68, pp. 37-51.

Guzman, L. A., Oviedo, D. & Rivera, C., 2017. Assessing equity in transport accessibility to work and study: the Bogotá region. *Journal of Transport Geography,* 1, Volume 58, pp. 236-246.

Hansen, W. G., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners,* 5, Volume 25, pp. 73-76.

Hu, Y. & Downs, J., 2019. Measuring and visualizing place-based space-time job accessibility. *Journal of Transport Geography,* 1, Volume 74, pp. 278-288.

Järv, O. et al., 2018. Dynamic cities: location-based accessibility modelling as a function of time. *Applied Geography.*

Kotavaara, O., Antikainen, H., Marmion, M. & Rusanen, J., 2012. Scale in the effect of accessibility on population change: GIS and a statistical approach to road, air and rail accessibility in Finland, 1990-2008. *The Geographical Journal,* 3, Volume 178, pp. 366-382.

Levine, J., 1998. Rethinking accessibility and jobs-housing balance. *Journal of the American Planning Association,* 6, Volume 64, pp. 133-149.

Lucas, K., 2012. Transport and social exclusion: where are we now?. *Transport Policy,* 3, Volume 20, pp. 105-113.

Lucas, K., Wee, B. & Maat, K., 2015. A method to evaluate equitable accessibility: combining ethical theories and accessibility-based approaches. *Transportation,* 3, Volume 43, pp. 473-490.

Neihaus, M., Galilea, P. & Hurtubia, R., 2016. Accessibility and equity: an approach for wider transport project assessment in Chile. *Research in Transportation Economics.*

Sanchez, T. W., 1999. The connection between public transit and employment: the cases of Portland and Atlanta. *Journal of the American Planning Association,* 9, Volume 65, pp. 284-296.

Shen, Q., 1998. Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and Planning B: Planning and Design,* Volume 25, pp. 345-365.

Social Exclusion Unit, 2003. *Making the connections: final report on transport and social exclusion.* London: Office of the Deputy Prime Minister.

Tenkanen, H. et al., 2016. Health research needs more comprehensive accessibility measures: integrating time and transport modes from open data. *International Journal of Health Geographics,* 15(1).

Transport for London, 2018. *Travel in London: report 11.* London: Transport for London.

Transport for London, 2019. *Plan a journey,* London: Transport for London.

Transport for London, 2019. *Transport for London unified API,* London: Transport for London.