**Introduction**

Cities support dense social, economic, physical and intellectual networks, which facilitate frequent and complex interactions that in turn drive high levels of productivity, efficiency and innovation (Bettencourt & West, 2010). However, inequality also tends to be greatest in cities, as people in different network locations experience different levels of connectivity to opportunities and resources, which in turn contributes to differentiated outcomes among individuals. This can be tangibly seen in how workers who live in different neighbourhoods experience uneven 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, workers burdened by lower accessibility may also experience 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 a highly salient aspect of the actual experience of urban commuting: day-to-day variability in travel times (Katz & Quealy, 2019), which may arise from congestion, varying service schedules and failures of parts of the network. This is because data on transport connectivity with sufficiently fine spatial and temporal resolution to capture such variability has only recently become freely available, particularly through real-time trip planning platforms. This study leverages on this type of data source to incorporate travel time variability into a job accessibility analysis.

In this study, I will analyse the how accessibility to jobs and variability in travel times via public transport affects working-age residents living in different areas under the jurisdiction of the Greater London Authority (GLA), with the aim of identifying areas that are burdened with poor accessibility to jobs, low travel time reliability and/or high social disadvantage. In addition to demonstrating a methodology to productively incorporate the novel angle of travel time variability into accessibility analyses, the findings from this study can also be used to target public transport improvements to increase accessibility to jobs and travel time reliability 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 shows that accessibility levels usually vary widely across a city. Also, most cities have locations where social disadvantage coincides with poor accessibility, although this problem is more serious in poorer cities with less extensive public transport systems than in more developed cities.

In the context of London, both the UK government (Social Exclusion Unit, 2003) and the GLA (2018) recognise that 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.

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 dimensions of disadvantage at the local scale (El-Geneidy, et al., 2016; Deboosere & El-Geneidy, 2018). In the UK, the Department for Housing, Communities and Local Government (DHCLG) has developed the Indices of Multiple Deprivation (IMD) to reflect the relative distribution of deprivation among small areas 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 areas, and will follow the DHCLG guidance to focus on the relative ranking of the areas by IMD index scores, rather than on the raw scores themselves.

Researchers have also developed many metrics to measure accessibility (Geurs & Wee, 2004). The “location-based” type, which regards accessibility as a property of different areas (as opposed to a property of travellers or of components in the transport network), 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). In particular, I will use a type of adapted cumulative potential measure (Geurs & Wee, 2004, p. 133) first proposed in the transport geography literature by Shen (1998) and further developed in the public health literature (Luo & Wang, 2003; Luo & Qi, 2009; Wang, 2012) as the two-stage floating catchment area method. This expresses accessibility as a ratio of opportunities (in the case of this study, jobs) per client (workers) who can reach them, and is thus a function of the spatial distribution of the supply of opportunities and the demand from clients, and of the connectivity between areas. More details about how this accessibility measure is calculated are given in the next section. However, an important property of this measure is that it allocates every opportunity exactly once to clients in a mutually exclusive way, so that there is no double-counting of clients or opportunities (Shen, 1998, pp. 363-364). Thus, unlike simpler location-based measures that only count the number of opportunities supplied within reach from any area without considering the distribution of demand, this ratio accounts for both connectivity and competition effects: as the number of opportunities in an area increases, the ratio will increase, but as the number of clients who can reach the area increases, the ratio will decrease. These properties are especially useful for assessing whether the local supply of opportunities is more or less than what is needed by the local demand.

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 fine-grained location and connectivity 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 that results from variation over time. Thus, extant analyses do not account for the uncertainty that actually exists around accessibility levels.

**Study Design and Data Sources**

This analysis will focus on accessibility to jobs from home locations. Even though not all trips (even during the peak hours) are commutes from homes to work, home-based work commutes are a primary driver of travel demand during peak periods, and thus significantly shape public transport investments that need to be scaled to handle peak demand. In addition, most of the opportunities that people seek (such as education, healthcare or shopping) also provide jobs, so accessibility to jobs will at least somewhat capture accessibility to opportunities in general (Deboosere & El-Geneidy, 2018, p. 56).

The two-stage floating catchment area method I use for this study involves 3 components:

1. The spatial distribution of jobs, stored in a column vector O of length n

where oi is the number of jobs in one of the n areas i;

1. The spatial distribution of working-age residents, stored in a column vector P of length n

where pi is the number of working-age residents in one of the n areas i, and;

1. A n by n matrix C of weights

where ci,j is the cost (in this study, time) of travelling from area i to area j, and f(ci,j) is a function that decreases as ci,j increases. Thus, f(ci,j) indicates the ease with which j can be reached from i.

Then, the accessibility values, stored in a column vector A of length n, is given by

where gives the weighted sum of the number of clients that can reach the opportunities in each area, divides element-wise the number of opportunities in each area by the weighted sum of the number of clients that can reach them, and is a weighted sum of the number of opportunities provided by each area to each client who can reach it.

The use of this method requires two choices. The first choice involves which spatial units to use for the analysis. 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; Wang, 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; Luo & Wang, 2003; 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 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.

The second choice involves which impedance function to use. Many studies use a power or exponential function to capture decreasing effects with increasing distance (i.e. distance decay), such that decreases smoothly as increases, resulting in a “gravity”-based measure where accessibility 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. However, origin-destination flow data and additional modelling is needed to estimate the distance decay functions, and 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 Luo & Wang (2003), 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 easy to understand 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 used interchangeably. For the purpose of this study, I will set X to 45min, following TfL’s (2018, p. 224) own analyses. Also, because the resulting quantities in the matrix A can be interpreted as ratios allocating jobs to working-age residents in a mutually exclusive way, I will term these quantities Job Provision Ratios (JPRs).

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 obtained 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 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 suited to provide information on travel time variability due to real-time fluctuations in transport service performance.

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 up to a maximum of 300 queries per minute. I used the API to compile the following databases:

1. The travel options for trips between every pair of MSOAs ending at 0830 on 22 March 2019. This constitutes the “baseline” travel option data for all trips for the weekday AM peak, assuming all services run as planned. Due to the API request limit, data collection for all 983 x 983 trips occurred over 1 month.
2. The travel options for selected trips beginning between 0600 and 0800 every weekday for 50 days between 29 January 2019 and 25 April 2019. This constitutes “real-time” travel option data during the weekday AM peak, reflecting actual day-to-day variability in transport service performance. Because of the API request limit, I only gathered data on trips starting from all 983 MSOAs that end at 7 MSOAs that correspond to 7 key transport nodes evenly spaced across London. See Appendix XX for details on how the 7 destinations were selected.

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.

**Results**

*Distributions of Jobs, Working-Age Residents, Inter-MSOA Connectivity and IMD Ranks*

As conceptualised in this study, each area’s accessibility to jobs is a function of 3 components: the spatial distributions of working-age population and jobs, and the area-to-area connectivity provided by the public transport system. Figure XX shows the first two components. It is clear that the distribution of jobs is extremely concentrated in the centre of the GLA area, where the City of London, Westminster and Southbank are located. However, there are also smaller concentrations of jobs around some stations in the GLA’s rail and tube network (black lines in the figure), most notably at Canary Wharf. In contrast, the distribution of working-age residents is more even, with lower residential densities in the areas with the most jobs near the centre, higher densities in a ring around the central job-rich core, and decreasing densities as one moves further away from the centre.

The third component, area-to-area connectivity, is stored as a n x n matrix of travel times T between each of the n = 983 MSOAs. The value in each cell of T, , is the minimum time (out of all available options) it takes to travel between the population-weighted centroid of MSOA i and the job-weighted centroid of MSOA j during the AM peak using the baseline database, that is, assuming all services run as scheduled. It is difficult to visualise the entire matrix in a meaningful way, but Figure XX visualises the travel times for all trips originating from the most populous MSOA: Stratford, MSOA ID E02000726, (i.e. the row vector from T where i = E02000726) with the distribution of jobs superimposed. The inset chart shows the total number of jobs that can be reached from Stratford within a certain travel time, i.e. . As the travel time threshold X increases, increases, until all jobs available in the GLA can be reached within X minutes.

The relative deprivation of different MSOAs is also important for us to identify which areas in the GLA area are burdened by both accessibility problems and social disadvantage. Figure XX shows the ranking of each MSOA by IMD scores, together with 3 raw indicators of social disadvantage, namely the modelled mean household income in 2015 (post-housing costs and equivalised for household size), the estimated proportion of households in poverty in 2014, and the unemployment rate among working-age residents in 2017. For all 4 maps, MSOAs are binned into deciles by the distributions of each variable, with more deprived MSOAs visualised in darker shades of red and less deprived ones in darker shades of green. In general, the IMD rankings correspond quite well to the 3 raw indicators of social disadvantage, with a swath of more deprived MSOAs running north-south and another swath running east-west along the northern bank of the River Thames.

*Analysis of Accessibility to Jobs*

In the inset of Figure XX, I plotted the curve of values for Stratford as the travel time cut-off increases. This exercise can be repeated for all MSOAs to yield Figure XX, which plots the value curve for each of the 983 MSOAs. 3 MSOAs are highlighted: Stratford (most populous), the City of London (most jobs) and Waterloo (the MSOA with among the highest value for any given travel time cut-off). The plot reveals 2 points. First, the time needed for residents of each MSOA to be able to reach a certain number of jobs (say half of all available jobs) varies quite widely, from around 45min for residents of the City and Waterloo, to more than 90min for the least well-connected MSOAs. Secondly, the rank-ordering of MSOAs by values does not vary significantly regardless of choice of travel time cut-off; an MSOA that has one of the highest values with a cut-off of 45min will also tend to have one of the highest values with a cut-off of 90min. This reinforces El-Geneidy et al (2016)’s point that contour and gravity-type accessibility measures are largely interchangeable, and indicates that the broad findings from this study are relatively robust against changes in the travel time cut-offs used.

I choose a travel time cut-off of 45min (marked with a red dashed line on Figure XX) to align with TfL’s own accessibility analyses. The value of where each curve intercepts the 45min cut-off line gives a vector of 983 values (one for each MSOA). This is plotted in the left panel of Figure XX, with the values binned into deciles. Clearly, areas in the centre of the GLA area with lots of jobs also have high values. However, the connectivity provided by the public transport network (particularly the rail network, visualised with black lines) extends the region with high values well beyond the City and Westminster, so much so that areas near the edge of the GLA area that are near rail stations may also have relatively high values. However, areas that are not directly served by rail tend to have lower values.

However, as noted above, values improve on as the former account for competition effects from residents of different areas for the same jobs. The right panel of Figure XX visualises the values of each MSOA, which can be straightforwardly interpreted as the following ratio: the number of jobs within 45min AM peak travel time from each MSOA that are allocated to each working-age resident in that MSOA. If jobs, workers and connectivity were all evenly distributed throughout the GLA area, then all the values will equal the ratio of the total number of jobs to the total number of working-age residents, . To express the extent to which actual values deviate from , values are binned into 5 classes: 0 to 50% of (jobs allocated << working-age residents), 50% to 80% of (jobs allocated < working-age residents), 80% to 125% of (jobs allocated ≈ working-age residents, this is a balanced state), 125% to 200% of (jobs allocated > working-age residents), and more than 200% of (jobs allocated >> working-age residents).

From the right panel of Figure XX, we can see that few MSOAs in the GLA area have values close to a balanced state. MSOAs near the centre have too many jobs allocated per working-age resident, while many MSOAs around the outskirts have too few jobs per working-age resident. This is because jobs are mostly concentrated in the centre, while working-age residents are relatively evenly spread out across the GLA, and the public transport system heavily prioritises fast connections to the centre. A notable exception is the relatively more balanced values in the west, where Heathrow Airport provides a significant number of jobs, and connections between Heathrow and areas in the west are relatively good.

*Analysis of Accessibility to Jobs by Sector*

The job distribution data disaggregates jobs by Standard Industrial Classification Codes (SICCs). This allows us to analyse how jobs are allocated to working-age residents by sector. First, I aggregate the SICCs to 6 overarching sectors:

1. Primary and Extractive
2. Industrial
3. Retail, Logistics, Food & Beverage and Hospitality
4. Finance, Insurance, Real Estate, Professional and Administrative Support
5. Public Sector
6. Others

The inset of Figure XX shows the proportion of all jobs in the GLA area that belong to each sector. The finance, insurance, real estate, professional and administrative support sector is the largest in the GLA area, followed by the retail, logistics, food & beverage and hospitality sector and the public sector. Figure XX plots the distribution of jobs in each sector as a dot density map. Industrial jobs are the most evenly spread out across the GLA area, while retail, logistics, food & beverage, hospitality and public sector jobs are somewhat less spread out with significant concentrations in the job-rich centre. Unsurprisingly, finance, insurance, real estate, professional and administrative support jobs are the most concentrated in the centre.

For each MSOA, I calculate values using the job distribution for each sector and plot maps of values per sector in Figure XX. The top left panel of Figure XX reproduces the map in the right panel of Figure XX for reference. Each map uses the same binning convention as the top left panel, except that the values of are now sector-specific. From the maps, we can see that accessibility to industrial jobs is most evenly distributed, while accessibility to public-sector and retail, logistics, food & beverage and hospitality jobs are also relatively evenly distributed albeit with significant over-provision of jobs per working-age population at the centre of the GLA area. In contrast, access to finance, insurance, real estate, professional and administrative support jobs is most unevenly distributed, with large swaths of MSOAs in the south, east and north having almost no jobs from this sector allocated per working-age resident.

When we plot Lorenz curves of the distribution of jobs allocated per sector to the working-age population (lower right panel of Figure XX), we can assess the extent to which the distribution of jobs in each sector contributes to the unevenness in overall accessibility to jobs. The Lorenz curves for the industrial, retail, logistics, food & beverage, hospitality and public sectors are closer to the diagonal than the curve for all jobs, but the curves for the finance, insurance, real estate, professional and administrative services and other sectors are below the curve for all jobs. This indicates that the distributions of the latter two sectors tends to make accessibility to jobs more uneven in the GLA area. In particular, because the finance, insurance, real estate, professional and administrative services sector is the largest in the GLA, the high concentration of jobs in this sector in the centre of the GLA area tends to drive the unevenness in the distribution of accessibility to jobs in the GLA area.

*Identifying Socially Deprived Areas with Poor Accessibility to Jobs*

Ideally, we should improve the public transport network and redistribute jobs and homes such that accessibility to jobs is even throughout the GLA area. However, in order to minimise social inequity, improvements should be prioritised to benefit the neediest areas first (citation). To do this, we need to identify areas which have high social deprivation and poor accessibility to jobs.

To this end, I carry out k-means clustering to classify MSOAs according to their sector-specific values and their IMD ranks, to identify a typology of MSOAs characterised by the number of jobs per sector allocated to each working-age resident and the relative level of deprivation of each area. The results for when k = 6 are presented in Figure XX. The upper left panel visualises the distribution of IMD ranks in each cluster. Cluster 0 is most deprived, while Clusters 1 and 5 tend to be more deprived, and Clusters 2, 3 and 4 tend to be less deprived. The upper right panel visualises the distribution of values for all jobs per cluster. Among the more deprived clusters, Cluster 0 has the lowest values, indicating areas that tend to be most deprived and to have the poorest overall accessibility to jobs. However, Clusters 1 and 5 tend to have values that indicate too many jobs allocated per working-age resident.

The lower left panel of Figure XX displays the distribution of sector-specific values per cluster. This sheds some more light on the job allocation situation for Clusters 1 and 5. Although these clusters tend to have plenty of jobs per working-age resident, most of these jobs are in the finance, insurance, real estate, professional and administrative support sector. Insofar as these sectors tend to require higher qualifications, there may be a skills mismatch between the more socially disadvantaged residents of MSOAs in Clusters 1 and 5 and the jobs that are available to them; in other words, though there may be enough jobs available, the types of jobs available may not match the types of workers available.

In contrast, MSOAs in Clusters 2, 3 and 4 tend to have lower deprivation but also insufficient jobs allocated per working-age resident. Faced with poor accessibility to the jobs they seek via public transport, residents in these MSOAs may be more likely to rely on cars for their commutes. Improving public transport accessibility to jobs for these areas may be a prerequisite to encouraging less car-dependent commute choices for residents of these areas, but the effectiveness of public transport improvements to encourage lower car usage remains in doubt.

*Analysis of Variability in Travel Times*

In addition to collecting data on the options to travel between every pair of MSOAs assuming all services run as scheduled, I collected AM peak travel option data on trips from all MSOAs to 7 strategically chosen destinations over 50 days in real time. Thus, for each of these 6881 trips, I obtained a distribution of 50 minimum travel times.

Each panel in Figure XX plots the width of the 90% highest probability density interval (hereafter the interval width) for trips originating in each MSOA and terminating at one of the 7 destinations (outlined in red in each map). The interval width indicates how compact the distribution of observed travel times is for each trip; for example, an interval width of 5min means that the 95th percentile observed travel time (for an unusually slow trip) is only 5min more than the 5th percentile (for an unusually fast trip), indicating that travel times on that trip are relatively reliable. To facilitate comparison of interval widths between different destinations, all maps use the same binning scheme to visualise the interval widths.

In general, we can see that areas that are within walking distance of each of the 7 destinations have the narrowest interval widths, indicating that walking tends to be a reliable travel mode. This makes sense, as the flexibility of walking means that routes are unlikely to be totally blocked, and even if a route is blocked, the high density of the walkable network means that travellers are likely to be able to find comparable alternative routes. Areas that are connected to the destination by rail also tend to have narrower interval widths, indicating that rail travel is generally quite reliable except in the event of a line disruption. Areas that are not directly connected to a destination by rail and that are too far to walk tend to have wider intervals, even if they are relatively near the destination. Such areas tend to rely more on buses to connect to each destination, and these wider intervals indicate that travel times by bus are less reliable. Also, locations further from a destination tend to have wider intervals, as these require longer journeys to connect to the destination, and the probability that a journey will be disrupted en route increases with the journey length.

Also, we can see that trips ending in destinations north of the Thames, particularly in the City but also in Finsbury Park, Harrow-on-the-Hill and Hammersmith, tend to have less travel time variability. This may be attributed to the more grid-like rail and tube network north of the Thames, which provides commuters with more alternative rail routes if a particular section of the rail network fails. In contrast, travel times to Lewisham and Streatham tend to be more variable, possibly reflecting the more tree-like structure of the rail network there, which provides less network redundancy against failures of line sections leading to these destinations. Finally, for all destinations, we can see that trips originating from the southeastern quadrant of the GLA area tend to have more variable travel times. This probably reflects the effects of less reliable rail services run by Southeastern and Southern Railways (+ citation).

*Analysis of Variability of Travel Times and IMD Rank*

An important question to investigate is the extent to which areas that are burdened by high travel times and/or low travel time reliability are also more socially deprived. To do this, I carry out K-means clustering on the MSOAs, considering the observed travel times and interval widths for travel to each MSOA’s local centroid (one of the 7 destinations that is closest to it) and to the City of London, together with each MSOA’s IMD rank.

Figure XX presents the characteristics of each cluster when 7 clusters are identified. Clusters 2, 3 and 4 tend to have lower IMD ranks that reflect higher social deprivation. Among these, Cluster 3 tends to have lower and less variable travel times to the City and to their local centroid. MSOAs in this cluster tend to be more centrally located and to have direct access to the rail system. Cluster 2 tends to have higher but less variable travel times to the City and their local centroid. These tend to also have direct access to the rail network, but are located further from the centre than MSOAs in Cluster 3. On the other hand, Cluster 4 tends to have higher and more variable travel times to the City and their local centroid. While MSOAs in this cluster may also be relatively centrally located, they tend to not have direct access to the rail system or to rail lines that run radially to the centre of the network, and thus may need to rely more on slower and less reliable buses or to make longer rail trips with more interchanges.

However, it is also important to note that MSOAs in clusters with higher IMD ranks (and thus lower social deprivation) also tend to have longer and more variable travel times to the City and the local centroid. With the exception of Cluster 0, MSOAs in these clusters tend to be located near the outskirts of the GLA area. The lower usefulness and reliability of the public transport network for connecting residents in these MSOAs to important destinations may contribute to these residents’ higher reliance on cars for their travel needs.

*Contributions to Travel Time Variability by Travel Mode*

As we have seen, travel time variability varies across different trips. But why do some trips have more variable travel times than others? One factor could be the distance covered for each trip; the probability that a disruption occurs at any point along a trip increases with the length of the trip. The results seen so far also indicate that some travel modes (such as walking, rail or tube) are more reliable than others (particularly bus). Also, we have seen that certain locations may be burdened by the low reliability of the service operators (particularly in the southeast).

To investigate the factors shaping travel time variability across trips, I carry out a set of linear regression analyses on n = 6881 trips, with trip-level interval widths as the dependent variable. In an initial model, I use the distance covered by each mode of transport for the fastest option from the baseline data as covariates – that is, I investigate how the interval width of the time needed to complete a trip is correlated with the distance travelled for that trip by each mode, assuming all services run as scheduled.

The results for this model are shown in Column 1 of Table XX, and indicate that for every additional kilometre travelled by National Rail, the interval width increases by 0.1min. This result is in line with the observations above that unreliability in the services operated by some rail companies could contribute to more variable travel times. Every additional kilometre travelled by Overground also increases the interval width by 0.1min. The Overground was conceived to take over the operation of some National Rail lines to increase service reliability, so this result is discouraging. However, freight trains do share tracks with Overground trains and freight train breakdowns are known to cause Overground disruptions. However, buses are by far the most unreliable mode of public transport, as every additional kilometre travelled by bus increases the interval width by 1min. The results also indicate that every additional km walked increases the interval width by 0.7min. This contrasts with the observation above that walking is the most reliable mode of transport, but can reflect that trips that involve more interchanges also require more walking, and the coefficient for distance walked reflects the increased vulnerability of trips with more interchanges to being disrupted at any single point en route.

However, because all the MSOAs are connected to each other via the same transport network, it is not true that the interval widths for all trips are mutually independent of each other. For instance, two MSOAs that are located along the same rail line leading to the City can be expected to have similar interval widths for trips ending in the City, as both will be affected by any disruptions along that line. One way to deal with such correlations among the dependent variables is to use a spatially autoregressive model, in which the value of the dependent variable in each area depends on the mean of the value of the dependent variable in its neighbours. To implement this, I first obtain for each MSOA a vector of travel times to the 7 destinations from the baseline data, and then calculated the Euclidean distance between every pair of vectors. Then, for each MSOA, I identify the 3 other MSOAs where this Euclidean distance is minimised; these are the MSOAs whose travel times to the 7 destinations are the most similar to those of the MSOA in question, assuming all services run as scheduled. Then, for each trip originating from the MSOA in question and terminating at each of the 7 destinations, I calculate the mean of the interval widths for trips starting from the 3 neighbours and ending at the same destination. I then include this local mean as a covariate in the linear regression model.

Column 2 in Table XX presents the results for this model. As expected, the inclusion of the local mean as a covariate significantly improves the fit of the model, as it allows it to better capture the correlation between the interval widths of MSOAs that arise due to their locations in the transport network. Specifically, for every additional minute in the local mean interval width, the interval width of the MSOA in question tends to increase by 0.6min. The inclusion of the local mean also reduces the estimated coefficient sizes for the other covariates, by about 40% for bus and 70% for National Rail and Overground. However, the coefficients still indicate that bus is the least reliable mode, followed by National Rail. Interestingly, the coefficient for distance walked is now no longer statistically significantly different from 0, indicating that the coefficient estimate for walking obtained from the previous model was indeed due to the network effects from trips involving more interchanges being more vulnerable to disruption. Finally, increasing travel by tram seems to decrease interval widths. In London, trams tend to run on segregated rights of way, and thus may not be as vulnerable to disruptions as buses or trams in other cities that run in mixed traffic. However, because trams only serve areas in southern London and run orbitally, they do not play a significant role in most residents’ commutes.

While the regression models in Table XX generate one global estimate for the correlation between distance travelled by each mode and interval width, we have seen signs that the mode-specific effect may vary across locations (recall that National Rail services seem particularly unreliable in the southeast). To investigate this, we can use locally-weighted regression, which uses only the trips originating from the closest neighbours for each MSOA and ending in the 7 destinations to fit a linear regression model to generate one estimate for each coefficient for each MSOA. I define the 30 MSOAs whose travel times to the 7 destinations are most similar to those of the MSOA in question as neighbours of that MSOA. Figure XX displays for Stratford, which MSOAs are included in the locally-weighted regression and which are further considered for calculating the local mean of the interval width.

Column 3 in Table XX displays a summary of the distributions of the coefficient estimates for each covariate obtained from the locally weighted regression. The distribution of local estimates contains the global estimates from Column 2, but demonstrates that the mode-specific effects may vary quite significantly. Figure XX maps out the R2 values for the locally fitted regression models, and Figure XX maps out the p-values for each coefficient estimate for each MSOA. Further, Figure XX maps out the coefficient estimates for each covariate and each MSOA. Green shades indicate that a mode contributes less to travel time variability in that MSOA, while red shades indicate that the mode contributes more to travel time variability. Grey MSOAs indicate where the coefficient estimates were not statistically significantly different from 0.

Overall, travel by tube, tram and Docklands Light Railway tends to reduce interval widths, reinforcing the finding that these rail-based modes are most reliable. Also, there are areas near the outskirts of the GLA area and in a western wedge where travel by bus seems quite reliable. However, in a swath of MSOAs running from the northeast to the south, bus travel seems to contribute disproportionately more to travel time variability. This could reflect the effect of bus congestion, particularly around the City, where too many buses each with few passengers could lead to slow and irregular services.

The results for National Rail are also interesting. Firstly, as mentioned previously, travel by National Rail in the south and southeast seems to disproportionately contribute to travel time variability in those areas. However, travel on National Rail also seems to increase travel time variability more in the northwest and northeast. This could be because National Rail lines from these areas terminate outside the City of London (at stations like Paddington, Marylebone, Kings Cross, Moorgate and Liverpool Street) and do not provide interchanges to orbital routes. Thus, passengers travelling via National Rail from these areas to any of the 7 destinations probably need to interchange at these terminals for more onward travel via tube or local bus services. This interchanging may make these trips more vulnerable to disruption. Finally, although travel by National Rail does tend to increase travel time variability overall, there are isolated locations in the southwest and the north where National Rail is dependable enough to actually reduce interval widths.

The results for Overground travel indicate that the Overground seems less reliable north of the Thames and more reliable in the south. This seems to make sense, as north of the Thames, tube services are much more available and reliable, while south of the Thames, where tube coverage is worse, the Overground may be an effective and more reliable alternative to less reliable National Rail and bus services. For the new TfL Rail service, the western sector seems to be particularly unreliable.

Finally, while walking may not have a statistically significant effect on travel time variability at the global level, there are locations with significant local effects. In the north, northeast and southwest, increasing distance walked tends to increase travel time variability, probably reflecting that trips from these areas to the 7 destinations tend to require more interchanging. In the southeast, however, increasing walking decreases travel time variability; it seems that rail and bus services in these areas are so unreliable that commuters who walk more may have more reliable travel times.

*Reliability of Travel Modes and Social Deprivation Levels of MSOAs*

A final question to investigate is the extent to which areas that where certain travel modes are particularly unreliable are also socially disadvantaged. Again, I carry out a K-means clustering analysis using the MSOA-level coefficient estimates from the locally weighted regression (considering only the estimates for the intercept, local mean interval width, bus, National Rail, tube and walking, since estimates for other covariates are not available for some MSOAs) and the MSOA-level IMD rank. The results for 6 clusters are shown in Figure XX.

Cluster 1 immediately stands out as the MSOAs where National Rail services are particularly unreliable. These are areas served by Southern and Southeastern services, which are known to have reliability issues. Clusters 0 and 4 tend to contain MSOAs with lower IMD ranks and thus higher social deprivation. Cluster 0 tends to have lower coefficient estimates for the mode-specific contributions to interval widths, indicating that travel by bus, tube, National Rail and walking in these areas tends to be reliable. On the other hand, Cluster 4 has higher coefficient estimates for distance travelled by bus and National Rail, indicating that these MSOAs are burdened by both high social deprivation and unreliable travel by these modes. Finally, Clusters 3 and 5 tend to contain less deprived MSOAs, with MSOAs in Cluster 5 having more reliable National Rail services than those in Cluster 3.

**Discussion**

# **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.