**Introduction**

A city’s vibrancy, productivity, efficiency and innovation depend in large part on how efficiently it can connect job opportunities and workers to each other through economic, social, intellectual and physical networks (Bettencourt & West, 2010). An approach to urban planning that integrates land use distribution and transportation strategies can facilitate this by maximising job accessibility (that is, matching the distributions of workers’ home locations and job opportunities to overcome spatial and temporal distance), so that the available transportation capacity can be used most efficiently to maximise the range of job opportunities that workers can access within reasonable travel times (Hansen, 1959).

An understanding of the macro-scale spatial patterns in the extent to which workers can access jobs is an important first step for identifying and prioritising problem areas, and optimising and coordinating intervention strategies (Geurs & Wee, 2004). However, it is also important to keep in mind that job accessibility varies across many dimensions besides space (Lucas, 2012). In particular, accessibility levels vary across different types of workers and jobs (depending on their characteristics like skills, salary and industrial sector, not every job is equally attractive or suitable for every worker), and fluctuates from day to day (as there is a margin of uncertainty in travel times that travellers need to account for when deciding how and when to travel).

This study aims to carry out an analysis of job accessibility in London that accounts for this multidimensional variability. I develop a typology of neighbourhoods that categorises them by overall job accessibility by public transport, their sectoral dependency for job opportunities, public transport travel time reliability and the relative social disadvantage of their working-age population. In addition to providing insights into how land use and transport interventions can adjust the large-scale spatial structure of London to improve job-worker spatio-temporal matching, the results can also provide guidance for where to focus measures that improve workers’ skillsets, encourage job growth in certain sectors and improve public transport reliability, thus allowing for closer coordination of strategies across a wider range of domains.

**Literature Review**

Hansen (1959) defines accessibility as a measure of the potential for interaction between people and the opportunities that they seek, and is a function of how people and the relevant opportunities are distributed across space and the cost of travel (or connectivity) between different locations. In his original conceptualisation, accessibility increases as the number of opportunities near a person increases and as the cost of travelling to the opportunities decreases.

Based on this definition, researchers have developed a wide range of indicators to analyse accessibility (Geurs & Wee, 2004). Of these, “location-based” measures are widely used for city-level analyses, because they treat accessibility as a property of different areas (as opposed to a property of travellers or of components in the transport network) and thus can be used to spatially aggregate and visualise the distribution of accessibility. In general, analyses of job accessibility using these measures have found wide within-city variations of accessibility (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). This is largely because job distributions tend to be highly clustered and connectivity is unevenly provided, as areas with higher density tend to generate more trips and thus tend to be better connected with faster, higher-frequency transport links.

One drawback of location-based accessibility indicators is that they match workers to jobs as if every worker and every job were interchangeable. To capture how accessibility levels vary across different types of workers and jobs, some studies separately measure accessibility for subsets of workers (such as low-income workers) matched to the jobs that are suitable for them (like jobs in sectors that tend to pay lower wages) (Shen, 1998; Deboosere & El-Geneidy, 2018), while others examine how accessibility levels vary by area-level indicators of social disadvantage (such as neighbourhood average income or an index that summarises multiple dimensions of social deprivation) (Sanchez, 1999; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018). These analyses have found that most cities have areas where more disadvantaged residents are burdened with poorer accessibility to jobs, though this seems to be a more systematic problem in poorer cities with less extensive public transport systems.

In the context of London, Ford et al (2015) showed that job accessibility varies significantly across space, but as far as I am aware there has not been a published analysis of how job accessibility varies across different types of workers and jobs. Nonetheless, both the UK government (Social Exclusion Unit, 2003) and the Greater London Authority (GLA) (2018) recognise that low accessibility may significantly hinder disadvantaged residents’ access to jobs as well as other services like healthcare, education and leisure, thereby constraining their options, increasing travel burdens, lowering quality of life and entrenching social deprivation. 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 increasingly recognised that the spatial distributions of workers, jobs and connectivity 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 three distributions over a typical day and calculated accessibility measurements for each area at each time point. However, these studies calculate accessibility based on “typical” (mean or ideal) values of these variables at each time point, without accounting for the stochastic distribution of possible values around the typical values. In particular, these studies use a single value to represent the travel time between any two locations, and thus do not capture the margin of uncertainty around travel times that travellers actually need to account for deciding when and how to travel.

Efforts to account for travel time reliability in accessibility analyses have been hindered by a scarcity of data on the stochastic distributions of travel times at the level of each trip, due to the labour intensiveness and expense of traditional methods to collect such data (such as floating traveller surveys, in which a person repeatedly conducts a trip at the same time over different days and records the time needed to complete it). However, recent advances in large-scale data collection and dissemination in public transport systems, such as through systems for automatic fare collection, automatic vehicle location and real-time trip planning, have overcome this scarcity and allowed for new measures of travel time variability to be developed and calculated at very fine levels of spatial and temporal resolution (+ citation). In particular, Uniman et al (+ citation) used fare collection data from London’s Oyster Card system to reveal the distribution of travel times for trips actually performed on the Underground network. On the basis of this, they define the Reliability Buffer Time (RBT) as a measure of travel time reliability. Simply put, the RBT is the difference between the 95th percentile of travel times (representing an unusually slow trip) and the median travel time (representing a typical trip), and reflects the extra time that a traveller needs to budget into her trip in order to be 95% certain that she will not arrive late on any given day.

In light of the existing literature, this study is positioned to make several contributions to the analysis of job accessibility. Firstly, it will conduct a job accessibility analysis of London that better captures how accessibility levels vary among working-age residents and jobs with different characteristics. Secondly, it will also consider how travel time variability affects different areas in London, adding a novel dimension to accessibility analyses. In the next section, I present the methodology of this study and the data sources I will use.

**Study Design and Data Sources**

This analysis will focus on accessibility to jobs from working-age residents’ home locations. Even though not all trips (even during 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 the layouts and operations of public transport networks. 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).

*Data on the Spatial Distributions of Jobs, Workers and Social Deprivation*

MSOA-level counts of working-age residents (those aged 16 to 64) and jobs for 2017 were obtained from the Office of National Statistics (+ citation). Job counts are disaggregated by Standardised Industrial Classification Code (SICC), allowing for sector-specific analyses of the job distribution. Unfortunately, counts of residents that are disaggregated by other characteristics (like educational qualifications, income or disability status) are only available from Census data, which are 8 years out of date at the time of this study.

In the absence of suitable disaggregated count data for the working-age population, many researchers use location-based measures to reflect spatial disparities in sociodemographic characteristics, such as neighbourhood average income (Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018) or an index combining multiple dimensions of social 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) (2015) to reflect the relative degree of social disadvantage experienced by residents in small areas across a wide range of domains, including income, employment, health, education, crime, housing and the living environment. In this study, I use the IMD data to capture the spatial distribution of deprivation, and I follow the DHCLG guidance to focus on the relative ranking of the areas by IMD index scores, rather than on the raw scores themselves.

*Measuring Job Accessibility through Job Provision Ratios*

Location-based accessibility measures are attractive as they can be easily used to spatially aggregate and visualise accessibility levels. The simplest location-based measure reports accessibility in terms of “cumulative potential” (Geurs & Wee, 2004, p. 133), summing up the total nuamber of jobs that can be physically reached from each area within a certain travel time. Essentially, this is the dot product of 2 components, , where:

1. O is a column vector of length n:

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

1. C is a n by n matrix 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 of travel from i to j.

However, such a measure treats accessibility to jobs as non-excludable; that is, the fact that a particular worker can access a job does not deprive any other workers of the ability to access the same job. This does not reflect the reality that each job typically can only be occupied by one person. To address this theoretical drawback, adjustments were proposed by Shen (1998) and elaborated in the public health literature (Luo & Wang, 2003; Luo & Qi, 2009; Wang, 2012) to account for:

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.

Then, the accessibility values, stored in a column vector R of length n, are obtained by:

where gives the weighted sum of the number of working-age residents within each area’s catchment, divides element-wise the number of jobs in each area among the weighted sum of working-age residents within its catchment, and is a weighted sum of the number of jobs provided by each area to each working-age resident within its catchment. Luo & Wang (2003) termed this the “two-stage floating catchment area” method, but because the resulting quantities in the matrix R for this study can be directly interpreted as ratios allocating jobs within each area’s catchment to working-age residents, I will term these quantities Job Provision Ratios (JPRs).

An important property of this measure is that it allocates every job exactly once (i.e. there are no leftover unallocated jobs) to potential workers in an excludable way (i.e. if a job is allocated to a particular worker, that job is no longer available to other workers) (Shen, 1998, pp. 363-364). Thus, this ratio accounts for both connectivity and competition effects: as the number of jobs within an area’s catchment increases, the ratio for that area will increase, but as the number of potential workers who can reach the same catchment (and thus compete for the same jobs) increases, the ratio will decrease. This property is especially useful for assessing whether the local supply of jobs is sufficient to satisfy the local demand, and methods to assess the degree of inequality in distributions such as Lorenz curves can be straightforwardly applied.

The use of this method requires two choices. The first choice involves which spatial units to use for the analysis. Some studies divide the study area 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 data to these squares requires very spatially disaggregated 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 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 that break up the UK’s land mass into areas that contain between 5,000 and 15,000 residents or between 2,000 and 6,000 households, and allow my analysis to directly incorporate a wide range of UK government statistics that are spatially disaggregated to the MSOA level.

The second choice involves how to define the function . Many studies use power or exponential functions so that decreases smoothly as increases (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 “distance decay” in the attractiveness of destinations that are further away. However, fine-grained origin-destination flow data and additional modelling are needed to estimate the parameters for 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 ratios that can be interpreted directly (Luo & Wang, 2003; Deboosere & El-Geneidy, 2018). At any rate, El-Geneidy et al (2016, p. 303) found that contour and distance decay-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.

*Inter-MSOA Travel Time Data*

A key dataset required for calculating accessibility measures 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, and can easily simulate alternative scenarios by adjusting the network and recalculating the costs. However, the approach requires an extensive data compilation and cleaning effort and is sensitive to the researchers’ choice of routing algorithm.

An alternative that has become feasible in recent years is to use automated fare collection data to obtain the full distribution of actual travel times between any two stations in the transport network, and then to calculate a single representative value like the mean or median travel time for each pair of stations. However, two factors limit its applicability in the context of London. First, while TfL’s automated fare collection system captures the locations of all travellers upon boarding a bus or rail vehicle, the alighting locations of customers using buses (about 45% of all tap-ins on TfL vehicles) (Transport for London, 2018, p. 29) are not recorded. While alighting locations can be imputed using modelling to some extent, this may introduce systematic distortions into the data. Second, the raw data is highly protected because it is possible to use it to identify individual travellers in the system, and recent studies using TfL’s automated fare collection data have had to rely on data that is many years out of date (+ citations). Also, a more general limitation of this approach is that the cost of first- and last-mile travel (from the trip origin to the first boarding location, and from the last alighting location to the final destination) is not reflected in the data, and more modelling is required to fill in these gaps.

Instead, I 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 walking, cycling or motorised public transport like buses or trains (but not by car or taxi). This data source has several compelling advantages. First, it 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. Secondly, the Journey Planner provides information for up to 6 route options per trip, including a bus-only option, a cycling-only option and a walking-only option, thus allowing for travel costs to be calculated under different circumstances (such as different levels of willingness to cycle). Thirdly, the travel option information is door-to-door: it includes both first- and last-mile travel information, as well as information for travel within the public transport network. Fourthly, the data is highly disaggregated: for each option, the overall fare is provided, together with further information on the time and distance covered by each travel mode. Finally, the data is freely available in bulk through an Application Processing Interface (API), subject to a limit of 300 trip queries per minute (Transport for London, 2019).

However, using it in this study 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 will exclude travel by car or taxi. Fortunately, TfL data (2018) 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 connectivity by car and taxi (and thus will tend to underestimate accessibility levels, particularly in areas with sparser public transport coverage), and do not account for trans-boundary effects (that is, the competition effects of the job and labour pools provided by areas outside the GLA boundary).

To obtain the inter-MSOA travel time data, I submitted requests for the travel options for trips from each MSOA’s population-weighted centroid to every MSOA’s job-weighted centroid. All requests were for trips ending at 0830 on 22 March 2019, reflecting trips that take place during the AM peak period on a typical weekday. To ensure that travel times reflect optimal service conditions in which there are no unplanned disruptions, I submitted the requests over 1 month from January to February 2019.

*Inter-MSOA Travel Time Variability Data*

As the Journey Planner provides travel guidance in real time, I also use it to obtain the distribution of travel times for selected origin-destination pairs over different days. This thus constitutes a measurement of the amount of travel time variability in each of the selected pairs. However, it is important to note that the travel time distributions obtained through this way probably underestimate the real level of travel time variability, due to two reasons.

First, the Journey Planner only partially accounts for the impact of service disruptions on routing options. If there are service suspensions (i.e. no services running) on Underground, Overground, Docklands Light Railway, TfL Rail and tram routes due to disruptions or planned shutdowns, the Journey Planner will not return routing options that require travel along the suspended sections. However, it assumes that walking, cycling, National Rail and bus routes are never suspended. In particular, the Journey Planner will still recommend using a bus service between two stops, even if that service is actually diverted and service to either or both of the stops in question is suspended (the fact that the service is diverted does get flagged in the routing results; however, alternative routing options taking the diversion into account are not provided).

Secondly, the travel times for each option do not account for congestion effects. Travel times incorporate access (the time needed to walk to and from stations or between platforms in stations), waiting and in-vehicle times. Access times are estimated based on assumed walking speeds between and within stations, while waiting and in-vehicle times are calculated based on service schedules. Further, the Journey Planner routing options are planned such that a customer will arrive at the first boarding location on her trip just as a vehicle on the appropriate route is scheduled to arrive. This means that factors such as slower walking speeds due to in-station or sidewalk congestion, increased waiting time due to late vehicle arrival or lack of capacity on vehicles, and increased in-vehicle time due to en route congestion are not reflected.

As such, the travel time distributions obtained from the Journey Planner only reflect two sources of variability: the impacts of severe disruptions that result in service suspensions on TfL rail-based modes, and fluctuations in wait times during transfers between services due to lack of synchronisation between schedules of different services at different times during the day. While this is a limitation of using the Journey Planner data to capture travel time variability, there are certain advantages. Particularly, though the variability captured probably constitutes a lower bound estimate of actual variability, the components captured in the lower bound are especially relevant in ensuring reliable services. Optimising service schedules is a prerequisite for improving travel time reliability, and the outsize impacts of service suspensions on travel times justifies substantial efforts to minimise their occurrence.

To obtain a distribution of travel times per trip, I requested for travel options for selected trips beginning between 0600 and 0800 over 50 weekdays between 29 January 2019 and 25 April 2019. I found that there was only enough time to process about 7,000 trip requests in the two-hour window in real time, so I only gathered data on trips starting from all 983 MSOAs’ population-weighted centroids that end at the job-weighted centroids of 7 MSOAs that correspond to 7 key transport nodes evenly spaced across the GLA area, totalling 6,881 origin-destination pairs. See Appendix XX for details on how the 7 destinations were selected. I varied the start time of each trip every day by randomising the order in which requests for the 6,881 trips were submitted to the API. This allowed the data to capture the variation in schedule synchronisation between different services for each trip at different times during the day, and the resulting distribution of travel times per trip can be thought of as representing the probability distribution of the travel time for each trip for any randomly selected start time between 0600 and 0800.

**Results**

*Distribution of Working-Age Residents by Location and Relative Social Deprivation*

Figure XX visualises the distribution of working-age residents as a dot density map. In general, working-age residents are quite evenly distributed across the GLA area. However, a ring pattern can be discerned, with lower residential densities near the centre, higher densities in a ring around the centre, and decreasing densities as one moves away from the centre.

Of course, residents living in different areas of London have different sociodemographic characteristics. While the resident counts in Figure XX are not disaggregated by characteristics like income, education level or disability status, variations in these characteristics at the aggregate level between MSOAs are reflected in the IMD data. 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) (+ citation), the estimated proportion of households in poverty in 2014 (+ citation), and the unemployment rate among working-age residents in 2017 (+ citation). For all 4 maps, MSOAs are binned into deciles by the distributions of each indicator, with the lowest decile of MSOAs (corresponding to the highest deprivation) visualised in dark red and the highest decile (the least deprived MSOAs) in dark green. In general, the spatial distribution of IMD rankings correspond quite well to the distributions of 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. The Spearman correlation coefficients between the IMD rankings and each of the other 3 indicators are also quite high at around XX. The distribution of working-age residents across MSOAs by IMD rank deciles is remarkably even, with each IMD rank decile containing MSOAs that are home to about 10% of all working-age residents.

*Overall Job Provision Ratios*

Each area’s JPR is a function of 3 components: the spatial distributions of the working-age population and jobs, and the connectivity provided by the public transport system. Together with the distribution of working-age residents, Figure XX also visualises the job distribution as a dot density map. The distribution of jobs is much more clustered than the population distribution, with extreme concentrations 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 network (black lines in the figure, including National Rail, Underground, Overground and Docklands Light Railway networks), most notably at Canary Wharf.

The third component, area-to-area connectivity, is captured in the matrix C. Each value of ci,j in C 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. 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 ), with all areas within 45min travel time from Stratford (i.e. the areas where f(cE02000726,j) = 1) outlined in blue and a dot density map of the distribution of jobs superimposed. In the JPR calculations, each MSOA j within the blue boundary will contribute some of its jobs to each working-age resident in Stratford; exactly how much each MSOA j contributes to Stratford depends on how many other working-age residents live in areas where they can also reach each MSOA j within 45min.

Figures XX, XX and XX visualise the steps needed to calculate the matrix R containing the JPRs for each MSOA. First, Figure XX shows the number of working-age residents who can reach each destination MSOA j within 45min, which is the result of . A dot density map of the distribution of working-age residents is superimposed. It is clear that the public transport system heavily emphasises fast connections to the centre of the GLA area, with central areas being within 45min travel time of the most working-age residents, and the number of residents able to reach each MSOA j within 45min declining as one moves away from the centre. Nonetheless, areas near major rail interchanges towards the outskirts may still be reachable by many residents within 45min, as they benefit from better rail connectivity by virtue of being en route to the centre from points outside the GLA.

Figure XX shows the number of jobs in each MSOA j that is provided to each working-age resident who can reach it, which is the result of . A dot density map of the distribution of jobs is superimposed. This ratio increases as the number of jobs in each MSOA j increases but decreases as the number of working-age residents who can reach it increases. A clear ring pattern can be seen, as in the most centrally located MSOAs, the sheer volume of jobs available offsets the high number of working-age residents who can reach them, while a ring of MSOAs around the centre have few jobs but benefit from good connectivity from areas around the GLA (which spreads the limited jobs available more thinly across larger catchments), and the MSOAs near the outskirts have few jobs but also suffer from poor connectivity from other areas (so that the few jobs available are shared only among small, sparsely populated local catchments).

Finally, Figure XX displays the JPR for each origin MSOA i, which is the sum of all the jobs contributed per working-age resident from each MSOA j within reach. This is the result from . In order to interpret whether an MSOA’s JPR is high or low, we need to compare it to an ideal case in which jobs, workers and connectivity were all evenly distributed throughout the GLA area, which would result in all MSOAs having JPRs equal to , the ratio of the total number of jobs to the total number of working-age residents. The value of in the GLA area is 0.88 jobs per working-age resident. To express the extent to which JPRs deviate from , JPRs are binned into 5 classes: 0 to 50% of (thus the number of jobs allocated is far too low for the number of working-age residents), 50% to 80% of (jobs allocated are somewhat insufficient for the working-age residents), 80% to 125% of (jobs allocated are roughly equal to working-age residents; this is a balanced state), 125% to 200% of (jobs allocated somewhat exceed working-age residents), and more than 200% of (jobs allocated far exceed working-age residents).

From Figure XX, we can see that few MSOAs in the GLA area have balanced JPRs. MSOAs near the centre have too many jobs allocated per working-age resident, while many MSOAs around the outskirts provide 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 JPRs in the west, where Heathrow Airport provides a significant number of jobs, and connectivity between Heathrow and areas in the west is relatively good.

Because the JPRs reflect a distribution of jobs among working-age residents in an excludable way, it is straightforward to construct a Lorenz curve for the distribution of jobs among working-age residents, by sorting the MSOAs in ascending order by JPRs and then summing up the cumulative proportion of the working-age population and the cumulative proportion of all jobs that are allocated to them. This neatly summarises the extent to which the job allocation is uneven. This Lorenz curve is in the inset of Figure XX. Because jobs are highly concentrated in the centre while working-age residents are relatively spread out throughout the GLA area, there is significant unevenness in the job allocation, with the bottom 50% of working-age residents in the more job-poor areas sharing only roughly 20% of the jobs, while the top 10% of working-age residents in the most job-rich areas share 20% of the jobs.

*Analysis of Job Provision by Sector*

The job distribution data disaggregates jobs by Standard Industrial Classification Codes (SICCs). Thus, we can calculate JPRs by sector and analyse the extent to which job accessibility varies by sector. First, I aggregate the SICCs to 6 overarching sectors:

1. Primary and Extractive
2. Industrial
3. Retail, Logistics, Food & Beverage and Hospitality (hereafter consumer services)
4. Finance, Insurance, Real Estate, Professional and Administrative Support (hereafter high-end services)
5. Public Sector
6. Others

The pie chart in Figure XX shows the proportion of all jobs in the GLA area that belong to each sector. 42% of all jobs in the GLA area are from the high-end services sector, while consumer services and the public sector make up 24% and 22% of all jobs respectively. Figure XX plots the distribution of jobs in each sector as a dot density map. The spatial distributions of industrial and consumer services jobs are the most decentralised, while public sector jobs are somewhat clustered around some rail interchanges. Unsurprisingly, high-end services jobs are the most concentrated in the centre and at Canary Wharf.

For each MSOA, I recalculate JPR values using the overall working-age population distribution and the job distribution for each sector, and plot maps of the sector-specific JPRs in Figure XX. This set of JPRs represents the number of jobs from each sector allocated to each working-age resident per MSOA. The top left panel of Figure XX reproduces the map of the overall JPRs from Figure XX for comparison. 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 consumer services jobs are also relatively evenly distributed albeit with significant over-provision at the centre of the GLA area. In contrast, accessibility to high-end services jobs is unevenly distributed, with large swaths of MSOAs in the south, east and north having much fewer jobs from this sector allocated per working-age resident than the GLA average, while MSOAs in the centre are allocated between 200% and 360% as many high-end services jobs per working-age resident as the GLA average. Job accessibility levels to primary & extractive and other sectors are also quite uneven, but there are so few jobs from these sectors in the GLA area that they do not make a significant difference to overall job accessibility.

When we plot Lorenz curves of the sectoral job allocation to the working-age population (lower right panel of Figure XX), we can assess the extent to which the unevenness of the accessibility to jobs in each sector contributes to the unevenness in overall accessibility to jobs. The Lorenz curves for the industrial, consumer services and public sectors are closer to the diagonal than the curve for all jobs, but the curves for the primary & extractive, high-end services and other sectors are below the curve for all jobs. This indicates that the distributions of the latter three sectors tends to make accessibility to jobs more uneven in the GLA area. In particular, because the high-end 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.

Finally, because JPRs allocate jobs in an excludable way, we can decompose each MSOA’s JPR sectorally and analyse the proportion of each MSOA’s overall JPR that is contributed by each sector. This indicates the extent to which each MSOA is dependent on each sector for its job allocation. Figure XX visualises these sectoral dependence proportions, with the MSOAs binned into quintiles for visualisation in each map. Firstly, it is clear that jobs in the primary & extractive and others sectors never make up a significant proportion of any MSOA’s job allocation. Secondly, areas near the outskirts in the northwest, north and east are especially reliant on industrial jobs, areas near the outskirts in the north, east and west (particularly around Heathrow) are especially reliant on consumer services jobs, and two swaths along the northern boundary of the GLA area and south of the Thames are particularly reliant on public sector jobs. Finally, the most centrally located areas are disproportionately reliant on the high-end services sector.

*Analysis of Variability in Travel Times*

For each of 6881 origin-destination pairs, I obtained a distribution of 50 minimum travel times. From this data, I then calculate the RBT for each origin-destination pair. Recall that the RBT measures the extra time that a traveller needs to budget for in order to be 95% confident that she will not arrive late, and thus measures the reliability of travel times.

Each panel in Figure XX plots the RBTs of trips originating in each MSOA and terminating at one of the 7 destinations (outlined in red in each map). To facilitate comparison of RBTs across all origin-destination pairs, all maps use the same binning scheme to visualise the RBTs, with green shades indicating lower RBTs that are under 10min. In general, we can see that areas that are directly connected to the destination by rail tend to have lower RBTs, reflecting the generally high frequency of rail services in the GLA area (which tends to reduce the amount of waiting time needed while interchanging between rail lines). Areas that are not directly connected to a destination by rail tend to have higher RBTs, even if they are relatively near the destination. Trips from such areas tend to require more interchanging, particularly if bus travel is needed, and this increases the probability of increased waiting times due to a lack of schedule synchronisation between different services. Also, trips originating from locations further from a destination tend to have wider intervals, as these trips also tend to involve more interchanging. The probability that a trip is affected by a disruption serious enough to suspend services on a rail-based line also increases as the distance covered by the trip and the number rail-based legs required increases.

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 lower RBTs. This may be due to the more grid-like rail network north of the Thames, which provides commuters with more alternative rail routes if service on a particular section of the rail network is suspended. In contrast, travel times to Lewisham and Streatham tend to be less reliable. This probably reflects two factors: the more tree-like structure of the rail network there (which provides less network resilience against failures along the rail lines leading to these destinations) and reliance on National Rail services to reach these two destinations (these services tend to run less frequently than TfL’s rail-based services, thus increasing the likelihood of increased waiting times during interchanging because of schedule desynchronisation). Finally, for all destinations, we can see that trips originating from the southeastern quadrant of the GLA area tend to higher RBTs. This probably reflects the effects of the less frequent and less reliable rail services run by Southeastern and Southern Railways (+ citation).

*Contributions to Travel Time Variability by Travel Mode*

As we have seen, the level of travel time reliability differs across trips. But why do some trips have less reliable travel times than others? Four factors have already been mentioned: the distance covered for each trip, (as the probability that a disruption occurs at any point along a trip probably increases with the length of the trip), the number of interchanges (which increases the probability of increased wait times due to schedule desynchronisation), the reliability of different travel modes (TfL rail-based services seem to be more reliable than the bus), and location-specific factors (like the extent to which the local network is gridlike rather than treelike, or the operators that provide services in the area).

To investigate how these factors shape travel time reliability across trips, I carry out a set of linear regression analyses with trip-level RBTs as the dependent variable. In an initial model, I use the number of transfers (that is, the number of non-walking legs minus 1) and the distance covered by each mode of transport for the fastest option from the baseline data as covariates – that is, I investigate how the characteristics of the fastest travel option for a trip assuming that all services run as scheduled are correlated with the reliability of travel times for that trip.

The results for this model are shown in Column 1 of Table XX. They indicate that every additional transfer during the trip adds about 0.5min to the RBT. In addition, every additional kilometre travelled by bus or National Rail increases the RBT by 0.4min and 0.06min respectively, indicating that travel by these modes tends to reduce travel time reliability. Although the Journey Planner data do not reflect the effects of congestion and diversions on travel times by bus, bus travel tends to involve more interchanging, and frequencies on bus services tend to be lower than on rail-based services, which increases the risk of schedule desynchronisation during interchanges. The result for National Rail is consistent with the fact that National Rail services are less frequent and may be more prone to serious disruptions than TfL rail-based services. On the other hand, travel by Underground, tram and Docklands Light Railway tends to reduce RBTs.

Because all the MSOAs are connected to each other via the same transport network, there is likely to be a mutual correlation structure between the RBTs of different trips. For instance, two MSOAs that are located along the rail line can be expected to have similar RBTs for trips via that line, as both will be affected by any disruptions along that line. If the non-independence of the RBTs between trips is not accounted for, the correlation structures among different RBT observations may distort the coefficient estimates for the covariates included in the model.

One way to deal with such correlations 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. This captures the idea that areas that are near each other tend to be similarly connected to the larger network of areas and thus should experience similar network-related effects. 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 RBT for trips starting from the 3 neighbours and ending at the same destination.

Figure XX displays the RBT and corresponding local mean RBT among the 3 nearest neighbours for each of the 6,881 origin-destination pairs, and Column 2 in Table XX presents the results for this model. Because of the high correlation between the two quantities, including this local mean as a covariate in the linear regression model significantly improves the fit of the model. The local mean RBT also allows the model to better capture the correlation between the RBTs of MSOAs that arise due to their locations in the transport network. Specifically, as the local mean RBT increases by one minute, the RBT of the MSOA in question tends to increase by 0.5min. The inclusion of the local mean also reduces the confounding of the estimated coefficients for the other covariates due to network effects, reducing the magnitude of the estimates by about 40% for the number of transfers, 35% for bus and 55% for National Rail. However, the coefficients still indicate that bus travel contributes the most to travel time unreliability, followed by National Rail, and that travel by Underground, tram and Docklands Light Railway tend to increase travel time reliability.

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 to contribute particularly significantly to travel time unreliability in trips from the southeast). To investigate this, I 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 as neighbours the 30 MSOAs whose travel times to the 7 destinations are most similar to those of the MSOA in question (using the same Euclidean distance calculation as the derivation of the local mean interval width). Figure XX illustrates 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. As the travel time vectors for each MSOA are unique, the local regression for each MSOA is very likely to use a subset of local neighbours that is unique to that MSOA, and this will thus yield a vector of coefficient estimates that is unique to each MSOA.

First, we examine some diagnostics of the locally weighted regression implementation. Figure XX maps out the adjusted R2 values for the locally fitted regression models. The median local adjusted R2 is 0.23, similar to the adjusted R2 of the model in Column 2 of Table XX, but the local adjusted R2 can vary from as low as 0.05 in the outskirts to 0.69 around Westminster. The local adjusted R2 tends to be higher near the centre where the rail network is densest, but is quite low towards the outskirts, particularly south of the Thames; this indicates that travel time variability in the outskirts is less correlated with the characteristics of the fastest travel option in the baseline database than with other factors that have not been included in the model. Next, Figure XX maps out the p-values for each coefficient estimate for each MSOA. Dark blue areas are where the coefficient estimates obtained were not statistically significantly different from 0. Grey areas are where p-values are missing; this indicates that there was no estimate for the coefficient for that MSOA, because there was no variation in that covariate among the MSOAs considered in the local regression.

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 distributions of local estimates contain the global estimates from Column 2, but capture the variation in each coefficient across space. Figure XX maps out the coefficient estimates for each covariate and each MSOA. Green shades indicate that a covariate contributes tends to increase travel time reliability in that MSOA, while red shades indicate that the covariate tends to decrease travel time reliability. Grey MSOAs indicate where the coefficient estimates were not statistically significantly different from 0 or were not available.

At the local level, increasing travel by bus tends to increase travel time unreliability almost everywhere in London, but particularly in the northeast around Stratford, Barking and Leytonstone, but also around the City, Westminster, Rotherhithe, Lewisham, Sidcup and in a swath from Waterloo to Wimbledon. Travel by National Rail contributes the most to travel time unreliability in a large swath in the southeast, adding further evidence of the effect of unreliability in services operated by Southern and Southeastern Railways. In fact, National Rail services in these areas seem to be so bad that increasing travel by bus and walking in these areas tend to reduce travel time unreliability, contrary to patterns elsewhere. Also, in this area, trips that involve more interchanges tend to be more reliable; this counter-intuitive result may be because trips from these areas that involve more interchanging tend to rely less on National Rail for long-distance travel. However, travel by National Rail also increases travel time unreliability in a northwestern corridor from Paddington to Ruislip, and around Stratford. This could reflect how National Rail lines north of the Thames tend to terminate outside the City (in stations like Paddington, Marylebone, Euston, Kings Cross, St. Pancras and Liverpool Street) and provide poor orbital connections, so National Rail users need to interchange more to complete their journeys to one of the 7 destinations. On the other hand, the local effects of travel by Underground, tram, Overground and Docklands Light Railway on travel time reliability tend to be positive or not statistically significant. Travel by the new TfL Rail service tends to increase travel time unreliability, particularly in the West.

*Typology of London MSOAs by JPR, Sectoral Dependence for Job Allocation, Travel Time Reliability and Mode-Specific Contributions to Reliability*

As we have seen above, there are uneven spatial distributions for JPRs, sectoral dependence for jobs, travel time reliability and mode-specific contributions to travel time reliability. In addition, we know that working-age residents living in different areas of London have different levels of social deprivation. It is possible to develop a typology of MSOAs in the GLA area to examine how these dimensions intersect with each other, to develop a more nuanced understanding of job accessibility and more precisely diagnose problem areas that may warrant additional interventions.

I carry out a k-means clustering analysis using 9 dimensions. Overall JPRs are included to capture overall job accessibility. The percent of each working-age resident’s JPR contributed by high-end services is included to capture sectoral dependence for job allocations, since we have seen that the high-end services sector is the largest single sector in the GLA area and it tends to make the job allocation more uneven. The RBT for travel from each MSOA to the City is included to capture travel time reliability to the most job-rich area in London. Additionally, I identify which of the 6 other destinations are closest to each MSOA in terms of baseline travel time, and then include the RBT for travel from each MSOA to its respective closest destination. In the absence of travel time distributions for every origin-destination pair, this RBT can represent to some extent the reliability of travel times from each MSOA to its immediate neighbourhood. To capture mode-specific contributions to travel time reliability at the local level, I include the MSOA-level coefficients for distance travelled by mode from the locally weighted regression; I consider only the coefficients for bus, Underground, National Rail and walking, as these are the modes where coefficient estimates are missing for no MSOAs. And finally, I also consider the IMD rank for each MSOA to capture the sociodemographic characteristics of the working-age population.

The results for a clustering with 8 clusters is shown in Figure XX. By social deprivation, three types of clusters are apparent: Clusters 0, 1 and 2 with higher IMD ranks reflecting lower deprivation, Clusters 5, 6 and 7 with lower IMD ranks reflecting higher deprivation, and Clusters 3 and 4 that contain MSOAs with IMD ranks from across almost the whole spectrum.

Cluster 3 (with about 8% of the working-age population) contains the most centrally located MSOAs in the City, Westminster and Southbank. As these areas are the most job-rich, MSOAs in Cluster 3 are distinguished by having the highest JPRs, indication extreme surpluses of jobs allocated per working-age resident, and among the highest dependence on the high-end services sector for jobs. Their central location gives them access to many public transport routes, especially rail-based routes, resulting in extremely reliable travel times to the City and to their local destinations. Reliability of travel by mode tends to be average, except for travel by National Rail, which seems to be less reliable. This could be due to scheduling desynchronisation between services by National Rail and other modes. In contrast, the MSOAs in Cluster 4 are distinguished by their combination of job poverty (low JPR values indication insufficient jobs per working-age resident) and high travel time unreliability, especially to the City. This unreliability is disproportionately due to the unreliability of National Rail services in the southeast, so much so that trips that involve more distance covered by bus, walking or Underground (and thus rely less on National Rail) tend to be more reliable. Fortunately, only about 2% of the working-age population lives in MSOAs in Cluster 4.

Of the three least socially deprived clusters, Cluster 0 (with 15% of the working-age population mostly located in the southwest in areas like Richmond, Kingston, Wimbledon and Battersea) tends to have the most balanced JPRs, indicating sufficient jobs allocated per working-age resident. These working-age residents tend to rely more on high-end services for their job allocations, but because of their low social disadvantage, they may be more likely to have the skills to take advantage of jobs in this sector. Travel times to the City and to local destinations tend to be quite reliable, and indeed this area seems to enjoy the most reliable Underground and National Rail services in London. Cluster 1 (with 23% of the working-age population and covering areas near the outskirts) also enjoys relatively reliable travel times to the City and local destinations, but experience a job shortage due to their greater distance from the centre and the relative scarcity of jobs in the outskirts. In contrast, Cluster 2 (with 10% of the working-age population) experiences both insufficient jobs and high travel time unreliability. In these areas near the outskirts, travel by National Rail seems to especially contribute to travel time unreliability.

The situations facing the most socially deprived clusters are quite interesting. Cluster 5 (with 18% of the working-age population) are located quite centrally, almost surrounding the MSOAs in Cluster 3. As a result, MSOAs in Cluster 5 enjoy similar conditions as Cluster 3, benefitting from excellent transport connectivity, high travel time reliability and a surplus of jobs allocated per working-age resident. However, these jobs also tend to be from the high-end services sector, which more socially disadvantaged people (who may lack educational qualifications, relevant working experience or social networks) may find more difficult to take advantage of. Of course, the high-end services sector does contain some jobs suitable for workers from less privileged backgrounds, but the degree to which the sheer volume of jobs available can compensate for the tendency for skills mismatch between high-end services jobs and more socially disadvantaged workers is unclear from this analysis.

MSOAs in Cluster 5 (containing 18% of the working-age population) are located further towards the outskirts, especially in the Lea Valley in the north, areas from Ealing to Hayes & Harlington along the TfL Rail line in the west, areas along the Overground line from Brockley to Croydon in the south, and Becontree and Dagenham in the east. They are characterised by relatively reliable travel times but a shortage of jobs allocated to workers. Travel seems to be reliable but too slow to connect these areas effectively to large job clusters. In contrast, MSOAs in Cluster 6 (containing 18% of the working-age population in Stratford, Ilford, Barking, the Royal Docks and Greenwich) have relatively balanced JPRs and reliable travel times to the City, but less reliable travel times to local destinations. The reliability of buses and National Rail in these areas seems especially low. These areas are also more dependent on the high-end services sector for jobs. Thus, workers living in these areas may be more vulnerable to skills mismatch with the available jobs.

In summary, the k-means clustering analysis allows us to develop a typology of MSOAs in the GLA area that captures a more nuanced understanding of job accessibility that expands on the usual perspective that is focused on job-worker spatio-temporal matching to also consider sectoral dependency for job allocations, travel time reliability and worker sociodemographic characteristics. This analysis reveals that besides simply having to many or too few jobs allocated to each working-age resident within a reasonable travel time, each area may also face challenges in terms of skills mismatch and travel time variability. The particular combination of challenges in each area suggests that intervention strategies should be tailored to each area’s characteristics.

**Discussion**

*Results in the Context of Extant Literature*

The analysis of job accessibility in the GLA area, as captured by MSOA JPRs, indicate that job accessibility is unevenly distributed across space, with locations nearer the centre and near major public transport network nodes having more jobs allocated per working-age resident, and locations towards the outskirts and with poorer connections to the public transport network having too few jobs per working-age resident. Other studies of Boston (Shen, 1998), Montreal (El-Geneidy, et al., 2016), Santiago (Neihaus, et al., 2016), Bogota (Guzman, et al., 2017; Guzman & Oviedo, 2018) and 11 major Canadian cities (Deboosere & El-Geneidy, 2018) also found similar patterns where job accessibility is highest where the public transport network is densest. This is an unsurprising finding, as the cost of building and operating public transport infrastructure means that connectivity cannot be evenly provided everywhere, and needs to be prioritised to connect areas that generate the most trips (usually high-density housing areas to job-rich city centres).

The analysis of each MSOA’s sectoral dependency for its job allocation gives some insight into how accessibility varies by job characteristics. In particular, more central areas are more dependent on high-end services for jobs, while areas nearer the outskirts are less dependent on any single sector. This contrasts with Shen’s (1998, p. 355) findings for Boston in 1990 that accessibility to jobs in manufacturing, administrative support, sales, services and agricultural sectors for low-income workers by car and public transport was highest in the city centre. This contrast may be accounted for by the structural differences in the transport networks of the two cities, where Boston’s public transport system is less extensive than London’s, and Boston’s road networks are more effective than London’s at providing fast connections to the city centre from the outskirts. It may also be accounted for by 3 decades of deindustrialisation and specialisation in central business districts, where high-end services firms have increasingly priced out firms in other sectors from the most central locations.

The measurements of travel time reliability for the 6,881 origin-destination pairs confirm the findings from Uniman (+ citation) and Ehrlich (+ citation) that travel times in the public transport network can be quite variable, even if the system operates entirely as planned. However, my RBT estimates seem lower than what Uniman found. For example, Uniman (+ citation) reports that a trip from Waterloo to Canary Wharf in 2007 had an RBT of 9min. While I did not collect travel time variability data for trips ending at Canary Wharf, the RBTs I measured for trips from Waterloo to Lewisham or Ilford (the two destinations with measurements that are closest to Canary Wharf) are 4min and 1min respectively. Also, Uniman (+ citation) found that RBTs for trips along the Victoria Line were around 8min, while I found that the RBTs for trips with origins along the Victoria Line are generally less than 5min. This is understandable, as the data I use to measure travel times in real time do not account for all sources of travel time variability, so the trip-level RBTs I calculate probably underestimate the real level of travel time unreliability. Nonetheless, my results still indicate that travel times by bus are significantly less reliable than travel times by rail-based modes. This is in line with findings from XX, XX and XX.

The k-means clustering analysis indicates that there are locations in the GLA area where more socially disadvantaged working-age residents suffer from a combination of insufficient accessibility to jobs, dependence on the high-end sector for jobs and travel time unreliability, especially by bus and/or National Rail. But there are also areas where socially disadvantaged workers benefit from high job accessibility and travel time reliability. This is similar to El-Geneidy et al’s (2016) and Deboosere & El-Geneidy’s (2018) findings in Canadian cities that indicate that low-income workers tend to have good accessibility to jobs that are suitable for them, although problem areas where social disadvantage coincides with low accessibility exist. In contrast, studies of South American cities (Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018) found that there is a systematic tendency for low-income workers to live in less accessible areas; this is probably due to their lack of affordable housing near job-rich centres and less extensive public transport systems.

*Recommendations for Urban Developments in the GLA*

One of the aims of this study is to obtain findings that can aid in developing multifaceted strategies to improve accessibility to suitable jobs in London, drawing on options from land use development, transport infrastructure and operations and economic and social policy. Below, I list some ways in which the findings discussed above can aid decision-makers seeking to improve job accessibility in London.

Firstly, the map of overall JPRs in Figure XX reveals locations where there are too few or too many jobs allocated per working-age resident. This can directly guide the distribution of future growth to shape the overall urban structure, with new housing supply being directed to locations with too many jobs per resident, and job growth being encouraged in areas with too few jobs per resident. Transport connectivity between areas with high JPRs and low JPRs can also be enhanced, to put more residents in job-poor areas within reach of areas with job surpluses. This may be especially effective where adjacent areas have very different JPRs.

Secondly, even though the measurement of travel time reliability in this study underestimates actual travel time unreliability, the results are still useful to identify areas where travel times to important destinations like the City are relatively unreliable. The mode-specific reliability analysis also indicates areas where services by different modes seem to be particularly unreliable. These results can be used to focus efforts to improve service operations, such as adjusting service schedules or increasing service frequencies to reduce the probability of schedule desynchronisation during interchanges, or adjusting the incentives for operators so that they put more effort into maintaining services on schedule.

Thirdly, the typologies of MSOAs derived from the K-means clustering analysis can be used to target social and economic policies to reduce the potential for skills mismatch between workers and available jobs. For instance, referring to Figure XX, the working-age residents of MSOAs in Clusters 5 and 7 may especially benefit from skills training programmes, which can help bridge the gaps between their relatively high social deprivation and dependence on high-end services jobs. Economic incentives to encourage job growth in other sectors, especially consumer services and the public sector, can also be targeted to MSOAs in these clusters.

Finally, the MSOA typology can also be used to prioritise interventions to benefit the most socially deprived areas first, especially since more socially deprived residents may have fewer resources to improve the accessibility situations that they face by themselves, and may have fewer options for travel besides public transport. In order to deliver results quickly, low-cost and easily implementable interventions that make the most out of existing resources and programmes can be found.

To a certain extent, these recommendations are in line with what the GLA already plans to do to improve transport connectivity and distribute growth more equitably throughout London. For instance, the upcoming Elizabeth Line is expected to greatly improve connectivity from the east and west of London to the centre, placing more workers within 45min travel time of the job-rich centre, and this should help to move the JPRs in areas along the route towards a more balanced state. The Elizabeth Line may be particularly important in improving job accessibility for areas in Dagenham and Becontree in the east, where a shortage of jobs coincides with higher social deprivation.

A similar effect can be expected from the proposed Crossrail 2 project, which runs from the north through the socially deprived and job-poor Lea Valley to the southwest (+ citation). The Mayor’s Transport Strategy (Greater London Authority, 2018) also calls for improvements to bus connectivity and reliability (pp.155-9), improvements to rail reliability south of the Thames (particularly through the conversion of National Rail services to Overground services through “metroisation”) (pp.180-1), and the encouragement of housing and job growth in areas with good connections by public transport (p.214). The findings from this study can add further justifications for pursuing these strategies.

Nonetheless, the findings can aid in prioritising areas for the implementation of these strategies. For instance, the leg of Crossrail 2 that runs north of the Thames benefits more deprived areas than the southern leg, so this could be a reason to prioritise the implementation of the northern leg. Also, the Transport Strategy (2018) recommends encouraging job growth in four strategic transport interchanges, Lewisham, Stratford, Clapham Junction and Willesden Junction, to put more jobs within easy reach of workers. This is a sensible strategy, as these four locations are surrounded by areas with low JPRs (from Figure XX). However, Figure XX also shows that the number of working-age residents that can reach Lewisham within 45min travel is relatively low, and the RBTs for travel to Lewisham (shown in Figure XX) are quite high; this indicates that travel speeds and reliability to Lewisham should be improved in tandem with the delivery of more jobs there. Similar RBT measurements are not available for Stratford, Clapham Junction and Willesden Junction, but all seem to be in areas where the locally weighted regression indicated buses are relatively unreliable, and Stratford and Willesden Junction are in areas with elevated National Rail unreliability, so similar strategies that deliver transport improvements with job growth may be applicable to these other 3 proposed centres as well.

*Recommendations for Further Research*

While this study has attempted to analyse job accessibility from multiple dimensions, the phenomenon of accessibility is complex enough that many more avenues remain to be explored. In particular, I list several ways in which future studies can better measure travel time reliability using real-time trip planning data, and how job accessibility measures like JPRs can explicitly integrate travel time reliability.

Firstly, this study has demonstrated how real-time trip planning data can be used to observe day-to-day travel time variability without having to build a full multimodal transport model in which to simulate trips and without having to access highly protected and sensitive datasets like automated fare collection records. However, the Journey Planner data does not reflect the effects of congestion on wait times and travel speeds, and thus measurements using this data tend to understate actual levels of variability. To a certain extent, this can be remedied by leveraging on other TfL APIs, particularly the Live Arrival Times API. For instance, after Journey Planner API results are obtained and the fastest travel option for a trip is identified, a script could be launched to request for live arrivals at stops along the route so that the overall trip time can be updated to reflect actual waiting times for vehicle arrivals. This will help to incorporate the effects of en route congestion on waiting times and in-vehicle travel times, though it will still not account for the additional waiting time that would result from the inability to board the vehicles that arrive due to in-vehicle congestion.

Secondly, instead of using the limited querying capacity to observe travel times from all 983 origins to 7 selected destinations, queries could be made for the 7,000 most travelled origin-destination pairs. The latter option would measure travel time reliability for the trips that affect the most commuters, and the regression models analysing the correlation between RBTs and trip characteristics may achieve better fits with this data. This in turn may enable us to predict the RBTs for trips without actual observations with better accuracy.

If a good model of RBTs is achieved, then this could enable us to incorporate travel time reliability directly into the calculation of JPRs. Specifically, rather than just using the ideal travel time between each origin-destination pair in the travel cost matrix C, we would add each trip’s predicted RBT to its ideal travel time, which would then give us the minimum travel time that a traveller needs to provide for a trip to be at least 95% certain that she will not arrive late. Some areas which may be accessible under ideal conditions may no longer be reachable once the RBT is provided for. This would change the number of working-age residents who can compete for each job and the areas that will contribute jobs to residents in each area, thus shifting the job accessibility picture.

Another angle is also worth exploring, namely the setting of the travel time cut-off. In this study, a 45min cut-off is used in line with TfL’s own analyses. However, travel time cut-offs can also be chosen in other ways. Firstly, it can be set to be in line with explicit policy goals, which may be expressed as maximising the accessibility to opportunities within a certain travel time cut-off; in this case, concerns that a sharp cut-off does not reflect actual behaviour can be countered by its relevance as a policy goal. Secondly, it is possible to set different cut-offs for different locations, to better reflect how the willingness to travel may differ contextually. As long as a n x n adjacency matrix C can still be calculated, it will still be possible to calculate JPRs and to maintain the ability to interpret them straightforwardly as ratios.

# **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.