* Literature Review
  + Job accessibility measures
    - Findings of overall job accessibility
  + Adding dimensions to job accessibility analysis
    - Workers’ characteristics: accessibility for subsets of workers, or accessibility by area-level variations in workers’ characteristics
    - Job characteristics: accessibility by subsets of jobs – wage levels or industrial sector
    - Connectivity
      * Accessibility by mode
      * Accessibility over time
        + By time period
        + Over various days: reliability
  + Current study: job accessibility with additional dimensions of jobs by sector, workers by area-level aggregate deprivation, and travel time reliability
* Methodology & Data Sources
  + Job accessibility measure
    - Limitations of Reach
    - JPR calculation method
    - Data sources
      * Job and worker counts
      * Connectivity data
  + Jobs by sector
    - Data source
    - Method: sector-specific JPRs and Lorenz curves, and sector dependency
  + Workers by area-level aggregate deprivation
    - Data source & aggregation method
  + Travel time reliability
    - Data source & caveats
    - Descriptive analysis
    - Modelling of factors that drive reliability
      * Linear regression
      * Spatially autocorrelated regression
      * Locally weighted regression
  + Synthesising findings into a typology of neighbourhoods with k-means clustering
* Results
  + Descriptive summary of job & worker distributions
  + Overall job accessibility by JPR
  + Sectoral JPRs, sectoral contributions to unevenness in JPRs, and sectoral dependency
  + MSOA IMD rankings
  + Travel time reliability
    - Descriptive analysis of variations in RBTs over space
    - Regression results
  + K-means clustering
    - Description of neighbourhood typologies
* Discussion
  + Limitations and further research
    - Capturing travel time reliability more accurately
    - Integrating travel time reliability directly into job accessibility measures
  + Recommendations for urban development in London

**Introduction**

A city’s level of vibrancy, productivity, efficiency and innovation depends on how effectively it can connect job opportunities and workers to each other through economic, social, intellectual and physical networks (Bettencourt & West, 2010). One important aspect of job accessibility (that is, the ease with which workers can take advantage of a suitable range of job opportunities) (Hansen, 1959) is the spatial and temporal distance between workers and jobs. An approach to urban planning that integrates land use distribution and transportation strategies can minimise this distance by optimising the distribution of homes and employment relative to the transport network’s connectivity, so that the available transport capacity can be used most efficiently to maximise the range of job opportunities that workers can access within reasonable travel times.

However, it is also important to keep in mind that job accessibility varies across many dimensions besides space and time (Lucas, 2012). In particular, accessibility levels are affected by the characteristics of workers and jobs (for example, depending on their skills, salary and industrial sector, not every job is equally attractive or suitable for every worker) and by day to day fluctuations in transport network performance (travellers need to account for a margin of uncertainty around potential arrival times when deciding how and when to travel).

This study aims to carry out an analysis of job accessibility by public transport in London that reflects its multidimensional nature. I develop a typology of neighbourhoods that categorises them by overall job accessibility, the degree to which they depend on particular economic sectors for jobs, the relative social deprivation of their working-age population and travel time reliability. 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 efforts to improve workers’ skillsets, encourage job growth in certain sectors and improve the reliability of public transport operations, and can help identify areas where interventions should be prioritised, thus allowing for closer coordination of strategies across a wider range of domains.

**Literature Review**

Hansen (1959) defines accessibility as the potential for interaction between people and the opportunities that they seek. Though people need many types of opportunities besides jobs, the literature on urban accessibility has devoted significant attention to accessibility to jobs in particular. Commutes to work form a significant proportion of all daily trips, especially during the AM peak period, so these trips play a major role in shaping the layouts and operations of transport systems. 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).

Job accessibility refers to the range of job opportunities that a person can choose from. This is a function of the spatial distributions of potential workers’ home locations and jobs on the one hand, and the connectivity (or cost of travel) between different locations on the other. Many indicators have been developed to study accessibility (Geurs & Wee, 2004), but “location-based” indicators are most frequently used for city-level job accessibility 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, studies of job accessibility using these measures have found wide within-city variations (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), with accessibility being highest in central areas where jobs are most concentrated and transport networks are densest.

Besides analysing the spatial patterns in overall job accessibility, researchers have also examined how accessibility varies by different types of workers, jobs and connectivity. To capture how accessibility varies by worker characteristics, some researchers separately calculate accessibility indicators for subsets of workers of different types (for instance, by considering the distribution of low-income workers only rather than all workers) (Shen, 1998; Deboosere & El-Geneidy, 2018), or examine how accessibility levels vary by area-level indicators of sociodemographic characteristics (like the average income or unemployment rate at the neighbourhood level) (Sanchez, 1999; Neihaus, et al., 2016; Guzman, et al., 2017; Guzman & Oviedo, 2018). These analyses have found that most cities have areas where high social disadvantage and poor accessibility to jobs coincide, though this seems to be a more systematic problem in poorer cities with less extensive public transport systems.

To study how job accessibility varies by job type, researchers disaggregate job distributions by economic sector (Shen, 1998) (+ other citations) or wage level (Deboosere & El-Geneidy, 2018), and find that the spatial patterns of accessibility by job type depend on the extent to which job clusters are specialised. In areas where jobs of similar economic sectors tend to cluster together and each cluster tends to be sectorally specialised, there may be areas with high overall job accessibility but low sectoral diversity, raising the possibility that the jobs available may not match the capabilities or characteristics of the workers nearby (+ citation).

Researchers also examine how accessibility varies by different types of connectivity. In general, these efforts fall into two categories. The first are studies that compare connectivity by different transport modes (Shen, 1998; Kotavaara, et al., 2012; Ford, et al., 2015; Neihaus, et al., 2016). They find that job accessibility by private transport is vastly superior to the accessibility by public transport. The second are studies that compare connectivity across different time periods (such as the AM peak and evening off-peak periods), to capture how differences in travel times throughout a day (for instance, because public transport services run less frequently outside of peak periods or because congestion slows down car travel during peak hours) can impact accessibility (Tenkanen, et al., 2016; Järv, et al., 2018; Hu & Downs, 2019). These find that accessibility varies considerably by time of day and by location.

However, connectivity for the same period (such as the AM peak) also varies across different days. Travellers have to account for a margin of uncertainty in travel times when planning how to travel each day, since a trip on any given day may be affected by unforeseen events like congestion, vehicles arriving late or breakdowns in the transport network (Katz & Quealy, 2019). This angle of travel time reliability is under-studied in the job accessibility literature, but has received increasing attention in the transport studies literature in recent years, driven by the increasing availability of large-scale transport data (such as automated fare collection records, automated vehicle location traces or real-time trip planning) that tracks the amount of time needed to complete a trip in real time (+ citation). These studies have proposed new measures of travel time reliability (+ citation Uniman) and found that reliability tends to be low for long trips, those using less frequent services and for bus routes (+ citations).

In the context of London, Ford et al (2015) show that job accessibility varies significantly across space, but as far as I am aware there has not been a published analysis of job accessibility that accounts for additional dimensions like different job and worker types and travel time reliability. Nonetheless, both the UK government (Social Exclusion Unit, 2003) and the Greater London Authority (GLA) (2018) recognise that low accessibility can intersect with social disadvantage, hindering less well-off 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. Transport for London (TfL) (2018) and the GLA (2018) also recognise the importance of improving travel time reliability in the public transport network, especially on bus and certain rail routes.

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 multidimensional 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 reliability 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.

**Methodology and Data Sources**

*Measuring Job Accessibility with Job Provision Ratios*

Several types of location-based job accessibility measures exist, with different levels of sophistication. The simplest type reports accessibility in terms of “cumulative potential” (Geurs & Wee, 2004, p. 133), summing up the total number of jobs that can be physically reached from each area within a certain catchment (usually defined as the area that contains all points that can be reached within a certain travel cost). Essentially, this is the dot product of 2 components, , where:

1. J is a column vector of length n:

where jd is the number of jobs in one of the n destinations d, and;

1. C is a m by n incidence matrix of a two-mode network consisting of weighted edges between m origin nodes and n destination nodes:

where co,d is the distance between one of the m origins o and one of the n destinations d, and f(co,d) is a function that decreases as co,d increases. Thus, f(co,d) indicates the strength of the connection between o and d. In principle, co,d can encode any distance metric (for instance, co,d can be the weighted mean between the physical distance from o to do and the Euclidean distance between the sociodemographic characteristics of the workers living in o and the workers working in d, so that f(co,d) increases as physical distance between o and d decreases and as the sociodemographics in o and d become more similar), but in practice, co,d only encodes spatial distance (in the case of this study, travel time) between o and d.

This measure is easy to calculate but is theoretically flawed, as it 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 drawback, adjustments were proposed by Shen (1998) and elaborated in the public health literature as the “two-stage floating catchment area” method (Luo & Wang, 2003; Luo & Qi, 2009; Wang, 2012) to account for:

1. The spatial distribution of working-age residents (or potential workers), stored in a column vector W of length m

where wo is the number of potential workers in one of the m origins o.

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

where gives for each destination d the weighted sum of the number of potential workers whose catchments include d, divides element-wise the number of jobs in each destination d among the previous weighted sum of potential workers, and is a weighted sum of the number of jobs provided to each potential worker over all destinations d within the catchment of each origin o.

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, ro (the accessibility value for each origin o) accounts for both connectivity and competition effects: as the number of jobs within o’s catchment increases, ro will increase, but as the number of potential workers who can reach the same catchment (and thus compete for the same jobs) as o increases, ro will decrease. This property is especially useful for assessing whether the local supply of jobs is sufficient to satisfy the local demand. Methods to assess the degree of inequality in distributions such as Lorenz curves can also be straightforwardly applied, by first calculating for each MSOA o how many jobs are allocated to it, which is given by , and then sorting the MSOAs in ascending order by JPRs and plotting the cumulative proportion of the potential workers against the cumulative proportion of all jobs that are allocated to them.

The use of this measure requires two choices. The first is a choice of which level of spatial aggregation to use. Ideally, since each value of co,d measures the travel cost between two points, point-level data should be used. However, the unavailability of population and job distribution data at the address point level in the public domain necessitates some level of spatial aggregation. Then each value of co,d represents the travel cost between area o and and area d by the travel cost between the two areas’ centroids. Some studies divide the study area using a grid of equally-sized squares, as this results in evenly-spaced centroids that are thought to reduce distortions in the travel cost calculations (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) to divide up my study area into pre-existing spatial units; specifically, the m origins and n destinations in this study are, respectively, the population-weighted and jobs-weighted centroids of the 983 Middle-layer Super Output Areas (MSOAs) within the GLA boundary. MSOAs 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 or smaller.

The second choice involves how to define the function . Many studies use continuous 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 (+ citation Chen & Jia). 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 that 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. A contour measure undoubtedly distorts reality, as it treats all locations where as if they were equally attractive and 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 at any rate, El-Geneidy et al (2016, p. 303) and Chen & Jian (+ citation) found that contour and distance decay-based measures are highly correlated. Most importantly, a contour measure maintains the raw accessibility values as simple ratios that can be interpreted directly (Luo & Wang, 2003; Deboosere & El-Geneidy, 2018). Specifically, this allows the quantities in the matrix R to be understood as ratios of the number of potential jobs allocated to each potential worker within each their catchment (viz. Hansen, 1959). I will term these quantities Job Provision Ratios (JPRs).

Having decided to use a contour measure, the choice of X then becomes significant, as larger values of X tend to result in greater extents of smoothing out of local unevenness in accessibility (+ citation). I follow Chen & Jia’s (+ citation) advice to choose X using empirical evidence of people’s actual willingness to travel and to reflect policy priorities. In 2017, a survey by the Department for Transport (+ citation, table tsgb0110) found that people who worked in London had an average commute time of 46min. TfL (2018, p. 224) also uses 45min as a travel time threshold for calculating job accessibility measures. As such, I will set X to 45min for this study (see Appendix X for an analysis of how sensitive my results are to other values of X).

*Job, Worker and Connectivity Data Needed for JPR Calculations*

Three components are needed to calculate the column vector R. The first two, MSOA-level counts of working-age residents (those aged 16 to 64) and jobs, are easily obtainable from the Office of National Statistics (ONS) (+ citation). I use the most recent data available, which is for 2017.

The third component is a matrix of inter-area travel costs, which is needed to build the matrix C. 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 for every origin-destination pair 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 to build a completely correct graph representation of the transport network, and is sensitive to the design of the routing algorithm.

An alternative that has become feasible in recent years is to use automated public transport fare collection data to obtain the mean or median travel time between any pair of origin and destination stations. This is especially effective for systems where travellers must validate their payment media upon both entrance and exit. However, a general limitation of this approach is that the cost of first- and last-mile travel (from the trip origin to the origin station, and from the destination station to the final destination) is not reflected in the data, and more modelling is required to fill in these gaps. Two other factors also limit its applicability in the context of London. First, while TfL’s automated fare collection system captures the locations of (most) travellers upon boarding a bus or entering and exiting a rail station, the alighting locations of all bus passengers (about 45% of all boardings of TfL vehicles) (Transport for London, 2018, p. 29) are not recorded. While the missing alighting locations can be imputed using modelling to some extent (+ citation), this may introduce systematic distortions into the data. Second, the raw fare collection 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).

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 or motorised public transport like buses, trams or trains (but not by car or taxi)[[1]](#footnote-2). This data source has several compelling advantages. First, it allows me to obtain travel time 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 guidance covers both first- and last-mile travel and travel within the public transport network. Thirdly, the data is highly disaggregated: besides information on the overall fare, distance and time needed for the trip, time and distance are separately reported for each leg of the trip. 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 the study 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 outside the GLA boundary).

I used the Journey Planner to build a “baseline” database of inter-MSOA public transport travel times during a typical weekday AM peak, assuming all services run as scheduled. To do this, I submitted requests for information on trips from each MSOA’s population-weighted centroid to every MSOA’s job-weighted centroid. All requests were for trips ending at 0830 on Friday, 22 March 2019, so that the trips have to occur during the AM peak, but the requests were submitted over 1 month from January to February 2019. As the requests were made far in advance of the intended travel date, they do not include the influence of actual failures in the network on 22 March 2019.

*Disaggregated Job Distribution Data*

The ONS reports not only the number of jobs in each MSOA, but also the breakdown of these jobs by Standardised Industrial Classification Codes (SICCs). This enables me to disaggregate job accessibility into 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

Then, I calculate sectoral JPRs by replacing O with O\*, a column vector of length n \* p:

where os,d is the number of jobs in sector s in MSOA d, and by replacing C with C\*, a m by (n \* p) matrix of travel costs:

where co,s,d is the time needed to travel from MSOA o to the jobs in sector s in MSOA d. I assume that the within-MSOA sectoral job distributions are all identical, such that co,s,d = co,d for all s.

Recall that JPRs allocate jobs in an excludable way. Thus, Lorenz curves of the sector-specific job allocations over all potential workers can be constructed and compared to the Lorenz curve for the overall job allocation, to see the extent to which the spatial distribution of each sector’s jobs contributes to the unevenness in the overall job allocation. Also, the sum of each MSOA’s sector-specific JPRs will be equal to its overall JPR, that is . This allows me to calculate , which is the proportion of each MSOA o’s JPR that is contributed by jobs in each sector s, which captures the extent to which MSOA o is dependent on sector s for jobs. This can help to illuminate two phenomena. First, high levels of dependence on a single sector for jobs may make an MSOA’s working-age residents more vulnerable to sector-specific shocks like sudden disruptions to its labour market. Secondly, if an MSOA’s working-age residents’ sociodemographic characteristics do not match the typical profile of employees in a sector, then increasing that MSOA’s dependence on that sector may increase the probability of skills mismatch between potential workers and the jobs that are available to them (+ citation).

*Capturing Variations in Characteristics of the Working-Age Population*

I also wish to investigate how job accessibility varies across different types of potential workers. Unfortunately, counts of residents that are disaggregated by sociodemographic 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. Thus, the type of disaggregated analysis that I can do for jobs is not as feasible for potential workers.

In the absence of suitably disaggregated population count data, many researchers use location-based aggregate 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 deprivation experienced by residents at the Lower Super Output Area level 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 variations in social deprivation among the working-age population. Following the DHCLG guidance, I aggregate the IMD scores to MSOA level using an average of the scores for LSOAs contained in each MSOA weighted by each LSOA’s 2012 population. I then focus on the relative ranking of the areas by IMD index scores, rather than on the raw scores themselves. Using this data, I can then identify MSOAs where residents experience both high social deprivation (indicated by low IMD ranks) and poor job accessibility.

*Inter-MSOA Travel Time Reliability Data*

As the Journey Planner provides travel guidance in real time, I also use it to obtain the travel times for selected origin-destination pairs over different days. The overall distribution of travel times per pair constitutes a measurement of its travel time reliability. However, it is important to note that the travel time distributions obtained through this way probably underestimate the real level of travel time variability.

Journey Planner 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 based on assumed walking speeds between and within stations, and do not reflect in-station crowding that may reduce walking speeds around and in stations. Waiting and in-vehicle times are calculated solely on service schedules and do not reflect live vehicle locations. In particular, as long as vehicles are running along a route, the Journey Planner assumes all vehicles are running as scheduled, even if they are in fact delayed due to breakdowns or en route congestion. Also, the Journey Planner assumes that travellers will always be able to board the first suitable vehicle that arrives, even though in-vehicle congestion may prevent boarding.

Finally, 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 rail routes (National Rail, Underground, Overground, Docklands Light Railway (DLR), TfL Rail and tram) due to disruptions or planned shutdowns, the Journey Planner will not return routing options that require travel along the suspended sections. However, if services are just delayed but not completely suspended, the Journey Planner will still continue to recommend routes via the affected services as if service schedules were not affected (albeit with an additional alert that service frequencies are affected by delays). Also, it assumes that walking 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 (if the routing results involve a service that is diverted, an alert describing the extent of the diversion is returned with the results; however, alternative routing options taking the diversion into account are not provided).

As such, the travel time distributions obtained from the Journey Planner only reflect two sources of variability: the need to reroute travel to avoid service suspensions on rail routes, and fluctuations in wait times during transfers between services due to a lack of synchronisation between schedules of different services at different times during the day. Nonetheless, these two sources of variability are arguably of special interest to transport operations planners, as rail service suspensions have an outsize impact on travel times, and service schedules are one of the few aspects of transport operations that planners can directly control.

To obtain a distribution of travel times per trip during the AM peak on a typical weekday, I requested for travel information on 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 in the two-hour window to process about 7,000 trip requests 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 points during the weekday AM peak, and the resulting distribution of travel times per trip represents the trip-specific probability distribution of the travel time for any randomly selected start time between 0600 and 0800 on a typical weekday.

With the trip-level travel time distributions, I can calculate what Uniman (+ citation) defined as the Reliability Buffer Time (RBT), which is the difference between the 95th percentile travel time (for an unusually slow trip) and the median travel time (representing a typical trip). The RBT captures the additional travel time that a traveller needs to budget for in order to be 95% confident that a trip on any given day will not end late.

Because I also have baseline data for each trip that lays out its characteristics assuming ideal conditions during the AM peak, I can analyse how RBTs vary according to the baseline characteristics of each trip. Specifically, I will fit a set of linear regression models, the first of which is:

where:

* Y is a column vector containing 6,881 values where yi is the RBT for trip i;
* T is a column vector of 6,881 values where ti is the number of transfers between vehicles (that is, the number of non-walking legs minus 1) needed for trip i assuming the trip ends at 0830 on a typical weekday and that all services run as scheduled;
* D is a 6,881 by m matrix where m is the number of travel modes and dk,i is the total distance covered by mode k for trip i assuming the trip ends at 0830 on a typical weekday and that all services run as scheduled;
* α is an intercept term, β is a scalar representing the partial effect of transfers on RBT, γ is a m x 1 column vector representing the partial effects of distance covered by each mode on RBT, and ε is an idiosyncratic disturbance.

This model investigates the impact on a trip’s travel time reliability resulting from the number of transfers (each transfer is an opportunity for travel times to increase due to schedule desynchronisation between services) and distance travelled by mode (so-called mode-specific effects).

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 same 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 (+ citation). 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 o 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 o, I identify the 3 other MSOAs (o1, o2 and o3) where this Euclidean distance is minimised; these are the MSOAs whose travel times to the 7 destinations are the most similar to those of MSOA o, assuming all services run as scheduled. Then, for each trip i originating from MSOA o and terminating at each of the 7 destinations d, I calculate the local mean RBT lo,d for trips starting from each neighbour o1, o2 and o3 and ending at the same destination. That is:

For an illustration of where the 3 neighbours are located for the case of one MSOA, Stratford, see Figure XX.

With the local mean RBTs, I then fit the following model:

where L is a column vector with 6,881 values where li is the local mean RBT for trip i and δ is a scalar representing the partial effect of local mean RBTs on the RBT.

While the previous two regression models generate one global estimate for the partial effects of each covariate on the RBT of a trip, there are indications that mode-specific effects vary across locations. For example, it is widely reported that National Rail services in the southeast of London are particularly unreliable (+ citations). To investigate this, I use locally-weighted regression (+ citation), 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. By mapping out the coefficient estimates for each covariate per MSOA, we can examine how the mode-specific effects vary across space.

*Developing a Typology of MSOAs with K-Means Clustering*

Finally, I will synthesise the various aspects of job accessibility into a single result. Some studies attempt to do this by making the accessibility value of each area a function of multiple dimensions including physical accessibility, but this is an aggregation that obviates the detail in the individual dimensions. For instance, if accessibility is a function of both physical accessibility to jobs and the sectoral diversity of the jobs available, then it is unclear the extent to which an area’s accessibility level is due to physical accessibility or job diversity, even though these two dimensions reflect qualitatively different phenomena that illuminate different types of problems that require different responses.

Instead, in order to perform the multi-dimensional synthesis without entirely sacrificing the nuances in each dimension, I use a k-means clustering analysis incorporating 9 dimensions:

* Overall JPRs, to capture overall job accessibility
* The percent of each MSOA’s JPR contributed by high-end services, to capture sectoral dependence for job allocations
* The IMD rank for each MSOA, to capture the sociodemographic characteristics of the working-age population.
* The RBT for travel from each MSOA to the City, to capture travel time reliability to the most job-rich area in London
* I also identify which of the 6 other destinations are closest to each MSOA assuming all services run as planned, 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 represents to some extent the reliability of travel times from each MSOA to its immediate neighbourhood.
* MSOA-level coefficients for distance travelled by bus, Underground, National Rail and walking from the locally weighted regression, to capture the effects on travel time reliability of the most commonly used travel modes.

The typology of MSOAs obtained can capture how these dimensions intersect dynamically across London, and highlight the different job accessibility situations faced by each area. These results can then be used to help tailor a suite of interventions for each area to match its circumstances, as well as to prioritise interventions for areas that face more serious problems first.

**Results**

*Overall Job Provision Ratios*

Each area’s JPR is a function of 3 components: the spatial distributions of potential workers and jobs, and the connectivity provided by the public transport system. Figure XX visualises the first two components as a dot density map. In general, potential workers 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. In contrast, the distribution of jobs is much more clustered, 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 DLR networks), most notably at Canary Wharf.

The third component, area-to-area connectivity, is captured in the matrix C. Each value of co,d in C is the minimum time it takes to travel from the population-weighted centroid of MSOA o to the job-weighted centroid of MSOA d to arrive at 0830 on a typical weekday, assuming that all services run as planned. It is difficult to visualise the entire matrix in a meaningful way, but Figure XX visualises the row vector of travel times for all trips originating from the most populous MSOA, Stratford, ID E02000726:

with all areas within 45min travel time from Stratford (i.e. the areas where f(cE02000726,d) = 1) outlined in blue and a dot density map of the distribution of jobs superimposed. In the JPR calculations, each destination d within the blue boundary will contribute some of its jobs to each working-age resident in Stratford; exactly how much d contributes depends on how many other working-age residents live in areas where they can also reach d 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 potential workers who can reach each destination d within 45min, which is the result of . A dot density map of the distribution of potential workers 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 potential workers, and the number of potential workers able to reach each destination d declining as one moves away from the centre. Nonetheless, areas near major rail interchanges towards the outskirts may still be reachable within 45min by many potential workers.

Figure XX shows the number of jobs in each destination d that is allocated to each potential worker 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 destination d increases but decreases as the number of potential workers who can reach d 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 potential workers who can reach them, while a ring of MSOAs around the centre have few jobs but are easily reachable from many areas around London (which spreads the limited jobs available more thinly across larger catchments), and the MSOAs near the outskirts have few jobs but are also relatively isolated 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 o, which is the sum of the per-worker job allocations over all destinations d that can be reached within 45min. 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, potential workers and connectivity were all evenly distributed throughout the GLA area, which would result in all MSOAs having JPRs equal to R0, the ratio of the total number of jobs to the total number of potential workers. This value in the GLA area is 0.88. To express the extent to which each MSOA’s JPR deviates from R0, JPR values are binned into 5 classes: 0 to 50% of R0 (thus there are far too few jobs allocated to meet the needs of the potential workers), 50% to 80% of R0 (the number of jobs allocated is somewhat insufficient for the number of potential workers), 80% to 125% of R0 (the number of jobs allocated is proportional to the number of potential workers; this is a balanced state), 125% to 200% of R0 (there are not enough potential workers within reach to fill the allocated jobs), and more than 200% of R0 (there are far too few potential workers within reach to fill the allocated jobs). In general, we can expect unbalanced JPRs to make longer commutes more likely, either because workers cannot find enough jobs near them, or because there are not enough workers nearby to fill all the available jobs.

From Figure XX, we can see how the interaction between unevenly distributed jobs and potential workers on the one hand and uneven connectivity on the other results in most MSOAs in the GLA area having unbalanced JPRs. Too few potential workers are able to reach the MSOAs near the centre within 45min to fill all the available jobs, while many MSOAs around the outskirts provide too few jobs within reach of each potential worker. A notable exception is the relatively more balanced JPRs in the west, where Heathrow Airport provides a significant number of jobs, and the connectivity between Heathrow and residential areas in the west is comparatively good. The unevenness of job accessibility in London, with accessibility generally highest in the job-rich and well-connected centre, is similar to what other studies found in Boston (Shen, 1998), Montreal (El-Geneidy, et al., 2016), Santiago (Neihaus, et al., 2016), Bogota (Guzman, et al., 2017; Guzman & Oviedo, 2018), 11 major Canadian cities (Deboosere & El-Geneidy, 2018) and Los Angeles (+ citation), and is unsurprising overall.

Because the JPRs reflect an allocation of jobs among potential workers that treats jobs as excludable, it is straightforward to construct a Lorenz curve for the distribution of jobs among potential workers. This Lorenz curve is in the inset of Figure XX, and it neatly summarises the inequality in the job allocation: the bottom 50% of potential workers in the more job-poor areas share only roughly 20% of the jobs, while the top 10% of potential workers in the most job-rich areas have their pick of 20% of the jobs. The shape of this Lorenz curve resembles the curves for job accessibility for The Hague (Lucas, et al., 2015) and for Bogota (Guzman, et al., 2017). However, these two studies use cumulative potential accessibility measures that do not account for the competitive effects from the distribution of potential workers and treat job accessibility as non-excludable, and it is unclear how to interpret a Lorenz curve for a distribution of a non-excludable good.

*Job Accessibility by Sector*

Next, I analyse how job accessibility varies by economic sector. The pie chart in Figure XX shows that 42% of all jobs in the GLA area are from the high-end services sector, while consumer services and the public sector make up another 24% and 22% 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. In contrast, high-end services jobs are mostly concentrated in the centre and at Canary Wharf.

Each of the 6 maps in Figure XX shows the sector-specific JPRs of each MSOA rs,o. The top left panel of Figure XX reproduces the map of the JPRs for all jobs from Figure XX for comparison. Each map uses the same binning scheme as the top left panel, except that the values of R0 are now sector-specific. From the maps, we can see industrial jobs are most evenly allocated across the potential workers, while public sector and consumer services jobs are also relatively evenly allocated albeit with significant over-provision at the centre of the GLA area. In contrast, high-end services jobs are very unevenly allocated, with potential workers living in large swaths of MSOAs in the south, east and north having almost no jobs allocated from this sector, while potential workers living in MSOAs in the centre are allocated between 200% and 360% as many high-end services jobs as the GLA average. The allocation of jobs from the primary & extractive and other sectors is also very uneven, but there are so few jobs from these sectors in the GLA area that they do not make a significant difference to the overall situation.

When we plot Lorenz curves of the job allocation to all potential workers by sector (lower right panel of Figure XX), we can assess the extent to which the sectoral job allocations contribute to the unevenness in the overall allocation of 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 jobs in the latter three sectors tends to make overall job allocations 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 accessibility to jobs in the GLA area.

Finally, we can decompose each MSOA’s JPR sectorally and analyse the extent to which each MSOA is dependent on each sector for its job allocation. Each map in 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 more than 11% 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 unusually 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.

Similar sectoral breakdowns of job accessibility are rare in the literature. However, Shen (1998, p. 355) and Cheng & Bertolini (+ citation) found for Boston and Amsterdam respectively that sectoral job distributions tend to resemble the overall job distribution. This contrasts with my findings, but the discrepancy may be accounted for by the structural differences in the economies of the three cities, where London’s much larger and more developed high-end services sector gives it a greater ability to price out firms in other sectors from the most accessible central areas, resulting in higher levels of sectoral specialisation in the city centre.

*Sociodemographic Differences among Potential Workers from Different MSOAs*

Potential workers living in different areas of London have different sociodemographic characteristics. While the population counts in Figure XX are not disaggregated by characteristics like income, education level or disability status, MSOA-level variations in these characteristics are reflected in the IMD data. Figure XX visualises the ranking of each MSOA by IMD scores, together with 3 other 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 their values for each indicator for the purposes of visualisation, with the 10% of MSOAs that are most deprived visualised in dark red and the 10% least deprived MSOAs in dark green. In general, the spatial distribution of IMD rankings corresponds quite well to the distributions of the 3 raw indicators of social disadvantage, with a swath of more deprived MSOAs running north-south along the Lea Valley and another swath running from Dagenham to Clerkenwell along the northern bank of the River Thames, and other significant clusters around Shepherd’s Bush, Elephant & Castle and Streatham. Also, the absolute values of the Spearman rank correlation coefficients between the IMD rankings and each of the other 3 indicators are between 0.8 and 0.9, further confirming that the IMD ranks effectively capture multiple dimensions of social deprivation in London.

*Analysis of Variability in Travel Times*

For each trip originating in all 983 MSOAs and ending in one of 7 destinations, I obtain a distribution of 50 minimum travel times over 50 days. I then calculate the RBT for each origin-destination pair, which indicates the degree to which travel times for that pair are reliable. Each panel in Figure XX plots the RBTs of all trips 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 5min.

As expected, my RBT estimates are lower than those previously 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 further from Waterloo than Canary Wharf) are 4min and 1min respectively. Also, Ehrlich (+ citation) found that RBTs for direct bus trips between stops (i.e. that involve only in-vehicle time with no interchanges) varied between 5min and 42min, whereas the RBTs I calculate for all trips (including trips that involve interchanges) range from 0min to 41min. This is understandable, as the data I use to measure travel times does not account for all sources of travel time variability.

Nonetheless, we can see that RBTs vary widely across space, from very low values for very short trips to more than 40min for trips from XXX to Streatham. This confirms 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. Some large-scale spatial patterns in RBTs are also discernible. For instance, areas that are directly connected to the destination by rail tend to have lower RBTs, as trips from these areas do not involve interchanges, so travel times are only affected by service schedules. 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 transfers, 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 higher RBTs, as these trips tend to involve more transfers too, and long-distance travel tends to rely more on rail-based modes and are thus more likely to be affected by service suspensions. These results are in line with findings by Uniman (+ citation), Ehrlich (+ citation) and Duran-Hormazabal & Tirachini (+ citation), who also found that trips by bus tend to have less reliable travel times than trips by rail, and that travel times for longer trips tend to be less reliable.

In addition, 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 schedule desynchronisation during interchanges). Finally, for all destinations, we can see that trips originating from the southeastern quadrant of the GLA area tend to higher RBTs. This may reflect 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 for a trip seems to depend on the distance covered, the number of transfers, the reliability of different travel modes and location-specific factors. 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. The results for the initial model considering only transfers and distance travelled by each mode 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 transfers, and frequencies on bus services tend to be lower than on TfL-operated rail-based services, which increases the risk of schedule desynchronisation during transfers to bus services. The result for National Rail is consistent with the fact that National Rail services are less frequent and more prone to service suspensions than TfL rail-based services. On the other hand, travel by Underground, tram or DLR tends to reduce RBTs.

Column 2 in Table XX presents the results for the spatially autoregressive model which includes the local mean RBT as an additional covariate. As expected, the local mean RBT 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 for a trip increases by one minute, the RBT of the trip 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 reduces travel time reliability the most, followed by National Rail, and that travel by Underground, tram or DLR tend to increase travel time reliability.

Next, we examine the results for the locally weighted regression, starting with some diagnostics. 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 reliability in the outskirts is less correlated with the characteristics of the baseline trip than with other factors that have not been included in the model (such as the frequencies of services at nearby stations or the distance covered by services under each operator). 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 (for example, there is no estimate for the effect of tram travel on RBTs for MSOAs north of the Thames because all trips from each of these MSOAs to the 7 destinations do not use trams at all).

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 (red) shades indicate that the covariate tends to increase (decrease) travel time reliability for trips originating in that MSOA. 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 decrease travel time reliability almost everywhere in London, but particularly in the northeast around Stratford, Barking and Leytonstone, and around the City, Westminster, Rotherhithe, Lewisham, Sidcup, and in a swath from Waterloo to Wimbledon. Travel by National Rail degrades travel time reliability in a large swath in the southeast, indicating further evidence of operational unreliability in Southern and Southeastern services. In fact, National Rail services in these areas seem to be so bad that increasing the number of transfers and increasing travel by bus and walking (and thus reducing reliance on National Rail) tend to improve travel time reliability, contrary to patterns elsewhere. However, travel by National Rail also reduces travel time reliability in a northwestern corridor from Paddington to Ruislip, and around Stratford. This could reflect how most National Rail lines north of the Thames terminate outside the City (in stations like Paddington, Marylebone, Euston, Kings Cross, St. Pancras and Liverpool Street) and do not provide direct orbital connections, so National Rail users need to transfer 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 reduce travel time reliability, particularly in the west; this may be because TfL Rail operates on tracks that are also used by National Rail services, and thus its operating patterns and reliability may resemble National Rail services more.

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

Finally, I carry out a k-means clustering analysis to explore how job accessibility, sectoral dependence for job allocations, social deprivation among potential workers and travel time reliability intersect among different MSOAs. 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 tend to contain MSOAs with lower deprivation, Clusters 5, 6 and 7 tend to contain MSOAs with higher deprivation, and Clusters 3 and 4 are more diverse.

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, indicating extreme surpluses of jobs allocated per potential worker. They are also among the most dependent 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 very 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. In contrast, the MSOAs in Cluster 4 are distinguished by their combination of job poverty 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 potential workers live 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 potential worker. These potential workers 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 and resources 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. MSOAs in Cluster 1 (with 23% of the working-age population and covering areas near the outskirts) also enjoy relatively reliable travel times to the City and local destinations, but experience job shortages due to their greater distance from the centre and the scarcity of jobs in the outskirts. In contrast, MSOAs in Cluster 2 (with 10% of the working-age population located near the outskirts) experience both insufficient jobs and high travel time unreliability, particularly due to National Rail services.

In London, the situations facing socially deprived residents are also quite diverse. MSOAs in Cluster 5 (with 18% of the working-age population) are located quite centrally, almost surrounding the MSOAs in Cluster 3. As a result, despite being home to more socially deprived residents, these MSOAs enjoy similar conditions as those in Cluster 3, benefitting from excellent transport connectivity, high travel time reliability and a surplus of jobs allocated per potential worker. However, these jobs also tend to be from the high-end services sector, which more socially disadvantaged workers may find more difficult to take advantage of due to a lack of educational qualifications, relevant working experience or social networks. 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 a mismatch between high-end services jobs and more socially disadvantaged workers is unclear from this analysis.

MSOAs in Cluster 6 (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 lack of jobs. Travel seems to be reliable but too slow to connect these areas effectively to large job clusters. In contrast, MSOAs in Cluster 7 (containing 7% 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 problematic. These areas are also more dependent on the high-end services sector for jobs. Thus, these areas may also be more vulnerable to a mismatch of characteristics between potential workers and the available jobs.

In summary, the k-means clustering analysis indicates that there are locations in the GLA area where more socially disadvantaged working-age residents experience a combination of insufficient accessibility to jobs, dependence on the high-end sector for jobs and/or travel time unreliability. 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 a significant proportion of low-income workers 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.

The k-means clustering analysis also allows us to develop a typology of MSOAs in the GLA area that captures a more nuanced understanding of job accessibility, allowing us to expand on the usual focus 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 potential worker within a reasonable travel time, each area may also face challenges in terms of skills mismatch and travel time reliability, especially by bus or National Rail. The particular combination of challenges in each area suggests that intervention strategies should be tailored to each area’s characteristics.

**Discussion**

*Limitations and Further Research*

Overall, the new findings from this study on job accessibility in London seem plausible in the context of the literature. However, we need to keep in mind some caveats when interpreting these results. I list three particularly important ones below, as well as preliminary ideas on how to design future research to better address them.

Firstly, the RBT measurements in this study probably underestimate the true level of travel time unreliability, because it is based on Journey Planner travel time estimates that do not account for real-time vehicle locations and congestion effects on travel time. To a certain extent, this can be addressed using data from other TfL API endpoints, especially for live arrival times at stops. For example, once the route for a trip is obtained from the Journey Planner, the stations where a traveller needs to board and alight can be identified, and separate scripts can then repeatedly request the live arrivals for each station to track the vehicles that the traveller would ride on. Actual waiting and in-vehicle times for each leg of the trip can then be obtained and updated into the database of trip characteristics. Nonetheless, without live vehicle loading data, we would still be unable to account for the effects of in-station and in-vehicle crowding that may reduce in-station walking speeds and prevent boarding of vehicles. In effect, this would be equivalent to tracking the progress of 6,881 ghost travellers (who can occupy the same space as other real travellers) through the system in real time. Another alternative is to use data from other trip planning applications that do take into account real-time vehicle locations, such as CityMapper. CityMapper data has the added advantage of also accounting for bus diversions when suggesting routing options for a trip. However, special permission is required to access CityMapper data in bulk through their API.

Secondly, there may be concerns that the travel time for a trip that ends at 0830 on 22 March 2019 may not reflect the ideal travel time for that trip during a weekday AM peak, assuming all services run as scheduled. This is because the options that exist for trips ending at 0830 may be unusual. For instance, a trip that ends at 0830 may not be able to use faster express services scheduled to arrive at 0815 and 0845, or an express service that is available for arrival by 0830 may only run rarely during the whole AM peak. If we plot for each of the 6,881 trips the travel time for a trip that ends at 0830 on 22 March 2019 against the median observed travel time (Figure XX), we do see that there are instances where the real median travel time is markedly more or less than the baseline travel time. Nonetheless, the two quantities are highly correlated, with a Pearson correlation coefficient of 0.987. As such, I believe that the baseline travel times are sufficiently good to represent typical travel times for all the origin-destination pairs. However, if it is important to ensure that the baseline times are truly representative of all the possible travel times for a trip, then the best way would be to collect real-time observations of travel times for that trip to construct the travel time probability distribution.

A third limitation lies in the definition of f(co,d). Workers living in different areas probably have different levels of willingness to travel; a worker living in the City may view a 45min commute as excessive, while one living in the outskirts may be quite willing to travel much more than 45min to work. In principle, we could use a different function to determine the value of each cell in C; if we view C as an incidence matrix of a two-mode network consisting of m origin nodes and n destination nodes, it is plain that the value in any cell of the matrix (i.e. the weight on the edge between any origin o and destination d) can be set independently from all other values. As such, we could craft C to reflect any network topology between origins and destinations that we would like. For instance, one could set the travel time cut-off for all trips from each origin o to the median travel time for workers living in o. Further, the functions for some cells could use contour measures while others could use distance decay functions. The creativity of the analyst on this front is basically limited only by her ability to justify empirically her design choices for the incidence matrix.

*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 policy-making in these fields.

Firstly, the map of overall JPRs in Figure XX reveals locations where there are too few or too many jobs allocated per potential worker. This can directly guide the distribution of future growth to shape the overall urban structure, with new housing supply being directed to locations with job surpluses, and job growth being encouraged in areas with too few jobs per potential worker. Transport connectivity between areas with high JPRs and low JPRs can also be enhanced, to put more potential workers 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 probably underestimates actual travel time variability, the results are still useful for identifying 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. Moreover, because the travel time variability measurements in this study only account for service suspensions and schedule desynchronisation and are not confounded by factors relating to travellers’ behaviours such as crowding, they can directly justify efforts to adjust service frequencies and arrival times to reduce the probability of schedule desynchronisation during transfers, or to adjust the incentives for operators so that they put more effort into preventing service suspensions.

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 non-spatial mismatch between potential workers and the available jobs. For instance, referring to Figure XX, the potential workers 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 or development strategies 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, we should look for low-cost and easily implementable interventions that make the most out of existing resources and programmes, such as extending the bus lane network or working with local institutes of higher learning to expand skills training.

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 place many more workers in the east and west of London 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 Ilford and Chadwell Heath 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; this should justify building the northern leg first. 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, the number of working-age residents that can reach Lewisham within 45min travel (shown in Figure XX) 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. This strategy of dovetailing transport improvements and development is probably applicable to the other 3 locations as well, though RBT measurements for trips to those destinations are not available from this study.

Overall, the results paint a picture of job accessibility in London that is hopeful. Problem areas do exist where high social deprivation, low job accessibility, high sectoral dependence and low travel time reliability coincide to varying degrees, but it is important to recognise that there are also areas where more disadvantaged residents enjoy job accessibility conditions that are similar to their more privileged neighbours. In addition, TfL, the GLA and UK governments have recognised the importance of job accessibility to enhancing residents’ quality of life, and have demonstrated a commitment to improving the situation, in particular to benefit more socially deprived areas, through measures that have already been implemented or plans that they are already pursuing. This study affirms the efforts made so far, and provides some guidance for how to tailor ongoing efforts to more effectively address circumstances that are particular to different locations.

# **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.

1. The Journey Planner also returns cycling options, which tend to be significantly faster than bus or walking. However, because only 2% of trips trips in London were by cycling (Transport for London, 2018), I omit these options from this study to avoid underestimating travel times. [↑](#footnote-ref-2)