# Deriving transport appraisal values from emerging revealed preference data

Panagiotis Tsoleridis[a,*], Charisma F. Choudhury[a], Stephane Hess[a]

[a] *University of Leeds, University road, Leeds, UK, LS2 9JT*

**Abstract**

Transport demand models are widely used to inform policy making and produce forecasts of future demand. A core output derived from demand models is the Value of Travel Time (VTT), which provides insights on the trade-offs that travellers are willing to make in terms of travel time and travel cost. VTT estimates are a critical input to cost-benefit analyses and feasibility assessments of potential projects and thus play a crucial role in transport planning and policy decisions. While much of the early work on VTT made use of revealed preference (RP) data, their use decreased due to growing concerns about reporting errors that may result in omitted observations and measurement errors in the model inputs. As a consequence, VTT measures have, for the last two decades, primarily been estimated using state-preference (SP) surveys. While SP methods can assess the individual trade-offs in a controlled manner, they are prone to behavioural incongruence. More recently, RP data from passively-collected data sources have raised the promise of accounting for some of the limitations of traditional RP surveys due to the minimal (or even no) active input from the respondent. The present study utilises such a dataset that combined a 2-week trip diary captured through smartphone GPS tracking with a household survey containing individual socio-demographic information. Mixed Logit models for mode choice were specified and more broadly representative VTT estimates were then applied on the National Travel Survey and using weights based on trip distances. This process resulted in estimates similar to the official UK guidelines used in transport appraisal that were obtained from SP data, where our results are not affected by concerns about response quality or survey artefacts. The findings hence strengthen the case for shifting towards passively generated RP data sources and are important for transport practitioners.

*Tags: transport appraisal, values of travel time estimates, emerging data sources, gps, mixed logit models*

*Corresponding Author

*Email addresses:* ts17pt@leeds.ac.uk (Panagiotis Tsoleridis), C.F.Choudhury@leeds.ac.uk (Charisma F. Choudhury), S.Hess@leeds.ac.uk (Stephane Hess)

## 1. Introduction

Transport projects and schemes can substantially impact our day-to-day lives, as well as mid-term decisions like whether or not a buy a car or long-term decisions like where to live. They also have a profound impact on the economic growth of the country, its productivity and people's well-being. Cost-benefit analyses (CBA) and feasibility assessment of potential transport projects are primarily based on the monetary savings from travel time reductions. It is estimated that savings on travel time are responsible for around 80% of the predicted benefits of a new transport project in the UK (Mackie et al., 2001; Fosgerau and Jensen, 2003; Daly et al., 2014). The Values of Travel Time (VTT) estimates, which are used to quantify the trade-offs that decision-makers are willing to make in terms of travel time and travel cost, are hence critical components of CBA. An accurate estimation of VTTs is thus important in order to properly evaluate the costs and the benefits of a new transport project and sufficiently forecast future demand for specific services, e.g. a new public transport route, leading to better informed decisions during the planning phase.

Estimates of the trade-offs that travellers would be willing to make in terms of travel time and cost were first produced in the 1960s. For a long period of time, VTT estimates were derived as relative values to the average wage cost (wage cost method or cost savings approach - CSA) and that method is still in use in several countries (Daly et al., 2014). Another approach involved Contingent Valuation, where VTTs were derived from direct questions about how much a participant would be willing to spend for a particular service or an improvement of a current one. In recent decades, most types of VTT analysis are based on the work of Daly and Zachary (1975), who introduced a Random Utility framework around the VTT estimation grounded on microeconomic theory (Daly et al., 2014).

Revealed Preference (RP) data, usually coming from travel diaries, would at face value provide the natural data source for estimating VTTs, and indeed were used in early studies (Beesley, 1965; Daly and Zachary, 1975). Nonetheless, while RP data provide the ability to capture real-world choices, most of the parameters influencing them are outside of the analyst's control. Furthermore, traditional RP data sources include recalled/reported data that are prone to issues like omitted trips (particularly short ones), perception and rounding errors, etc. - often leading to large measurement errors. This in turn led to the growing popularity, over the last two to three decades, of Stated Preference (SP) data as the main input to models, with RP data often being used only in limited scale for verification of the SP results. SP surveys present respondents with a number of hypothetical scenarios, where they are asked to choose among a set of alternatives. This approach has a long tradition for example in the United Kingdom (UK) with the first major SP survey conducted in 1984 (MVA et al., 1987) and follow-up studies in 1994 (Accent and Hague Consulting Group 1996) with the same data then re-analysed by Mackie et al. (2003) before the most recent study involving primary data collection taking place in 2014-2015 (Batley et al., 2019; Department for Transport, 2015; Hess et al., 2017).

SP surveys are generally seen to have the advantage of providing the analysts with an environment where they have control over a large number of parameters that could influence VTT estimates, such as the attributes of the alternatives. On the other hand, SP surveys are prone to behavioural incongruence and hypothetical bias and are often criticised for being too sensitive to the experimental design and the representation of the SP scenarios (Brownstone and Small, 2005; Daly et al., 2014; Haghani et al., 2021). Concern in a VTT context has also been raised in relation to the use of overly simplistic settings in some countries (Hess et al., 2020).

In a recent study examining the impact of hypothetical bias in SP surveys within several domains including transport (Haghani et al., 2021), the authors concluded that although it is more sensible to assume that individuals would likely overstate their Willinness-to-pay (WTP) in a hypothetical scenario (Li et al., 2020), there are a number of transport studies showing the opposite (Nielsen, 2004; Brownstone and Small, 2005; Krcal et al., 2019). As an example, we refer to the study of Brownstone and Small (2005) examining WTP for toll pricing among a range of road corridors in the United States, where the estimates obtained from hypothetical SP-surveys systematically underestimated the VTTs compared to RP-based estimates. A potential cause could be that VTT estimates will depend to a large extent on the travel time and cost range values of the SP-scenarios, which could differ substantially in a real-life scenario, such as the case of severe congestion during the morning peak, forcing individuals to place higher valuations of time so as to avoid arriving late at work.

Evidence from neuro-imaging studies also suggests that individuals would often react differently in a stressful situation compared to the lab setting of an SP survey, e.g. be willing to pay more in order to avoid an unpleasant outcome (Loewenstein, 2005; Kang and Camerer, 2013; Haghani et al., 2021). In addition to that, Kang and Camerer (2013) showed that a certain part of the brain was more strongly activated when participants had to make a real choice compared to a hypothetical one. Other psychological effects can also come into play during an SP survey influencing participants' responses, such as the desirability to appear more socially acceptable to the analyst, the feeling that their choices will lack of any real-world consequences, or the opposite with respondents deliberately giving misleading answers to avoid a potentially harmful outcome resulting from that study, e.g. a road pricing scheme (Haghani et al., 2021).

With that evidence in mind and considering that VTT estimates are to be used for the purpose of project evaluation during a CBA, it is only sensible to assume that policy makers would be mostly interested in the trade-offs individuals are willing to make under real-life conditions, sometimes stressful, while taking into account real distributions of travel time and cost and not the ones imposed by the analyst (Louviere and Hensher, 2001; Brownstone and Small, 2005). This thus motivates an increased interest in revealed preference (RP) data for VTT studies. For example, in a review of transport appraisal studies performed in various countries, Daly et al. (2014) concluded that despite SP data have been the standard approach so far, researchers and practitioners should reconsider the use of RP data due to the benefits they can provide, while also taking advantage the new emerging and more robust data collection methods.

Emerging data sources, primarily from sensors, such as GPS and mobile phone data have provided new breakthroughs and challenges to researchers. Travel diaries captured through GPS tracking are able to produce large panels of RP data per participant at a very high spatial and temporal resolution. Compared to traditional pen-and-paper diaries, GPS-based surveys offer the advantage of capturing an increased number of daily trips giving a more representative depiction of individual mobility behaviour without resulting in user fatigue. Though there have been limited efforts to infer VTTs from anonymous RP data sources (e.g Bwambale et al. (2019)), the absence of socio-demographic information of the travellers and trip characteristics (e.g. trip purpose) have meant that it is not possible to capture the heterogeneity in the VTTs among different socio-demographic groups of users or due to the differences in trip purpose (e.g. commute, business, leisure) from such data.

A GPS travel diary linked to a background household survey can help to account for the limitations of GPS-based RP trip diaries. Several studies have used GPS datasets complimented with a background survey, but most of them have limited their analysis on descriptive statistics of individual mobility behaviour based on the observed choices (Arifin and Axhausen, 2012) or estimated models of mode choice, but without reporting VTT estimates (Schuessler and Axhausen, 2009; Montini et al., 2017; Huang et al., 2021). An exception is the study of Calastri et al. (2018), who estimated mode choice models based on GPS data for the purpose of uncovering latent mode availability and consideration constraints of the individuals during their decision-making process. Their study also reported VTTs based on the estimated parameters, however, this was purely as a means of validating their proposed approach, with no emphasis on extrapolating the findings to a representative sample, as required for official VTT values.

The focus of GPS studies so far in the literature has thus not been on the estimation of behaviourally accurate VTTs, representative of the country's population, which are derived from GPS tracking, and more importantly they have not been compared with national official estimated SP-based VTTs before. That limitation in the current literature and the lack of empirical evidence could partly explain the reluctance of policy makers to accept the use of new emerging GPS data for VTT estimation for appraisal purposes, a task that is still heavily reliant on SP surveys. Aiming to address that limitation, the current study utilises such an emerging data source, namely a 2-week GPS trip diary including 540 participants and 12524 trips, collected as part of the European Research Council funded "DECISIONS" project, for the purpose of estimating a behavioural model of mode choice. The estimated parameters are then applied to the National Travel Survey (NTS) data and VTT estimates are derived, which are weighted by distance to ensure proper representativeness of the UK's population. The main aim of the study is to compare the final distance-weighted VTT estimates with the latest official SP-based VTTs currently used in appraisal in the UK.

The remainder of the paper is as follows. In the following section, the datasets used in the current study are described, while in the third section, the modelling framework is outlined. In the next two sections, the modelling outputs and the derived VTT estimates are analysed. Finally, in the last section, the policy implications alongside the conclusions and limitations of the current study are summarised and the scope for

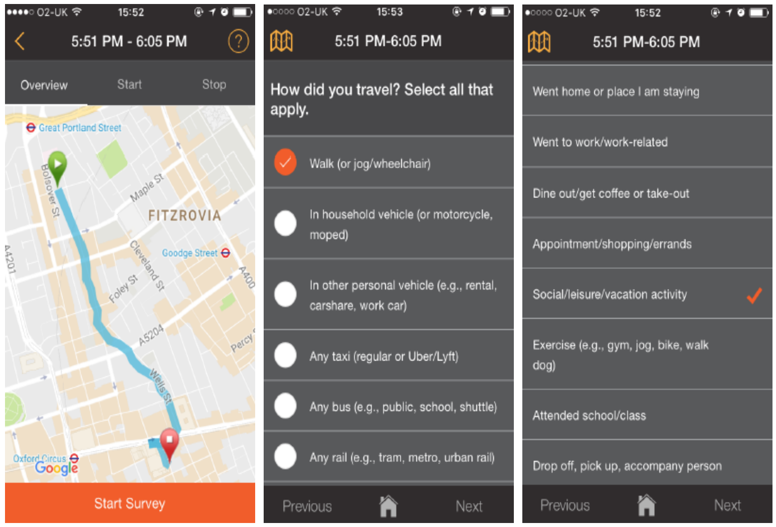


Figure 1: User interface of smartphone application used for the GPS trip diary (Calastri et al., 2020)

future studies is outlined.

## 2. Data

### *2.1. DECISIONS data*

Several datasets are utilised in the current study. A behavioural model of mode choice is estimated using the labelled GPS dataset, which was collected between November 2016-March 2017, as part of the "DECISIONS" research project aiming to understand individual transport and energy choices. A detailed description of the dataset (referred to as DECISIONS dataset in the remainder of this paper) is presented in Calastri et al. (2020). That survey consists of several submodules including a trip diary captured through GPS tracking using a smartphone application and a household survey capturing important sociodemographic information of the participants. The GPS trip diary includes the participants' trips during a 2-week period, in which additional information on the purpose and the chosen mode had to be provided at the end of each trip (*Figure 1*).

The GPS diaries initially included 721 unique individuals and 56,693 observed trips around the UK with the vast majority of those being around the region of Yorkshire and the Humber, and predominantly around the city of Leeds. As a result, only trips within the region of Yorkshire were selected for the subsequent analysis to avoid larger estimation errors for less represented areas, such as London. The spatial distribution of trips initially included in the dataset, represented as interzonal flows between MSOA zones across the UK, is depicted in *Figure 2*. The observed modes of transport included car, bus, rail, taxi, cycling and walking.

A significant effort was undertaken during the cleaning phase with an emphasis on detecting inconsistencies between consecutive trips, in terms of time (following trip starting before the end of the previous trip) and space (space gaps between two consecutive trips). Furthermore, a large number of trips were left *untagged*, meaning that participants did not provide mode and purpose information, and these thus had to be removed from the analysis. No pattern was identified for the erroneous observations or untagged trips that were removed and those were considered as missing at random for the subsequent analysis. Unique activity locations were defined by clustering the observed latitude/longitude coordinate pairs. Hierarchical Agglomerative Clustering (HAC) was used for that purpose, since it does not require the analyst to predetermine the number of clusters. A distance threshold of 200m was selected for the observed destinations to be considered in the same cluster, which resulted in the most plausible clusters for the sample after testing thresholds between 50-500m. As a result from that process, home and work locations were identified based on the tagged purposes for those locations, the time of the day that those locations were visited and the time spent there. Trips were then assigned into tours, starting from and finishing at the home location, per individual and for each day of the survey.

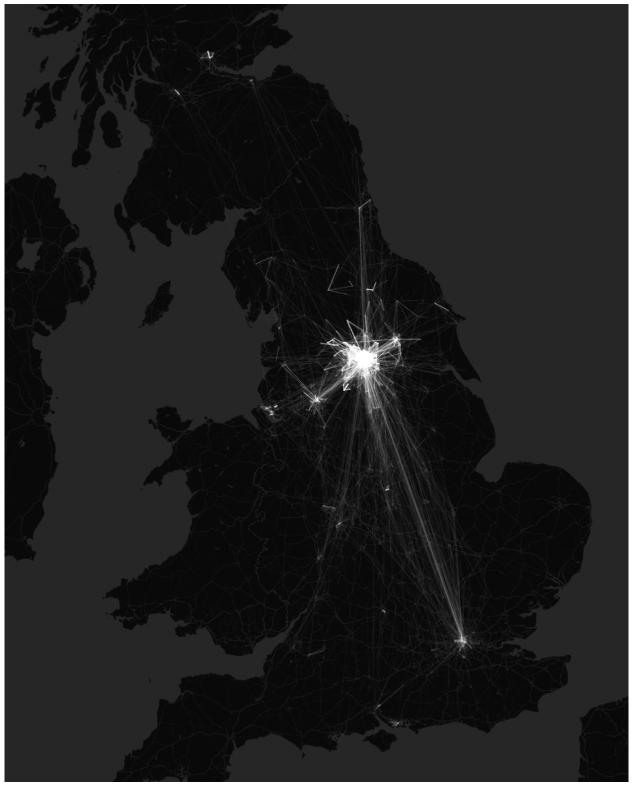

Figure 2: Spatial distribution of interzonal flows between MSOAs across the UK

The aforementioned process allowed us to adopt a tour-based approach in terms of mode availability per individual and choice task. In that sense, if an individual chooses car for the first trip of the tour, then only car will be available for the remaining trips of the tour, since it has to be returned back home. Therefore, in such a choice task only the first trip is relevant in a mode choice context and therefore included in the analysis. Conversely, if any other of the available modes was chosen for the first trip, i.e. bus, rail, taxi, cycling or walking, then car would become unavailable for the remaining trips of the tour.

A significant problem inherent to RP data and especially to those derived from new emerging data sources, such as GPS, is the lack of any information on the non-chosen alternatives. To overcome that obstacle, the "Directions" Google API was implemented providing travel times and distances for a range of modes[1]. Travel times/distances for both chosen and unchosen mode alternatives were re-calculated using the API for consistency reasons and to ensure that all values would come from the same data generation process (Calastri et al., 2018). The requests passed on to the API were for two weeks in the future from the date of analysis and for the same day of the week and time of day as in the initial dataset. Two weeks in the future were chosen, so as to avoid the short-term negative impacts of a recent traffic disruption and because the API cannot be used for past dates.

The API provided travel times for car/taxi based on traffic information for the specific time of day when the observed trip was performed. Travel times on the shortest distance routes are calculated for walking and cycling trips. For bus and rail trips, a timetable approach is used and a detailed breakdown of the whole route is provided including walking segments, waiting times and transfers between different services. That level of detail was essential in order to quantify in-vehicle and out-of-vehicle travel times and to be in line with the official VTT estimates. Bus and rail trips for which the API returned only walking segments due to short distance or lack of service, were assigned as unavailable for those choice tasks.

Travel cost was also missing for all alternatives. Car costs were calculated using the official WebTAG

[1]More details can be found here: https://developers.google.com/maps/documentation/directions/overview

specifications regarding fuel and operating costs (Department for Transport, 2014). Parking costs were also calculated based on the location of the observed destination (central areas, local high streets etc.) and the activity duration there. Information on hourly and fixed parking costs was obtained from the local authorities in the region of Yorkshire. Fuel, operating and parking costs were added together to form the total car travel cost used for estimation. Bus and rail costs were calculated based on a distance-based fare of the most popular bus and rail operators in the region and a discount was applied for season ticket holders. Finally, taxi cost was calculated using fixed, hourly and distance-based average costs for different cities around Yorkshire.

The final dataset used for model estimation contained 12,524 trips and 540 unique individuals, which is significantly smaller than the SP sample used for analysis in the official study consisting of 7,692 individuals and 15 choice tasks per individual (Hess et al., 2017). Regarding the observed/chosen modes, 47.6% were car trips, 14.6% bus, 5.2% rail, 3.2% taxi, 3.3% cycling and 26.1% walking trips. Commuting and business trips were observed for 19.1% and 9.7% of the sample, respectively. The majority of respondents were female (59.6%) of an average of 40 years old, 75.4% had at least one car in their household and finally 20.7% and 13.3% had a bus and rail season ticket, respectively. The average trip distance across modes per choice task is 2.4 miles (6.9km) with a maximum of 61.1 miles (98.4km) depicting the urban nature of the trips captured in this survey and the absence of longer-distance interurban trips.

### *2.2. NTS dataset*

The estimates derived from the model estimated on the DECISIONS were applied on the NTS dataset. The NTS is the official annual survey in the UK providing invaluable long-term information on travel behaviour and mobility trends[2]. Three consecutive years of the NTS data were used, namely 2015-2016-2017, to ensure a representative sample, while also providing an overlap with the period of the DECISIONS survey. A non-sensitive version of the NTS dataset was acquired[3], where information on personal income was missing. Furthermore, household income was not reported for more than 65% of participants for each year rendering it practically unusable. In addition, since the parameters were estimated on a dataset containing trips mostly around the region of Yorkshire (DECISIONS), it was decided to exclude trips in London from the NTS data, due to the individuals there generally having a different set of available modes, e.g. underground.

The NTS dataset did not provide any information on travel cost for car trips and for a large number of bus and rail trips. For the former, only information on parking cost was provided and fuel and operating costs were imputed using the same approach as in the DECISIONS dataset (Department for Transport, 2014). Fare cost for bus and rail was missing or reported as zero for 49% and 13.4% of bus and rail trips, respectively, which were performed by season ticket holders. Since it is not reasonable to assume a zero VTT for season ticket holders, an average daily cost of a season ticket was applied for bus and rail, based on the cost calculations performed for the DECISIONS data.

The final NTS dataset used for the analysis included 453,438 trips performed by 29,127 unique individuals, with 52.6% of those being female and with an average age of 50 years old. The average trip distance is 7.9 miles (12.7km) with a maximum of 719 miles (1,157.1 km) showing a more representative depiction of mobility behaviour including both urban and interurban trips. Regarding the observed modes, 80.4% were car trips, 5.1% bus, 1.2% rail, 1.2% taxi, 1.9% cycling and 10.1% walking trips. Finally, commuting and business trips account for 17.4% and 4.0% of NTS trips, respectively. Detailed descriptive statistics of DECISIONS and NTS trips per mode and purpose are presented in *Table 1*.

## 3. Modelling framework

The VTT estimates presented in the current study are derived from a behavioural model based on the Discrete Choice Modelling (DCM) framework (Ben-Akiva and Lerman, 1985; Train, 2009). A DCM framework based on Random Utility Maximisation assumes that each individual $n$ has a preference for a specific alternative $i$ among a set of $J$ alternatives in a choice task $t$ represented as a latent utility $U_{int}$ consisting of a deterministic part $V_{int}$ and a disturbance term $\epsilon_{int}$. Different distributional assumptions about

[2] Details can be found here: https://www.gov.uk/government/collections/national-travel-survey-statistics
[3] The NTS dataset was acquired from https://beta.ukdataservice.ac.uk

Table 1: Number of DECISIONS and NTS trips per mode and purpose

| **Mode** | **Commuting** | **Business** | **Other (non-work)** | **Total** |
|---|---|---|---|---|
| DECISIONS trips | | | | |
| *Car* | 1,015 (8.1%) | 693 (5.5%) | 4,253 (34.0%) | 5,961 (47.6%) |
| *Bus* | 510 (4.1%) | 201 (1.6%) | 1,117 (8.9%) | 1,828 (14.6%) |
| *Rail* | 243 (1.9%) | 55 (0.4%) | 350 (2.8%) | 648 (5.2%) |
| *Taxi* | 23 (0.2%) | 30 (0.2%) | 352 (2.8%) | 405 (3.2%) |
| *Cycling* | 121 (1.0%) | 19 (0.2%) | 269 (2.1%) | 409 (3.3%) |
| *Walking* | 477 (3.8%) | 214 (1.7%) | 2,582 (20.6%) | 3,273 (26.1%) |
| *Total* | 2,389 (19.1%) | 1,212 (9.7%) | 8,923 (71.2%) | **12,524 (100%)** |
| NTS trips | | | | |
| *Car* | 62,750 (13.8%) | 16,199 (3.6%) | 285,594 (63.0%) | 364,543 (80.4%) |
| *Bus* | 5,054 (1.1%) | 406 (0.09%) | 17,580 (3.9%) | 23,040 (5.1%) |
| *Rail* | 2,133 (0.5%) | 488 (0.1%) | 2,916 (0.6%) | 5,537 (1.2%) |
| *Taxi* | 725 (0.2%) | 143 (0.03%) | 4,786 (1.0%) | 5,654 (1.2%) |
| *Cycling* | 3,475 (0.8%) | 254 (0.1%) | 4,927 (1.1%) | 8,656 (1.9%) |
| *Walking* | 4,729 (1.0%) | 487 (0.1%) | 40,792 (9.0%) | 46,008 (10.1%) |
| *Total* | 78,866 (17.4%) | 17,977 (4.0%) | 356,595 (78.6%) | **453,438 (100%)** |

the disturbance term would yield a different specification form. The most commonly used specification is the Multinomial Logit model (MNL) assuming a type-I (Gumbel) Extreme Value distributed disturbance term (McFadden, 1973). The deterministic part $V_{int}$ consists of alternative- and individual-specific attributes, $x_{int}$ and $z_n$, respectively, as shown in *Equation 1*. The choice probabilities of an MNL model are derived from *Equation 2*.

$$U_{int} = V_{int} + \epsilon_{int} = f(\beta, x_{int}, z_n) + \epsilon_{int} \tag{1}$$

$$P_{int}(\beta) = \frac{e^{V int}}{\sum_{j=1}^{J} e^{V_{jnt}}} \tag{2}$$

where $\beta$ is a vector of parameters to be estimated.

The basic MNL specification assumes that individuals will have the same sensitivity to the specified parameters. Deterministic taste variation in response to specific attributes can be captured as shifts from their base level for specific types of individuals or choice tasks. In the present study, deterministic heterogeneity was captured by specifying shifts from the base level of the alternative specific constants (ASCs) for specific sociodemographic attributes. Furthermore, shifts were also included for the base time and cost parameters of level-of-service (LOS) variables for business and commuting trips. An elasticity specification was used for interactions with continuous sociodemographic attributes, such as age and income, with a separate beta being specified for respondents who did not provide any income information.

Even in the case of accounting for deterministic heterogeneity, however, it is reasonable to assume that some degree of heterogeneity would still remain uncaptured among and/or within individuals leading to biased estimates. Mixed Logit models (McFadden and Train, 2000) can be used to account for that, offering a more flexible specification, where parameters are allowed to vary randomly across individuals. Mixed Logit models are considered as the most general form of a Logit, since they are able to approximate any other specification (McFadden and Train, 2000). The results, however, will largely depend on the distributional assumptions for each random parameter, a task bestowed on the analyst.

The choice probabilities in a mixed MNL model are now given by an integral over the distribution of individuals' sensitivities (which follow a density function $\phi(\beta|\Omega)$), where this integral does not offer a closed form solution. Simulated log-likelihood estimation is an alternative way of calculating the integral of choice probabilities, based on drawing random numbers from a pre-specified distribution. From that process, the choice probabilities can be calculated as the average over the draws (*Equation 3*) and the simulated log-likelihood can be computed as shown in *Equation 4*.

$$\widehat{P_{int}(\Omega)} = \frac{1}{R}\sum_{r=1}^{R} P_{int}(\beta^r) \tag{3}$$

$$SSL(\Phi) = \sum_{n=1}^{N} ln(\widehat{P_{int}(\Phi)}) \tag{4}$$

where $\beta^r$ is a random draw from a distribution with $\phi(\beta|\Omega)$.

It is reasonable to assume that the impact of LOS parameters should be strictly negative, indicating that an additional minute spent travelling or an additional unit of cost spent for a trip will decrease the utility and therefore the choice probability for a certain mode alternative. The specified distribution for the random LOS parameters should be able to account for that, with the negative log-normal distribution being the most applied one for that purpose. In the current study, the long tails of the log-normal distribution resulted in numerical issues during estimation, prompting us to take a different approach. As a consequence, the negative log-uniform distribution was chosen instead, with its shorter tails ensuring no problems during estimation, similarly to the official UK study, which provided the first large scale application of that distribution (Hess et al., 2017). Under that distribution, a variable $x$ is log-uniformly distributed, if $y = log(x)$ is uniformly distributed. The log-uniform distribution is defined by two additional parameters, $a$ and $b$ denoting its lower bound and spread, respectively. The mean and the variance of the log-uniform distribution are calculated as following (Hess et al., 2017):

$$E(\beta_0) = \frac{e^{a+b} - e^a}{b} \tag{5}$$

$$Var(\beta_0) = e^{2a}\left[\frac{e^{2b}-1}{2b} - \frac{(e^b-1)^2}{b^2}\right] \tag{6}$$

In total, nine parameters were specified as random, namely travel time for car, taxi, walking and cycling, in-vehicle (IVT) and out-of-vehicle (OVT) travel times for bus and rail and finally travel cost. Due to the multidimensionality of the integral, Modified Latin Hypercube Sampling (MLHS) draws were chosen over Halton draws to avoid the multicollinearity issues identified with multidimensional Halton sequences (Hess et al., 2006). For the simulated log-likelihood estimation, 1,000 MLHS numbers $r_U$s were drawn from a uniform distribution for each randomly distributed $\beta$, which was specified as $\beta_{LU(a,b)} = e^{a+b*r_U}$. At that number of draws, a sufficient level of stability was observed among the estimates and model fit, hence it was decided not to increase the number of draws any further.

## 4. Modelling results

The outputs of the behavioural models estimated on the DECISIONS dataset, base MNL and mixed MNL, are presented in *Table 2*. The mixed MNL model with 9 additional parameters provided significant improvements in model fit reducing the LL by 1153.6 units. The adjusted $\rho^2$ of 0.7691 also signifies that the model is able to explain a significant portion of the variation in the dataset. The Alternative Specific Constants (ASCs) reveal that, all else held equal, individuals have a negative inherent preference for bus, taxi and cycling over the base mode of car. The preferences for walking and rail are higher than car, although the ASC for the latter is not statistically significant at the 95% confidence level. Bus is a more preferred option for travellers with lower personal income or no employment. Taxi is more preferable for travellers who are younger (18-24) and is less preferable for individuals of lower education (with no undergraduate degree), which can be considered as a proxy of income. Cycling is more preferred by men and students, while it is also more preferred by both the lowest and the highest personal income bands. Finally, walking is more likely to be chosen by students and younger ages in general.

Regarding the LOS parameters, all estimates have behaviourally reasonable signs. Travel time for car, taxi, cycling, walking, OVT bus and OVT rail were specified using a box-cox transformation, as $\frac{x^{\lambda}-1}{\lambda}$, capturing significant non-linearities. A log specification of travel cost led to the best results depicting the presence of cost damping effects in the sample (Daly, 2010). For IVT for bus and rail, however, only linear sensitivities were supported by the sample. Commuters and business travellers have increased time sensitivities and the latter have decreased cost sensitivities, as well, although not statistically significant. Different combinations of personal and household income were specified as elasticities for the LOS parameters, with only the elasticity of personal income and OVT for bus and rail resulting in statistically significant estimates. The sign of the parameter is positive, meaning that as income increases, the sensitivity to OVT also increases, which is behaviourally sensible. Age elasticity was also specified and estimated separately for mechanised modes, cycling and walking resulting in negative signs for all of them meaning that time sensitivity decreases as age increases. That could be potentially attributed to individuals being less stressful as they age or that their improved occupational position enables them to be more flexible with their work schedule.

Table 2: Outputs of base MNL and mixed MNL models

| **Fit statistics** | **Base MNL** | | **Mixed MNL** | |
|---|---|---|---|---|
| *Number of individuals* | 540 | | | |
| *Number of observations* | 12,524 | | | |
| *Log-likelihood (0)* | -14,974.45 | | | |
| *Log-likelihood (model)* | -4,554.805 | | -3,401.148 | |
| *Adjusted* $\rho^2$ | 0.6927 | | 0.7691 | |
| *AIC* | 9,203.61 | | 6,914.3 | |
| *BIC* | 9,553.07 | | 7,330.68 | |
| **Parameter** | **Base MNL** | | **Mixed MNL** | |
| | **Estimate** | **Rob. s.e.** | **Estimate** | **Rob. s.e.** |
| Alternative-specific constants | | | | |
| *Constant Bus* | -1.2881 | 0.5195 | -1.1993 | 0.5336 |
| *Constant Bus shift for personal income>50k* | -1.1251 | 0.8598 | -3.7820 | 0.5592 |
| *Constant Bus shift for weekend* | -0.6553 | 0.1773 | -0.9027 | 0.2472 |
| *Constant Bus shift for unemployed individuals* | 0.7659 | 0.7068 | 0.9377 | 0.6303 |
| *Constant Rail* | 1.5491 | 1.2068 | 1.4606 | 1.0056 |
| *Constant Taxi* | -1.7511 | 0.4737 | -1.6004 | 0.7625 |
| *Constant Taxi shift for male* | -0.6777 | 0.2791 | -0.5530 | 0.4405 |
| *Constant Taxi shift for age 18-24* | 1.8740 | 0.3668 | 2.6163 | 0.6202 |
| *Constant Taxi shift for age 25-29* | 1.2315 | 0.3702 | 1.9059 | 0.4610 |
| *Constant Taxi shift for lower education levels* | -1.3278 | 0.5271 | -1.8896 | 0.7506 |
| *Constant Taxi shift for personal income 40k-50k* | -0.7643 | 0.3528 | -1.0749 | 0.5581 |
| *Constant Cycling* | -4.2831 | 0.5517 | -4.6385 | 0.6567 |
| *Constant Cycling shift for male* | 1.0556 | 0.4014 | 1.9054 | 0.4982 |
| *Constant Cycling shift for personal income 10k-20k* | 0.7866 | 0.4022 | 1.9500 | 0.4933 |
| *Constant Cycling shift for personal income 75k-100k* | 3.3556 | 0.9845 | 5.4345 | 1.1033 |
| *Constant Cycling shift for not reported income* | -2.1447 | 1.4537 | -1.5846 | 1.1415 |
| *Constant Cycling shift for weekend* | -0.8289 | 0.3269 | -1.4539 | 0.5118 |
| *Constant Cycling shift for student* | 1.2804 | 0.5169 | 2.4162 | 1.0758 |
| *Constant Cycling shift for unemployed individuals* | -1.3606 | 0.5488 | -5.1260 | 0.6626 |
| *Constant Walking* | 2.8632 | 0.6160 | 3.3554 | 0.6615 |
| *Constant Walking shift for age 18-29* | 0.9189 | 0.3692 | 1.2792 | 0.8351 |

Continued on next page

Table 2 – continued from previous page

| **Parameter** | **Base MNL** | | **Mixed MNL** | |
|---|---|---|---|---|
| | **Estimate** | **Rob. s.e.** | **Estimate** | **Rob. s.e.** |
| *Constant Walking shift for lower education levels* | -0.8210 | 0.2233 | -1.1357 | 0.2550 |
| *Constant Walking shift for weekend* | -0.6254 | 0.1994 | -0.7236 | 0.2879 |
| *Constant Walking shift for student* | 0.7642 | 0.3120 | 1.4677 | 1.1081 |
| LOS parameters | | | | |
| *Travel time Car (box-cox)* | -0.5539 | 0.1898 | – | – |
| *Travel time Car shift for commuting* | -0.3194 | 0.0968 | -0.4344 | 0.1125 |
| *Travel time Car shift for business* | -0.0445 | 0.0815 | -0.3240 | 0.1106 |
| *IVT Bus (linear)* | -0.0488 | 0.0065 | – | – |
| *IVT Bus shift for commuting* | -0.0007 | 0.0049 | 0.0061 | 0.0068 |
| *IVT Bus shift for business* | 0.0175 | 0.0091 | -0.0175 | 0.0123 |
| *IVT Rail (linear)* | -0.0346 | 0.0122 | – | – |
| *IVT Rail shift for business* | 0.0026 | 0.0223 | -0.0389 | 0.0373 |
| *Travel time Taxi (box-cox)* | -0.9412 | 0.2976 | – | – |
| *Travel time Cycling (box-cox)* | -0.5696 | 0.2048 | – | – |
| *Travel time-Age elasticity for Cycling* | -0.3148 | 0.2260 | -0.5840 | 0.2372 |
| *Travel time Walking (box-cox)* | -1.0894 | 0.3707 | – | – |
| *Travel time-Age elasticity for Walking* | -0.0895 | 0.0830 | -0.1331 | 0.0766 |
| *Travel time-Age elasticity for Car, IVT Bus, IVT Rail, Taxi* | -0.2338 | 0.1098 | -0.0275 | 0.1168 |
| *Box-cox lambda for Travel time for Car, Taxi, Cycling, Walking* | 0.3616 | 0.1164 | 0.3985 | 0.1038 |
| *OVT Bus (box-cox)* | -1.1735 | 0.2126 | – | – |
| *OVT Bus for income non respondents* | -1.2168 | 0.2408 | -1.3852 | 0.2094 |
| *OVT Rail (box-cox)* | -1.7890 | 0.5181 | – | – |
| *OVT Rail for income non respondents* | -1.5545 | 0.5154 | -1.8108 | 0.3789 |
| *Box-cox lambda for OVT Bus, OVT Rail* | 0.1180 | 0.0919 | 0.1888 | 0.0644 |
| *Income elasticity for OVT Bus, OVT Rail* | 0.0727 | 0.0376 | 0.0661 | 0.0430 |
| *Travel cost (log)* | -0.8464 | 0.0803 | – | – |
| *Travel cost shift for business* | 0.2811 | 0.1952 | 0.1672 | 0.2360 |
| Random LOS parameters | | | | |
| *a of Travel time Car (box-cox)* | – | – | 0.2195 | 0.2423 |
| *b of Travel time Car (box-cox)* | – | – | -2.2598 | 0.4025 |
| *a of IVT Bus (linear)* | – | – | -1.5341 | 0.2598 |
| *b of IVT Bus (linear)* | – | – | -2.3331 | 0.5843 |
| *a of IVT Rail (linear)* | – | – | -1.5271 | 0.1132 |
| *b of IVT Rail (linear)* | – | – | -4.6136 | 1.1123 |
| *a of Travel time Taxi (box-cox)* | – | – | 1.0313 | 0.2579 |
| *b of Travel time Taxi (box-cox)* | – | – | -1.2045 | 0.1556 |
| *a of Travel time Cycling (box-cox)* | – | – | 2.0943 | 0.3919 |
| *b of Travel time Cycling (box-cox)* | – | – | -3.3665 | 0.7597 |
| *a of Travel time Walking (box-cox)* | – | – | -0.1530 | 0.3560 |
| *b of Travel time Walking (box-cox)* | – | – | 0.7139 | 0.1131 |
| *a of OVT Bus (box-cox)* | – | – | 1.2246 | 0.1446 |
| *b of OVT Bus (box-cox)* | – | – | -1.6430 | 0.1844 |
| *a of OVT Rail (box-cox)* | – | – | 0.2068 | 0.2570 |
| *b of OVT Rail (box-cox)* | – | – | 1.0589 | 0.1625 |
| *a of Travel cost (log)* | – | – | -1.6712 | 0.3711 |
| *b of Travel cost (log)* | – | – | 2.7359 | 0.4583 |

1

## 5. Values of Travel Time estimates

The estimated parameters were applied to the NTS dataset, which provides mobility-related information on a sample of the UK population, and expanded using appropriate distance-based weights in order to calculate VTT values representative for the population of the UK (with the exclusion of London). Distance weighting was performed using mode-specific information on travel distances for work trips obtained from the Census of 2011. The process is detailed below.

Sample level VTT calculation

VTTs were computed using sample enumeration over the car, bus and rail trips of the NTS sample. VTTs are calculated as the relative importance of one unit of change in time relative to one unit of change in cost. In mathematical terms, that is represented as the ratio of the partial derivatives of travel time over travel cost, as shown in *Equation 7*.

$$VTT = \frac{\frac{\partial V_i}{\partial tt_i}}{\frac{\partial V_i}{\partial tc_i}} = \frac{\partial V_i}{\partial tt_i}\frac{\partial tc_i}{\partial V_i} \tag{7}$$

Deriving representative VTT for different distance-bands

In the guidance provided by DfT, the VTT values were weighted based on the trip distance acknowledging the impact of distance on the VTT (Batley et al., 2019). In the current study, a similar approach was followed, with distance weighting factors being derived based on the relative importance of the mode-specific distance bands from the Census 2011 over the NTS distance bands and applied on the sample level VTTs (*Equation 8*). For that purpose, the NTS car, bus and rail trips were allocated to the same eight distance bands as the ones in the Census, namely 0-2km, 2-5km, 5-10km, 10-20km, 20-30km, 30-40km, 40-60km and over 60km, and the trip distributions over those distance bands were calculated for both datasets. The distance weighting factor $w_t$ for trip $t$ was computed as $w_t = \frac{perc^{C}_{d_{t,i,p,m}}}{perc^{NTS}_{d_{t,i,p,m}}}$, where $perc^{C}_{d_{t,i,p,m}}$ amd $perc^{NTS}_{d_{t,i,p,m}}$ are the Census and the NTS distributions of distance band $i$ for trip $t$ of purpose $p$ and mode $m$, respectively [4]. Due to the lack of any additional information regarding non-work trips, the overall trip distance distribution of the Census 2011, regardless of mode, was applied in those cases.

$$\overline{VTT_{p,m}} = \frac{\sum_t (w_t VTT_{t,p,m})}{n_{p,m}} \tag{8}$$

where $VTT_{t,p,m}$ and $\overline{VTT_{p,m}}$ are the VTT for choice task $t$, purpose $p$ and mode $m$ and the weighted average VTT, respectively, while $n_{p,m}$ is the total number of trips for that purpose and mode combination.

Confidence intervals and standard errors for the estimated VTTs were also calculated using multivariate normal draws based on the estimated parameters and the covariance matrix of the behavioural model (Train, 2009). Specifically, 2,000 draws for the estimated parameters were drawn and the VTTs were calculated for each of the 2,000 samples. The 95% confidence interval was then calculated for the resulting sampled VTT distribution per mode and purpose using the percentile interval method and the standard errors were calculated as the standard deviations of the sampled VTT distributions. Finally, the t-stat of the difference between the estimated VTT means were calculated based on *Equation 9*.

$$t - stat_{diff} = \frac{\overline{VTT_{GPS}} - \overline{VTT_{SP}}}{\sqrt{s.e._{GPS}^2 + s.e._{SP}^2}} \tag{9}$$

[4] Details can be found here: https://www.nomisweb.co.uk/census/2011/dc7701ewla

Table 3: Official VTT estimates per mode, purpose and distance band based on the latest SP survey (Batley et al., 2019) and the respective derived GPS-based VTT estimates (£/hour)

| **Distance band** | **Commute trips** | **Other trips** | **Business trips** | | | |
|---|---|---|---|---|---|---|
| | **All modes** | **All modes** | **All modes** | **Car** | **Other PT** | **Rail** |
| Official SP values | | | | | | |
| *All distances* | 11.21 | 5.12 | 18.23 | 16.74 | 8.33 | 27.61 |
| *<20 miles* | 8.27 | 3.62 | 8.31 | 8.21 | 8.33 | 10.11 |
| *20-100 miles* | 12.15 | 6.49 | 16.05 | 15.85 | 8.33 | 28.99 |
| *>=100 miles* | 12.15 | 9.27 | 28.62 | 25.74 | 8.33 | 28.99 |
| GPS-based values | | | | | | |
| *All distances* | 11.92 | 6.62 | 13.29 | 14.15 | – | 20.00 |
| *<20 miles* | 11.12 | 5.63 | 12.45 | 12.70 | – | 13.92 |
| *20-50 miles* | 18.45 | 24.00 | 18.71 | 21.11 | – | 41.41 |

VTTs in the official study were segmented into distance bands of 0-20 miles, 20-100 miles and over 100 miles, however, those were later revised by the DfT to 0-50 miles, 50-100 miles, 100-200 miles and over 200 miles (Batley et al., 2019). The behavioural model in the current study was estimated on a dataset where the maximum trip distance was 61.1 miles, therefore it was decided not to apply the estimates to trips of longer distances. In order to stay in accordance with the official distance segmentations, as much as possible, 50 miles was chosen as the maximum trip distance on which the estimates were applied. Therefore, the derived VTT estimates in the current study range from 0-50 miles and the distance segmentation includes bands of <20 miles and 20-50 miles. In addition to having different distance bands, when comparing the derived VTTs of the current study with the official ones, the exclusion of London is an important factor to be taken under consideration, as well. As a consequence, in the current study there is no "Other PT" as a mode alternative, which was included in the official VTTs and mainly referred to London-specific mode alternatives, such as light rail and the underground.

The official VTT estimates based on the latest nationwide UK SP survey are presented in *Table 3*, both overall and distance segmented values, and are compared with the respective GPS-based VTT estimates of the current study. In the official VTT study, bus was not included as an alternative for business trips, hence we decided to follow the same approach here, as well, for consistency reasons. It can be seen that the overall VTT for "other (non-work)" trips is 6.62, which is very similar to the official value of 5.12, which served as validation during the official study, as reported in Batley et al. (2019). The overall VTT values, which are to be used for appraisal, are generally close to the official ones considering that longer distance trips are missing in the currently derived GPS-based VTTs that would partly explain the lower VTTs for business trips. The distance segmented values are mainly used for reporting purposes (Daly et al., 2014), however, interesting findings can be extracted by their examination. Firstly, the comparison for the <20 miles distance band can provide a direct comparison between the two sets of values, with rail VTTs being the most similar, but the rest of them are also comparable. In general, the VTTs in shorter distance bands have only small discrepancies. Therefore, it can be said that the SP surveys can sufficiently capture individual mobility behaviour in such hypothetical scenarios. For longer distances, the average business VTT across modes and the one specific to car are also fairly close. Starker differences, however, are observed across the remaining VTTs in longer distance bands and especially for "other (non-work)" purpose trips and rail business trips. Specifically, in the last distance band of 20-50 miles, most VTTs in the current study are larger than the ones in the >=100 miles distance band of the official study. That can be attributed to the inherently more unpredictable nature of longer trips that is more challenging to be sufficiently accounted for in the hypothetical setting of SP survey. The differences of VTTs in trips with "other (non-work)" purposes, especially, is significantly large. This can be an SP data issue (i.e. the respondents having difficulty in visualising non-work trips) or it may be due to the lack of explicit information applied during distance weighting for those types of trips.

The standard errors of the estimated VTT means and their 95% confidence intervals for the overall VTTs are presented in *Table 4* along with the t-statistic of the difference of the means between the GPS-based and the official SP-based VTTs. Overall, standard errors of the GPS-based VTTs are higher and that can be attributed to a higher degree of heterogeneity captured in our study compared to the SP survey or it could

Table 4: Confidence intervals and standard errors of the mean estimates for the overall official VTT estimates per mode and purpose (Batley et al., 2019; Hess et al., 2017) and the respective derived GPS-based VTT estimates (£/hour)

| **Mode-Purpose** | **VTT** | **St.error** | **95% Confidence Interval** | | **t-stat diff** |
|---|---|---|---|---|---|
| | | | **Lower bound** | **Upper bound** | |
| Car SP-based values | | | | | |
| *Commuting* | 11.70 | 1.97 | 7.84 | 15.56 | – |
| *Business* | 16.74 | 1.95 | 12.91 | 20.57 | – |
| *Other* | 4.91 | 1.76 | 1.47 | 8.35 | – |
| Bus SP-based values | | | | | |
| *Commuting* | 3.15 | 0.46 | 2.25 | 4.05 | – |
| *Other* | 3.26 | 0.41 | 2.45 | 4.06 | – |
| Rail SP-based values | | | | | |
| *Commuting* | 12.42 | 0.87 | 10.73 | 14.12 | – |
| *Business* | 27.61 | 2.39 | 22.93 | 32.29 | – |
| *Other* | 8.68 | 0.61 | 7.49 | 9.86 | – |
| Car GPS-based values | | | | | |
| *Commuting* | 12.53 | 2.33 | 8.53 | 17.70 | 0.273 |
| *Business* | 14.15 | 7.03 | 6.47 | 31.88 | -0.355 |
| *Other* | 6.75 | 1.52 | 4.32 | 10.17 | 0.793 |
| Bus GPS-based values | | | | | |
| *Commuting* | 5.69 | 1.09 | 3.84 | 7.98 | 2.145 |
| *Other* | 4.09 | 0.68 | 2.94 | 5.51 | 1.043 |
| Rail GPS-based values | | | | | |
| *Commuting* | 8.07 | 2.78 | 4.40 | 14.74 | -1.495 |
| *Business* | 20.00 | 9.20 | 3.66 | 38.49 | -0.800 |
| *Other* | 9.50 | 3.23 | 5.18 | 17.35 | 0.250 |

also simply be due to the smaller sample size. For almost all VTTs presented in *Table 4*, we cannot reject the null hypothesis of difference from the official valuations, with the exception of bus VTT for commuting trips, which can be attributed to several factors, such as the low number of observations for those trips and the fact that stated bus fares from the NTS dataset were used to calculate the VTTs, while imputed fares were used during the estimation of the behavioural model. Finally, as mentioned before, the lack of longer distance trips could be a reason for the lower commuting and business VTTs in our analysis, where rail is more likely to be preferred for those types of trips on longer distances, although those differences are still not statistically significant.

## 6. Conclusions

The current paper presents a study of deriving VTT estimates in a manner comparable to the official values currently used in appraisal. Though the level of detail included in the initial GPS trip diaries provided challenges during the data cleaning phase, there were significant advantages in terms of accuracy. For example, the time-stamped geo-locations provided the ability to better capture individual mobility behaviour by making it possible to get precise travel times for the chosen modes and to extract travel times between specific latitude/longitude pairs (opposed to between TAZ centroids as done in traditional RP survey data). It also enabled the estimation of more behaviourally rich models by offering a more comprehensive representation of tours, where even very short stops and/or trips have been included.

It may be noted though that the overall sample size of the finally utilised dataset was smaller compared to the SP survey, as a large share of trips had to be removed from the original DECISIONS data during the cleaning phase in order to exclude inconsistent, incomplete and untagged trips. Furthermore, the trips recorded in the DECISIONS dataset were mostly urban trips, while the official SP survey included longer distance trips, as well. That hindered the comparison of the VTT values for longer distance bands in the current study, but such limitations can be overcome in nationally important studies by designing a more comprehensive data collection process.

Despite those limitations, the study has two key findings which are of importance to transport planners and policy makers:

1. The study demonstrates that the overall VTT estimates were similar to the official SP-based values used for appraisal, even with a smaller sample size.
2. Segmenting the VTTs by distance bands, larger discrepancies start to become evident among longer distance bands with the SP-based VTTs being smaller in most cases. That hints to a downward bias for SP surveys potentially originating from the hypothetical nature of the longer trips, which made them difficult to comprehend. In contrast, the smaller differences of VTTs for shorter distance bands could mean that the design of the SP survey was sufficient enough to capture realistic mobility behaviour.

The results hence demonstrate that by harnessing recent technological advances in data collection, transport planners and policy-makers can make a successful shift to RP data sources, which have more behavioural validity compared to SP. Furthermore, the findings of the current study also demonstrate that smaller sample sizes derived from GPS smartphone data could be sufficient for the estimation of behaviourally accurate VTTs for the whole population. That finding could lead to a more frequent data collection process for the purpose of updating the national VTT estimates, compared to the so far slower update rate of traditional SP-derived VTTs (approximately every 10-20 years). GPS-derived VTTs could also be used by policy-makers to complement the official SP-derived VTTs, since the more frequent GPS studies could help to detect any significant deviations from the previously SP-based estimated VTTs due to income increase or other unforeseen circumstances that could occur in the meantime, such as economic recession or the introduction of new disruptive modes/technologies into the transport market. Technologies like online shopping and its ever-increasing popularity especially in a post-Covid world, electric vehicles, shared ride modes (Uber, Lyft etc.) and policy initiatives like Mobility as a Service are constantly changing the transport sector, which has become more volatile than ever before. Transport is rapidly changing and the usual update rate of nation-wide official VTTs might be too slow nowadays to provide insights into the current trade-offs or even capture the sensitivities on new technologies in hypothetical scenarios and in a behaviourally realistic manner. As a result, new transport projects might not be properly evaluated if the appraisal is based on individual trade-offs that no longer represent the current behaviour of the target population.

The findings in the current study can provide practitioners and policy makers with additional confidence when it comes to using new emerging data sources for future nationwide VTT studies. The small differences across the overall VTT estimates, regardless of distance bands, showcase that RP data captured through new emerging data collection methods –GPS in this case– can provide behaviourally reasonable VTT estimates that are in line with the official SP-based values currently used in appraisal. Of course, a reader may ask why RP data should be used if the results are no different from SP data. The simple answer to this question is that RP data provides the truth, and the fact that the findings in this case are in line with the SP results thus arguably also serves as a validation of SP rather than RP. Furthermore, our results are achieved using a much smaller sample size during estimation, compared to the SP study, which can lead to a reduced cost or more frequently performed surveys in general. Performing a large scale GPS-based RP study at the country level will result in a significantly more accurate representation of individual mobility behaviour, capturing choices over a large number of real-life scenarios, independent of the researcher's assumptions, while also resulting in less fatigue for the respondents.

This study comes at a time where ubiquitous sensing data sources are steadily gaining ground in transport research and provides empirical evidence for their further adoption into the field of practice. Nonetheless, more studies are required offering similar practical applications, even in small sample sizes, before we can see a departure from the current state of practice that has been dominant over the last several decades.

## Acknowledgements

The current research was funded by the Advanced Quantitative Methods (AQM) scholarship of the Economic and Social Research Council (ESRC). Charisma Choudhury acknowledges the financial support of her UKRI Future Leader Fellowship MR/T020423/1-NEXUS. Stephane Hess acknowledges the financial support by the European Research Council through the consolidator Grant 615596-DECISIONS.

## References

Accent and Hague Consulting Group 1996. *The value of travel time on UK*. London: Department of Environment, Transport and Regions.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M.O., Chintakayala, P.K., Choudhury, C.F., Daly, A., Dekker, T., Drivyla, E., Fowkes, T., Hess, S., Heywood, C., Johnson, D., Laird, J., Mackie, P., Parkin, J., Sanders, S., Sheldon, R., Wardman, M. and Worsley, T. 2019. New appraisal values of travel time saving and reliability in Great Britain. *Transportation*. **46**, pp.583–621. 10.1007/s11116-017-9798-7.

Arifin, Z.N. and Axhausen, K.W. 2012. Investigating Commute Mode and Route Choice Variabilities in Jakarta Using Multi-Day GPS Data. *International Journal of Technology*. **3**(1). 10.14716/ijtech.v3i1.1087.

Beesley, M.E. 1965. The Value of Time Spent in Travelling: Some New Evidence. *Economica*. **32**, pp.174–185. 10.2307/2552547.

Ben-Akiva, M. and Lerman, S. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Massachusetts: MIT Press.

Brownstone, D. and Small, K.A. 2005. Valuing time and reliability: Assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*. **39**, pp.279–293. 10.1016/j.tra.2004.11.001.

Bwambale, A. and Choudhury, C.F. and Hess, S. 2019. Modelling Long-Distance Route Choice Using Mobile Phone Call Detail Record Data: A Case Study of Senegal. *Transportmetrica A: Transport Science*. **15**(2), pp.1543–1568. 10.1080/23249935.2019.1611970.

Calastri, C., Crastes dit Sourd, R. and Hess, S. 2020. We want it all: Experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. *Transportation*. **47**, pp.175–201. 10.1007/s11116-018-9858-7.

Calastri, C., Hess, S., Choudhury, C.F., Daly, A. and Gabrielli, L. 2018. Mode choice with latent availability and consideration: Theory and a case study. *Transportation Research Part B: Methodological*.

Daly, A. 2010. *Cost Damping in Travel Demand Models - Report of a study for the Department for Transport*. RAND Europe.

Daly, A., Tsang, F. and Rohr, C. 2014. The Value of Small Time Savings for Non-business Travel. *Journal of Transport Economics and Policy*. **48**(2), pp.205–218.

Daly, A. and Zachary, S. 1975. *Commuters' values of time*. Reading: Local Government Operational Research Unit.

Department for Transport 2014. *Values of Time and Vehicle Operating Costs TAG Unit 3.5.6*. Department for Transport.

Department for Transport 2015. *Provision of market research for value of travel time savings and reliability - Phase 2 Report*.

Fosgerau, M. and Jensen, T.L. 2003. Economic appraisal methodology - controversial issues and Danish choices *In*: *European Transport Conference, 4-6 October*. Strasbourg, France.

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H. and Lancsar, E. 2021b. Hypothetical bias in stated choice experiments: Part I. Integrative synthesis of empirical evidence and conceptualisation of external validity.

Hess, S., Daly, A. and Börjesson, M. 2020. A critical appraisal of the use of simple time-money trade- offs for appraisal value of travel time measures. *Transportation*. **47**, pp.1541–1570. 10.1007/s11116-020-

10097-w.

Hess, S., Daly, A., Dekker, T., Cabral, M.O. and Batley, R. 2017. A framework for capturing heterogeneity, heteroscedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological*. **96**, pp.126–149. 10.1016/j.trb.2016.11.002.

Hess, S., Train, K.E. and Polak, J.W. 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. *Transportation Research Part B: Methodological*. **40**(2), pp.147–163. 10.1016/j.trb.2004.10.005.

Huang, Y., Gao, L., Ni, A. and Liu, X. 2021. Analysis of travel mode choice and trip chain pattern relationships based on multi-day GPS data: A case study in Shanghai, China. *Journal of Transport Geography*. **93**. 103070.

Kang, M.J. and Camerer, C.F. 2013. fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Frontiers in Neuroscience*. **7**(104). 10.3389/fnins.2013.00104.

Krcal, O., Peer, S., Stanek, R. and Karlinova, B. 2019. Real consequences matter: Why hypothetical biases in the valuation of time persist even in controlled lab experiments. *Economics of Transportation*. **20**(100138). 10.1016/j.ecotra.2019.100138.

Li, Z., Hensher, D.A. and Ho, C. 2020. An empirical investigation of values of travel time savings from stated preference data and revealed preference data. *Transportation Letters The International Journal of Transportation Research*. **12**(3), pp.166–171. 10.1080/19427867.2018.1546806.

Loewenstein, G. 2005. Hot-cold empathy gaps and medical decision making. *Health Psychology*. **24**(4S), pp.49–56. 10.1037/0278-6133.24.4.S49.

Louviere, J.J. and Hensher, D.A. 2001. Combining sources of preference data *In*: D. A. Hensher, ed. *Travel Behaviour Research: The Leading Edge*. Pergamon Press: Oxford, pp.125–144.

Mackie, P.J., Jara-Diaz, S. and Fowkes, T. 2001. The value of travel time savings in evaluation. *Transportation Research Part E: Logistics and Transportation Review*. **37**(2-3), pp.91–106. 10.1016/S1366-5545(00)00013-2.

Mackie, P.J., Wardman, M., Fowkes, A.S., Whelan, G.A. and Nellthorp, J. 2003. *Values of Travel Time Savings in the UK. Report to Department for Transport*. Institute for Transport Studies, University of Leeds.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior *In*: P. Zarembka, ed. *Frontiers in Econometrics*. New York: Academic Press, pp.105–142.

McFadden, D. and Train, K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*. **15**, pp.447–470.

Montini, L., Antoniou, C. and Axhausen, K.W. 2017. Route and mode choice models using GPS data *In*: *Transportation Research Board 96th Annual Meeting*. 8-12 January 2017, Washington D.C., United States.

MVA, ITS Leeds and TSU Oxford 1987. The value of travel time savings', Newbury Berkshire. *Policy Journals*.

Nielsen, O.A. 2004. Behavioral responses to road pricing schemes: Description of the Danish AKTA experiment. *Intelligent Transportation Systems*. **8**(4), pp.233–251. 10.1080/15472450490495579.

Schuessler, N. and Axhausen, K.W. 2009. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*. **2105**(2105), pp.28–36. 10.3141/2105-04.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge, Massachusetts: Cambridge University

Press.

Zhang, G. and Laird, J.J. 2014. Local values of travel time savings - should we forget about them? *In*: *The 19th International Conference of Hong Kong Society for Transportation Studies*. Hong Kong, 13-15 December, 2014.