USING ROAD-SIDE INTERVIEWS TO MODEL EMPLOYER'S BUSINESS AND NON-HOME-BASED TRAVEL IN THE WEST MIDLANDS

James Fox, Marco Kouwenhoven, Andrew Daly
RAND Europe
Hugh Gunn
HGA Ltd Cambridge

1. INTRODUCTION

Travel made in the course of an individual's work forms a key component of total demand. While the number of such employer's business trips may be small relative to other travel purposes, such trips are typically much longer than the overall average, and are more likely to be made by car than the overall average. Furthermore appraisal procedures attribute a high value of time to such trips, which implies that they should be forecast with particular accuracy, as they can have a significant impact on the benefits associated with a particular scheme.

A standard form of surveying for the development of choice models for large urban areas is to use home interview surveys. Such surveys have the advantage of collecting detailed person and household information, which makes the data suitable for the development of detailed choice models. However the use of such surveys is problematic for modelling employer's business travel. Firstly, because employer's business travel is a small fraction of overall household travel, the volume of data recorded is typically small which makes the development of detailed models difficult. Secondly, there is often confusion of employer's business with commuting and personal business, which may further reduce the volume of employer's business travel recorded. In the West Midlands, for which the models described in this paper were developed, very little employer's business travel was recorded in the home interview data.

Non-home-based travel contains a significant component of employer's business travel, but is also often poorly reported in home interview surveys, even for other travel purposes. In the West Midlands data, non-home-based travel was indeed under-reported, but the same data source and modelling approach developed for home-based employer's business could also be applied efficiently to other non-home-based travel, for both business and other travel purposes

A convenient source of transport data is *en route* surveys, such as road-side interviews for car, or on-board surveys for public transport. Non-responses and no-contacts in such surveys are lower than for home interview surveys, and so the biases introduced into the data are determined more strongly by the survey design.

With *en route* surveys, the probability of a trip being surveyed depends upon the traveller's origin, route, mode, destination and potentially on the time of the journey. For public transport on-board surveys, it can be difficult to determine what the probability of a traveller making specific trips would be. However, road-side interviews are undertaken at well defined points, often forming a cordon, which allows the probability of intercepting a trip with a given origin and destination to be determined. Because the bulk of employer's business travel is made by car, this paper focuses on the development of choice models from road-side interview data. With this type of data the intercept probabilities can be defined reasonably clearly; for public transport travel the issues are typically more complicated, unless the surveys have been explicitly designed to interview travellers crossing cordons.

A key concern in the development of choice models from road-side interview data is the need to account for the sampling procedure used. The probability of a trip being intercepted at a road-side interview site depends on the length of the trip and therefore the trip length distribution for road-side interview data is different to the distribution for the household based data. While the procedures for dealing with complex choice-based samples were set out in theory some time ago, they have not been used widely in practice and the implications of their use for travel demand modelling have not been fully established.

The work presented in this paper was undertaken as part of the development of the PRISM model system for the West Midlands area of the United Kingdom. This model system has been developed by RAND Europe and Mott MacDonald on behalf of the seven local authorities in the West Midlands Metropolitan County and the Highways Agency. The model area covers some 5300 sq. km. and has a population of about 3,950,000, with its main centre in Birmingham.

This paper describes analysis of the volume of employer's business and non-home-based data in the available household interview data, and proceeds to analyse the data available from road-side interview data, highlighting the biases in the road-side interview data. The paper discusses the approaches that can be used to deal with choice-based data, and then goes on to describe the particular approach used in the West Midlands study. The procedure adopted identifies the routes that can be used for travel between origins and destinations, and identifies the screenlines that would be crossed for these journeys, enabling the probability of interception to be represented in the modelling.

2. DATA

2.1 Household Interview Data

The 2001 household interview data was collected as part of the West Midlands Transportation Surveys 2001. A travel diary was used in the survey, which recorded all journeys made by persons on the survey day. The diary is completed by all household members aged five years and above, and person and household information is collected together with travel information. For the model system developed for the West Midlands study (the PRISM model system) information from around 27,000 persons and 11,700 households was available for analysis. This interview clearly forms a substantial data base.

The household interview data has been used for the development of the majority of the choice models used in the PRISM model system, specifically models of car ownership, public transport pass ownership, travel frequency and mode and destination choice. The detailed person and household information recorded, such as information on the age, licence holding and employment status of individuals, and the number of cars and licences owned by the household, make the data suitable for the estimation of detailed disaggregate choice models. These models incorporate parameters to reflect differences in choice preferences across person and household segments, which enables policies which impact on specific segments of the population to be tested.

2.1.1 Analysis of Business Shares in the Household Interview Data

The unit of travel in the mode-destination models in the PRISM model system is home-based *tours*. A home-based tour is a series of linked journeys starting and finishing at home. Each tour is associated with a main purpose, which is determined by the *Primary Destination* (PD) of the tour. Most tours consist of two journeys, from the home to the PD, and then back home again. However if more than one non-home destination is visited during a tour, then a hierarchy is used to determine the PD, with work at the top, followed by employer's business and then all other purposes. Ties are resolved by taking the destination at which the most time was spent. A review of tour based approaches is given by Gunn *et al.* (2001).

The numbers of tours by purpose recorded in the 2001 household interview data are summarised in Table 1, for all modes and for car-driver only. The tour purpose shares are also presented.

Table 1: Tour Purpose Shares in the 2001 Household Interview Data

Purpose	All Modes Shares		Car Driver only Shares	
Commuting	7,773	35.6 %	5,281	54.0 %
Employer's Business	141	0.6 %	114	1.2 %
Education	4,699	21.5 %	441	4.5 %
Shopping	5,178	23.7 %	1,948	19.9 %
Personal Business	2,033	9.3 %	1,017	10.4 %
Visiting Friends	1,046	4.8 %	538	5.5 %
Recreation/leisure	973	4.5 %	445	4.5 %
Total	21,483	100.0 %	9,784	100.0 %

Only 0.6 % of tours are for employer's business, and the total volume of tours (141) is insufficient to enable the estimation of a detailed tour based model, or to allow aggregate matrix estimation techniques to be employed. The car driver only share is higher, because a high percentage of employer's business tours are made by car driver.

A common problem in such surveys is confusion between employer's business travel and commute and personal business travel. Therefore analysis was undertaken of the home interview data to investigate whether travel recorded as 'work' could be recoded as 'employer's business'.

This analysis applied a series of rules to look at destinations visited in the course of commute tours. If two different destinations recorded as work are visited during a tour, then the main workplace is taken to be the location at which the most time was spent. The location of the second workplace visited during the tour can then be examined. If the second workplace visited is in a different location to the main workplace, then it must be a different workplace and so can be recoded as an employer's business location. The journey purpose shares before and after recoding are highlighted in Table 2.

Table 2: Journey Purpose Shares after Recoding

	Before Recoding		After Recoding		Difference	
Home	22,527	47.0%	22,527	47.0%	0	0.0%
Workplace	8,576	17.9%	8,432	17.6%	-144	-1.7%
Employer's business	255	0.5%	399	0.8%	144	56.5%
Education	5,184	10.8%	5,184	10.8%	0	0.0%
Shopping	5,903	12.3%	5,903	12.3%	0	0.0%
Personal business	2,793	5.8%	2,793	5.8%	0	0.0%
Visit friends	1,445	3.0%	1,445	3.0%	0	0.0%
Recreation/leisure	1,224	2.6%	1,224	2.6%	0	0.0%
Total	47,907	100.0%	47,907	100.0%	0	0.0%

As a result of the analysis to recover miscoded journeys, the number of employer's business journeys increased by more than 50 %, from 255 to 399. The overall business share remains low however, at 0.8 % of all journeys and 1.6 % of journeys with a non-home destination.

This results of this processing were evidence that a significant volume of employer's business travel had been miscoded as travel to work. However, the processing could only recover employer's business journeys made during work tours. It is also likely that a significant proportion of work tours (and possibly of personal business tours) are in fact employer's business tours. For example if an individual travels from home to a business location other than their main workplace, and then returns directly home, this should be an employer's business tour, but may have been miscoded as work. The processing cannot 'recover' such tours.

Following this recoding, analysis was made to compare the employer's business share in the household interview data to other data sources. The aim of this analysis was to determine the extent to which employer's business travel had been under-reported in the household interview data. Home-based trips were only included in the analysis if the home end lay in the model area, non-home-based trips were only included if one end of the trip lay in the model area. Thus external trips were excluded from the analysis.

A comparison was undertaken of *kilometrage* shares recorded in the household interview data to 2001 road-side interview data, which is discussed in more detail in Section 2.2. The road-side interview data only interviews travel made by car, and so the shares for the household interview data are based on car driver observations only.

Table 3: Comparison of Employer's Business Kilometrage Shares

	Household Interview		Road-Side Interview	
Share of:	Obs.	KM share	Obs.	KM share
Home-Based	302	4.9 %	11,462	9.6 %
Non-Home-Based	125	27.2 %	19,925	40.9 %
Total	427	6.0 %	31,387	18.6 %

This comparison suggests that the 2001 household interview data only records about one-third of the true business kilometrage. However, when only home-based journeys are considered the figure is closer to a half, suggesting the lack of recording of non-home-based business travel is a greater problem than the miscoding of home-based journeys as work. Indeed the volume of non-home-based travel recorded in the 2001 household interview data is low for all journey purposes. Only 5.2 % of car driver journeys recorded in the home interview were non-home-based, compared to 28.9 % in the 2001 road-side interview data.

The conclusion from these analyses was that the business share in the 2001 household interview data was significantly under-reported.

2.1.2 Comparison of Business Shares with Other Data Sources

The observed journey purpose shares were also compared with data from the 1998/2000 National Travel Survey (NTS) data. To ensure comparability, journeys from the West Midlands household interview data were classified into purposes using the rules applied in the NTS, and the seven day NTS data specific to the West Midlands was re-scaled to reflect an average weekday. The conclusion of this analysis was that only one-third of the expected volume of employer's business travel has been recorded in the 2001 household interview data.

Comparison was also made of the employer's business share in another large UK urban area. This comparison demonstrated the share of employer's business travel recorded in the London LATS data to be more than four times higher than the HI data. However, London may have a higher business share than average.

A further comparison was made of the volume of employer's business travel in other international data sets. Data from the Netherlands and Sydney, Australia was analysed. This comparison again demonstrated the volume of employer's business data in the household interview data to be very low in comparison with other data sources.

As the analysis above has noted, the volume of non-home-based travel recorded in the household interview data was very low compared to other data sources, both for employer's business and other travel purposes. Indeed the under-reporting of non-home-based travel is a significant reason why the employer's business share is so low.

Overall it was concluded that the volume of employer's business recorded was too low to be credible, and is affected by both miscoding of work travel as employer's business, and a general under-reporting of non-home-based travel. The low volume of employer's business and non-home-based travel meant that there was insufficient data to estimate mode and destination choice models from the household interview data for these purposes, or indeed to employ aggregate matrix estimation methods. Therefore it was decided to use road-side interview data, described in the following section, for the modelling effort instead. Merging employer's business travel with commuting was not considered desirable, given the differences in values-of-time, and the need to represent non-home-based business travel in the model framework.

2.2 Road-Side Interview Data

The 2001 road-side interview (RSI) data was also collected as part of the West Midlands Transportation Surveys 2001. A total of 100 RSI sites were located on main roads across the West Midlands, and a further 46 sites were placed on motorway slip roads to survey traffic joining the motorway network.

Analysis was undertaken of the volume of RSI data available for modelling. At this stage it is useful to define how the purposes were defined for the modelling in the RSI data:

- home-based (HB) employer's business trips have one end at home, and the other end at an employer's business location;
- non-home-based (NHB) employer's business trips have neither end at home, but one or both ends at an employer's business location;
- NHB other trips have neither end at home, and neither end at an employer's business location.

It should be noted that some analysis presented earlier in this paper defined the purpose of a NHB trip on the basis of the destination purpose only.

The mean trip lengths observed in the RSI data were compared to those observed in the much smaller sample of data in the 2001 household interview data. This comparison is presented in the following table.

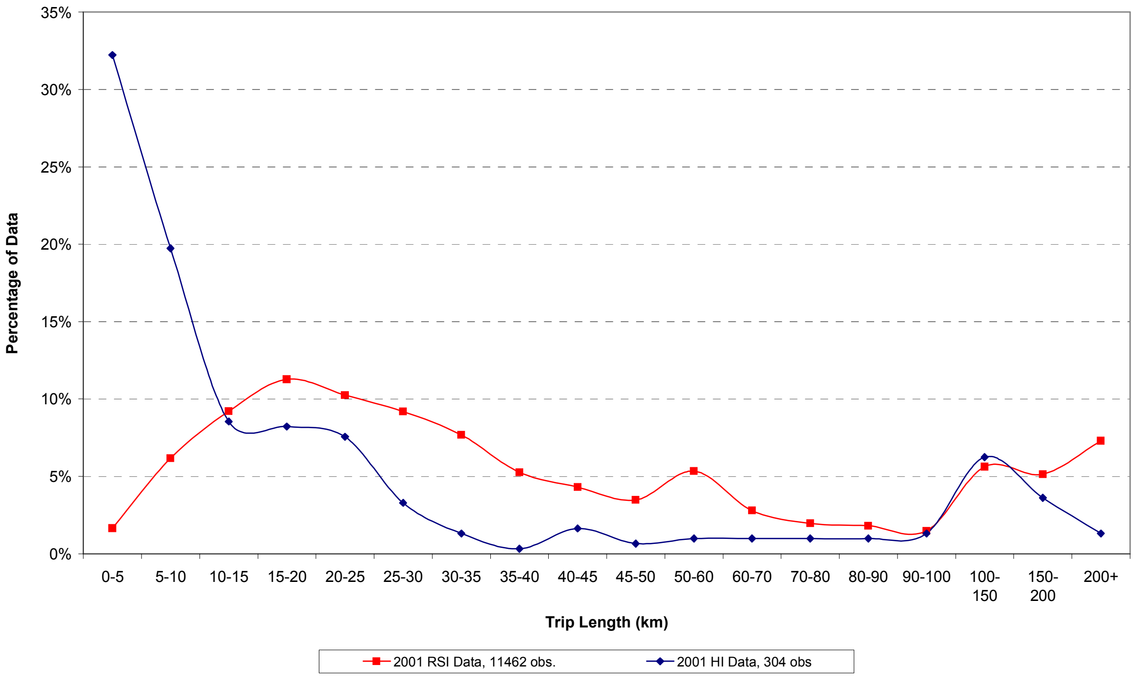
Table 4: Comparison of Mean Trip Lengths (km) by Purpose

	Household Interview Data		Road Side Interview Data	
	Observations	Mean Length	Observations	Mean Length
Home-Based Employer's Business	304	30.0	11,462	41.9
Non-Home-Based Employer's Business	189	19.1	25,764	36.7
Non-Home-Based Other	664	7.5	22,969	20.8

It can be seen that the mean trip lengths observed in the RSI data are considerably longer than those observed in the household interview data, reflecting the increased probability of sampling longer trips in the RSI data.

A comparison was also made of the observed trip length distributions in the household interview and RSI data sources. The comparisons are plotted in the following figures. Although the volume of household interview data is low, the data provides a better estimate of the 'true' trip length distribution, because it is not affected by the choice-based sampling approach that influences the RSI data.

Figure 2: Observed Trip Length Distributions, HB Employer's Business



This figure clearly highlights the low proportion of short distance trips in the 2001 RSI data compared to the 2001 household interview data, with significantly fewer trips in the 0-5 km and 5-10 km distance bands in the 2001 RSI data. There is much more RSI data in the 200+ km band than household interview data.

Figure 3: Observed Trip Length Distributions, NHB Employer's Business

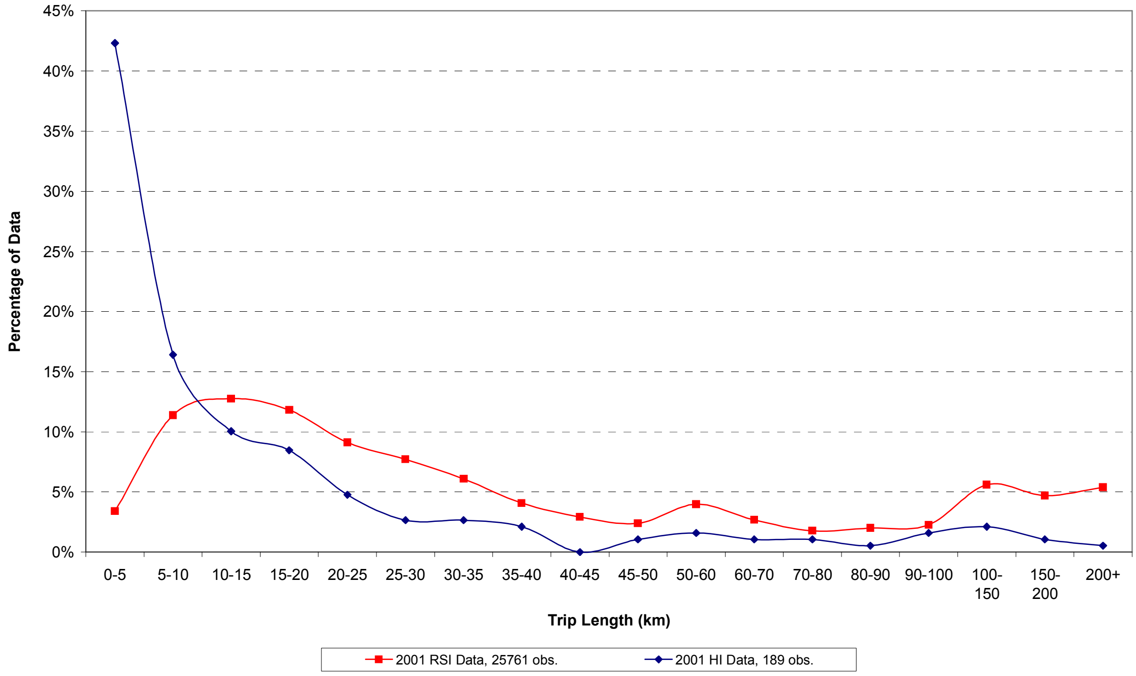
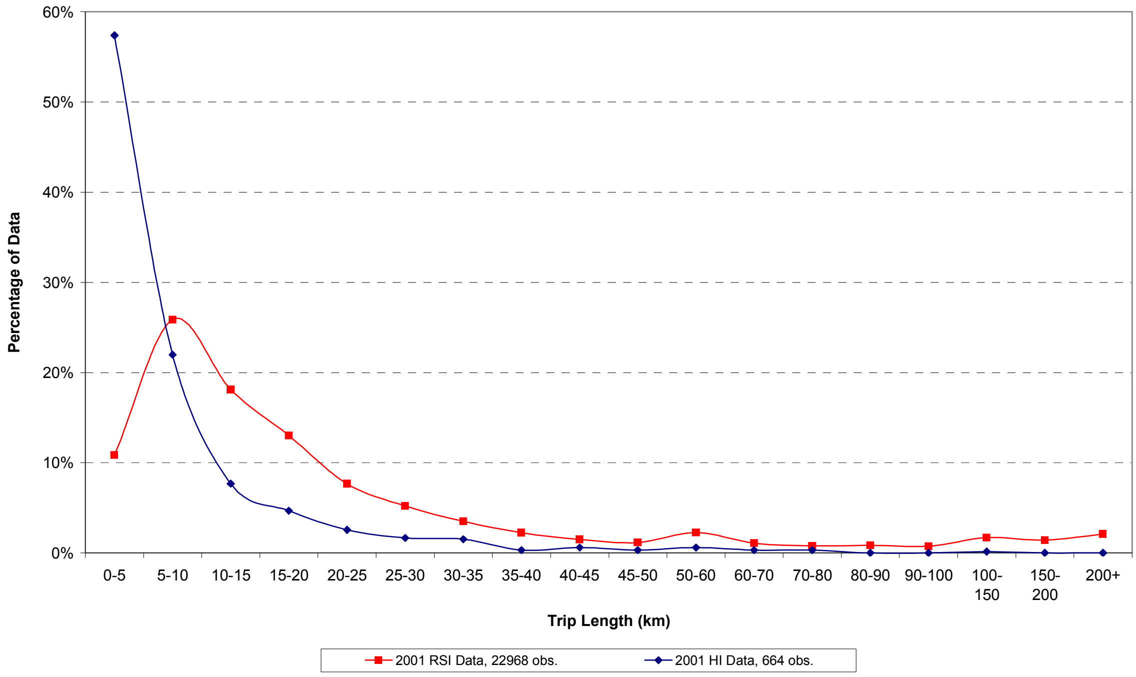


Figure 4: Observed Trip Length Distributions, NHB Other



Both these figures clearly highlight the low proportion of short distance trips in the RSI data relative to the household interview data. For NHB employer's business, trips in the 0-5 km and 5-10 km bands are noticeably lower, whereas for NHB other the under-representation is concentrated in the 0-5 km band. In both plots the percentage of long distance trips, over 100 km in length, is noticeably higher in the RSI data.

It was clear from the analysis of mean trip lengths, and trip length distributions, that the screenline based survey approach results in significant differences in the trip length distributions compared to data sampled independently of choice at the household level. The screenline based surveying approach used means that longer trips have a higher probability of being interviewed. The impact of this effect on the observed trip length distributions is particularly noticeable for short distance trips. A further cause of these differences is that traffic on main roads, on which most RSI sites are located, will tend to contain a higher proportion of longer distance traffic than average.

The implication of this analysis was that it was essential that the differences in RSI trip length distributions relative to the 'true' trip length distributions were corrected for in the modelling. The next chapter describes the approach adopted to develop models from the choice-based RSI sample.

3. MODEL ESTIMATION

3.1 Estimation with Choice-Based Samples

Discrete choice modelling generally aims at estimating the unknown parameters of a model using the criterion of maximum likelihood. This criterion estimates the values of the parameters that would make the observation of the specific data most likely. The likelihood function for a data set can be defined as

$$L = \Pi_{obs} \Pr \{ \text{observe obs} \}$$

or, usually more conveniently,

$$\log L = \Sigma_{obs} \log \Pr \{ \text{observe obs} \}$$

The logarithmic form is more convenient because addition is more convenient than multiplication. Because the logarithmic function increases monotonically, maximising its log has the same effect as maximising the original function.

The probability of a given observation has two components. First, there is the probability that an individual will behave in a certain way and second there is the probability that, given the behaviour, the observation will be captured by the survey. Thus

$$\Pr \{\text{observe behaviour}\} = \Pr \{\text{behaviour}\} . \Pr \{\text{observe, given behaviour}\}$$

In many cases, the probability of being observed does not depend on the behaviour at all. For example, in a home interview survey, it is reasonable to assume that all kinds of travel behaviour by the residents of a household will be observed. The probability of being observed in the survey is uniform across the sample and can be omitted from the maximisation. However, in contexts such as road-side interviews it is necessary to include the probability of observation because it is not uniform. For example, choice of a distant destination makes the chance of being observed in a road-side interview much higher than the choice of a nearby destination.

Because of the need to correct for the effect of the sampling procedure, special estimation procedures have to be used in the modelling. Depending on the specific circumstances, different procedures are available. When the true numbers choosing each alternative are known, weighted or conditional maximum likelihood estimators can be used. However, in this case, we do not know the true numbers of trips or tours with destination in each zone and estimators of the type originally developed by Cosslett (1981) have to be used.

In work by Hague Consulting Group in The Netherlands in 1982-84, the Cosslett estimators were enhanced, simplified and successfully applied for a large-scale model transfer study. The estimators require the data to be split into a number of segments, with each segment having a specific probability of being observed. In this case, the probability of being observed depends on the screenlines that are crossed when a trip from a given origin goes to a given destination. The process of estimation then requires specific correction terms to be estimated for each screenline. The authors believe it is the first time this method has been applied in a UK study.

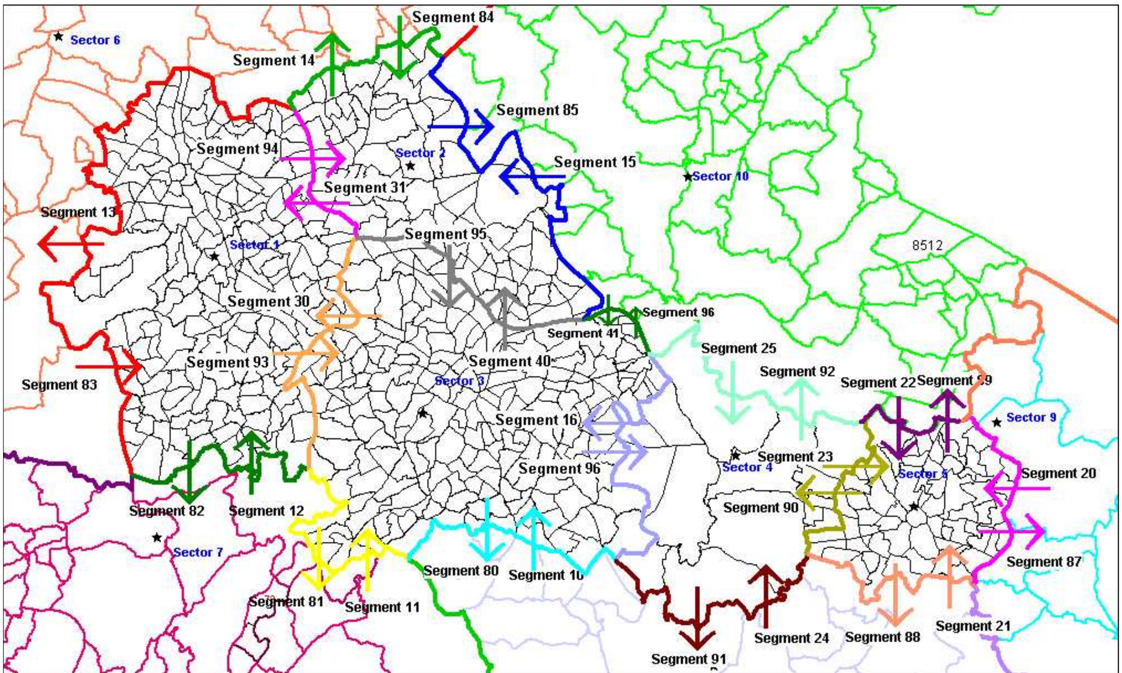
3.2 Estimation with Screenline Correction Terms

As discussed in the previous section, the approach that has been used to deal with the choice-based nature of the road-side interview data was to use screenline correction terms. Screenline correction terms represent the probability of an individual being sampled by a given screenline, and as such correct for the choice-based sampling that exists in the data. The probability of being sampled at a given screenline is proportional to the sampling rate at a screenline, and this sampling rate will vary between screenlines, and may also vary with the time of day.

Analysis was undertaken to investigate variation in the sampling rate at the West Midlands road-side interview sites with time of day, to assess whether it was necessary to explicitly correct for variations in sampling rate with time of day in the modelling. For the main road side interview sites, the sampling rate did not drop noticeably in the peak periods, although there was a slight decline sampling rates as the day proceeded which may be due to interviewer fatigue. For the motorway slip road sites, there was a slight decline in sampling rates in the AM and PM peaks. Overall it was concluded that the time of day variations were not significant enough for warrant explicit correction during model estimation.

A unique screenline correction term was added to the models for each of the screenline segments shown in Figure 5. In the text following the figure, screenline segments are simply referred to as screenlines.

Figure 5: Screenline Segments



In order to estimate the screenline correction terms, it is necessary to determine which screenlines are crossed for a given origin-destination (OD) pair. For the sampled trip, the individual has clearly been sampled at a given location, but it is possible that they could make a long trip and cross another screenline without being sampled. Furthermore in order to estimate choice models, it is necessary to know which screenlines would be crossed for other choices of destination zone.

To determine the information on which screenlines are crossed for a given OD pair, an all-or-nothing shortest path assignment was made on a congested network loaded with car, light-goods-vehicle and heavy-goods-vehicle matrices. When the models were estimated, it was necessary to take account of the fact that OD pairs can be observed only if they cross a screenline according to the assignment. It should be noted that in some cases the individual had been surveyed at an RSI site, but the assignment suggested that for the chosen OD pair a different route would be taken which did not cross a screenline. Such observations were excluded from the modelling. It should also be noted that it was possible for a given OD pair to pass more than one screenline according to the assignment.

It is important to emphasise that the screenline correction terms are *only* used in model application, in order that the choice-based bias is represented during model estimation. When the models are applied, so that the predicted trip length distributions can be investigated, the screenline correction terms are dropped from the model. A second key difference is that in model estimation, only OD pairs which cross at least one screenline are included as alternatives in the models, whereas in model application all OD pairs are available irrespective of whether or not they cross a screenline.

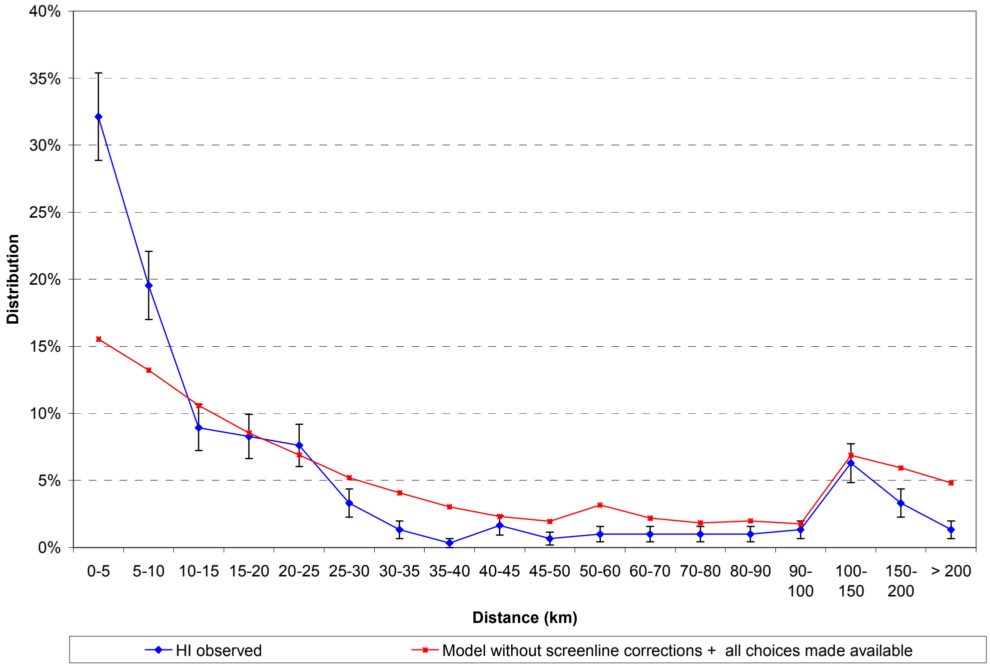
3.3 Model Development

The majority of the development work described in this section was undertaken on the home-based employer's business model. Once the model tests were complete, similar model formulations were tested for non-home-based (NHB) employer's business and NHB other, and then these models were refined accordingly. The RSI data only collected limited information on the purpose, origin and destination of the trip. Therefore the choice models developed were simple in nature, with no socio-economic parameters incorporated.

Once the screenline correction terms had been added to the home-based employer's business model, model validation runs were made to compare the predicted trip length distributions to those observed in both the 2001 household interview data and the 2001 RSI data. The aim of these comparisons was to determine the impact that the addition of the screenline correction terms had upon the predicted trip length distributions.

The comparison of the model predictions to the observed trip length distribution in the 2001 RSI data demonstrated that the trip length distribution predicted by the model closely matched the observed RSI distribution. Furthermore, the addition of the screenline correction parameters did not have much impact on the predicted trip length distribution. This latter finding was unexpected, as it was believed that the addition of the screenline correction parameters would overcome the bias towards longer trips in the observed data. The predicted trip length comparison was also compared to the observed distribution in the 2001 household interview data. This comparison is given in Figure 6. Error bars have been plotted for the household interview data due to the low volume of available data (304 observations).

Figure 6: Comparison of Screenline Correction Model Predictions to 2001 Household Interview Data



This comparison clearly demonstrates that the introduction of screenline correction parameters had not overcome the low representation of short distance trips in the RSI data.

Given these discrepancies, model tests were made with distance correction parameters added in *addition* to the screenline correction parameters. In the first model test made, correction parameters were added for four distance bands where the under-representation in the RSI data is most noticeable:

- 0 – 5 km
- 5 – 10 km
- 10 – 15 km
- 15 – 20 km

The estimated values for these correction parameters are given in Table 5. It is emphasised that these parameters are only used during model estimation to correct for differences in the trip length distributions. When the models are applied, the distance correction parameters are dropped.

Table 5: Distance Correction Parameter Estimates

Parameter	Distance Band	Estimate (t-ratio)
Dist_1	0 – 5 km	- 1.670 (12.1)
Dist_2	5 – 10 km	- 0.940 (13.0)
Dist_3	10 – 15 km	- 0.571 (11.2)
Dist_4	15 – 20 km	- 0.282 (6.9)

As expected all parameters are negative, and the shorter the distance band the more negative the parameter as the probability of being surveyed at a screenline declines further. The predicted trip length distribution from this model was determined and compared to the observed distribution in the 2001 household interview data. This comparison demonstrated a good match between the predicted trip length distribution and the household interview data. In particular, the percentage of short distance trips in the 0-5 km and 5-10 km bands now matched the observed distribution well.

It was clear from this result that additional distance correction terms were necessary to fully account for the low proportions of short distance trips in the RSI data. However the distance correction dummies are not ideal from a theoretical perspective, because the correction introduced to utility is not continuous with distance, but instead jumps suddenly at each distance interval. Therefore later model tests searched for distance correction parameters that were continuous with distance, so that the correction applied to utility is also continuous with distance.

At this stage in model development, the implied values-of-time were calculated for different model formulations. These values were lower than expected for employer's business travel, lying in the region of £ 0.76 to £ 3.50 (UK pounds) per hour, and varied considerably between different model formulations. These problems were in part due to the high correlation between the time and log-cost parameters in the model (-0.908), and large changes were observed in the values of both parameters, and hence the implied values-of-time, between different model formulations.

These problems were overcome by using a generalised time formulation, with car costs converted into generalised time units by assuming a fixed value of time from the 1998 Transport Economics Note (inflated to 2001). Both linear and logarithmic generalised time parameters were introduced into the models. The logarithmic parameter was more significant and gave a better fit to the data than a model formulation with a linear generalised time parameter alone.

Once the generalised time formulation had been introduced, the model development returned to the issue of introducing a continuous distance correction parameter to the model. A combination of linear (dist_lin) and quadratic (dist_quad) functions were used, defined only over the short distance interval as follows:

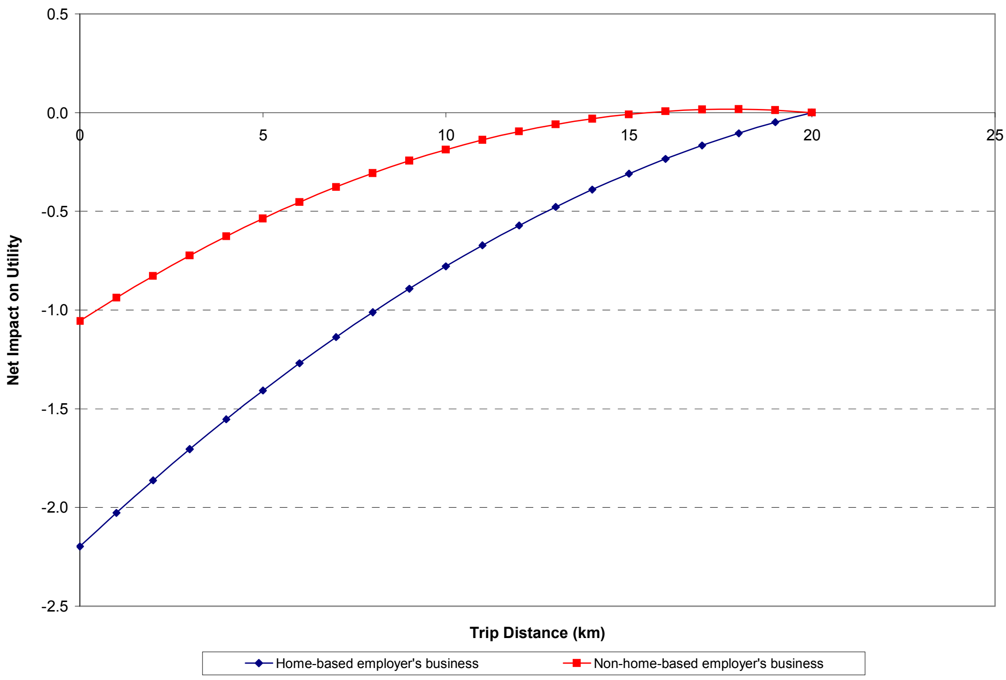
dist_lin	=	K – D	0 ≤ D ≤ K	(1)
dist_lin	=	0	D > K	
dist_quad	=	dist_lin * dist_lin		(2)

where: D is the distance measure from the highway level-of-service (D is never less than zero)
K is the upper limit for the distance correction function

Parameters were then estimated in the models which are multiplied by the linear and quadratic functions, so that the relative importance of each term was represented in the models. For the home-based employer's business model, values of K of 20 and 25 km were tested. For each value of K, the predicted trip length distribution was plotted to assess how well the distance correction parameters had overcome the short distance bias in the RSI data. The best model fit was obtained using correction parameters defined over the 0-20 km interval. For non-home-based employer's business, values of K of 20 and 25 km were also tested. Again the best model results were obtained using a value for K of 20 km.

The net impact of the two distance corrections on utility for both of these models are plotted in Figure 7.

Figure 7: Impact of Distance Correction Functions on Utility



In both cases the effect of the functions is to introduce a negative correction to 'utility' for shorter trips, reflecting the context of the RSI data collection, with this correction increasing as the trip length approaches zero. In application, the distance correction parameters are dropped and so the utility of short distance trips increases, and therefore more short distance trips are predicted.

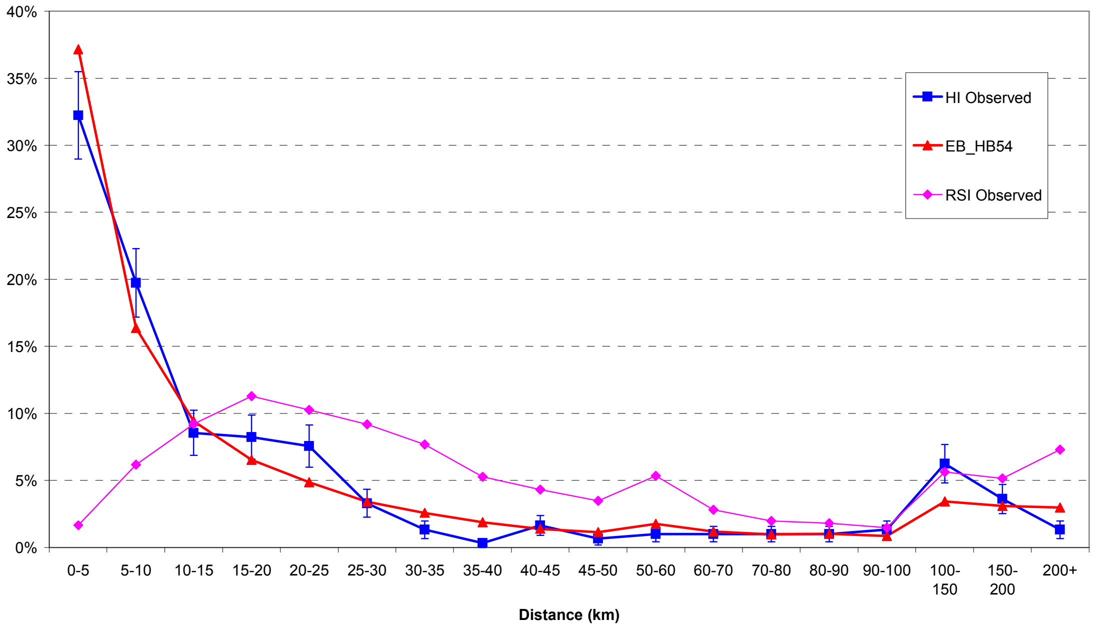
For the non-home-based other model, the mean trip length is shorter and so the continuous distance correction functions were tested over shorter intervals. Values for K of 10, 15 and 25 km were tested. The best match to the trip length distribution observed in the 2001 household data was obtained with a value for K of 15 km. When the distance correction functions were plotted for each value of K, they demonstrated large negative corrections to utility over the 0-5 km band, where trips are significantly under-sampled in the RSI data.

The next section presents validation of the final model formulation for each of the three purposes modelled.

3.4 Model Validation

The following figures compare the predicted trip length distributions from the final three final models to the observed distributions from the 2001 household interview data, and the observed distributions in the 2001 RSI data. Error bars are used for the 2001 household interview distribution due to the low volume of data. It is emphasised again that the screenline correction dummies and the continuous distance correction functions are only used in mode estimation, are therefore are dropped when the models are applied.

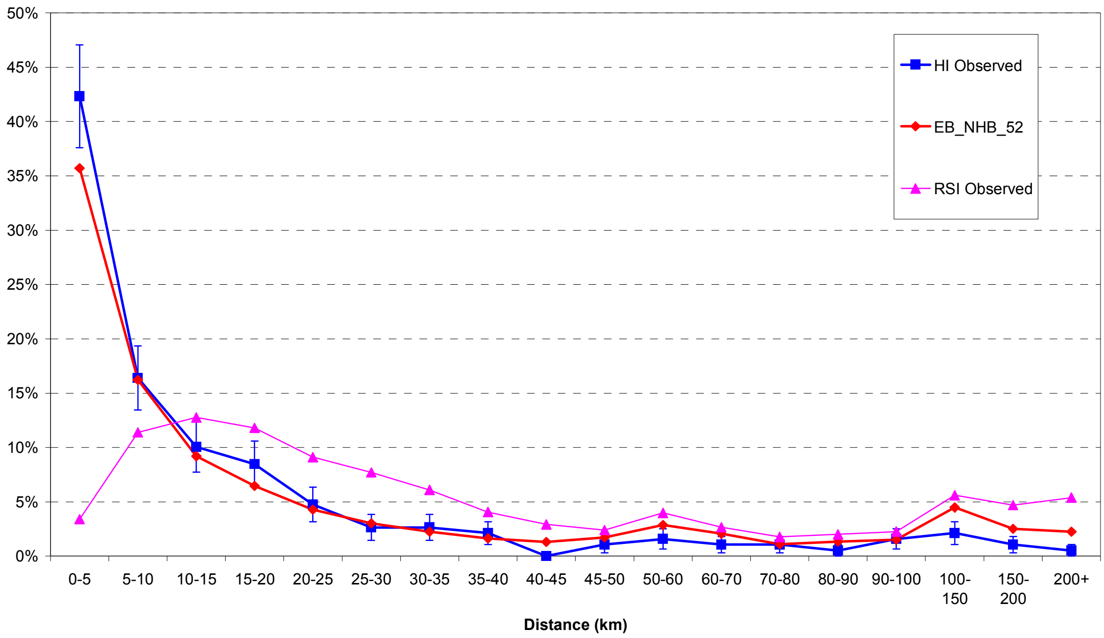
Figure 8: Validation of the Home-Based Employer's Business Model



It can be seen from Figure 8 that the model predictions match the observed distribution in the 2001 household interview data well. In particular, the low representation of short distance trips in the RSI data has been corrected.

The mean trip length predicted by the model is 30.4 km, which compares well to the mean observed value of 30.0 km from the 2001 household interview data.

Figure 9: Validation of Non-Home-Based Employer's Business Model



The under-representation of short-distance trips in the RSI data is not present in the model predictions, which match the observed distribution from the 2001 household interview data well. The model predicts more long distance trips (over 100 km) relative to the 2001 Household Interview data. This is highlighted by the comparison of mean trip distances predicted by the model given in Table 6.

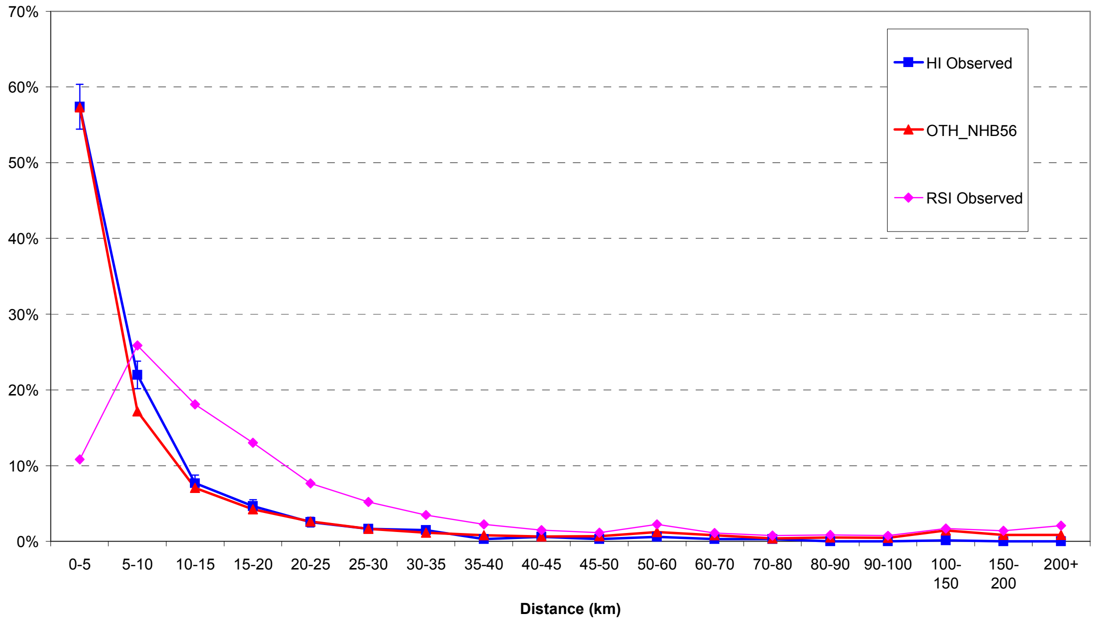
Table 6: Mean Trip Length Validation of NHB EB Model

	Average distance (km)	Average distance (only trips up to 100km)
Model	30.9	15.8
HI Observed	19.1	13.6

The mean trip distance predicted by the model is significantly longer than observed in the HI data. However, when only trips up to 100 km in length are included in the comparison, the match between predicted and observed trip distances is much better.

The over-prediction of long distance trips reflects the fact that the RSI sample contains persons from across Great Britain, whereas the 2001 household interview data only contains persons resident in the West Midlands. Trips are only included in the analysis if one end lies in the model area. In the RSI data, the NHB trip surveyed can form part of a chain of trips starting from a home anywhere in Great Britain. However in the household interview data, the NHB trip will form part of a chain of trips with a home in the model area, and for this reason the mean trip distance recorded will be shorter.

Figure 10: Validation of Non-Home-Based Other Model



The predicted trip length distribution matches the observed distribution from the 2001 Household Interview data well, and it can be seen that the under-representation of short distance trips has been successfully corrected.

The mean predicted trip distance for the model is 14.6 km. This is significantly longer than the value of 7.5 km observed in the 2001 household interview data. This difference is due to a lack of long distance trips in the 2001 household interview data, only one NHB trip greater than 80 km in length was observed in the 2001 household interview data. The RSI data will contain NHB trips made by persons from across Great Britain, and not just those resident in the West Midlands. For this reason the RSI data will contain a higher proportion of longer distance trips than the 2001 household interview data, for the reasons explained for the business model.

In summary, once distance correction terms had been added to the models in addition to the screenline correction parameters, the systematic difference between the data sources was successfully overcome, so that the predicted trip length distributions matched those observed in the household interview data well.

4. CONCLUSIONS

Modelling employer's business trips accurately is important because appraisal procedures attribute a high value of time to such trips. However such trips form a low proportion of the overall total, a problem frequently compounded by confusion with commuting and personal business travel. As a result, the volume of employer's business data available from household interview data may be insufficient to allow detailed choice models to be estimated.

In the West Midlands study, analysis demonstrated a very low employer's business share, which was around one-third of values in comparable data sources. This analysis also demonstrated a significant under-recording of non-home-based travel in the data. As a result this lack of data, it was necessary to use road-side interview data for model estimation. However, the choice-based nature of this data has a significant impact upon the trip length distribution and this needs to be explicitly corrected for when estimating choice models.

A number of approaches were reviewed to correct for the sampling procedure in the roadside interview data. The approach selected was to use screenline correction parameters, which provide a representation of the probability of interception to be incorporated into the model formulation. However comparison of model predictions to the reference household interview data demonstrated that these correction parameters did not fully overcome the low proportion of short distance trips in the RSI data. Additional distance correction functions were necessary to fully correct for the under-representation of short distance trips.

The approach adopted in the West Midlands work could be applied in other studies where choice-based data is to be used to estimate choice models. This study clearly demonstrates the need to fully understand the patterns of introduced by sampling procedures in the observed data, and to validate that the estimation procedure adopted fully corrects for these. It would be possible in future work, where the home interview contains more relevant information than in this study, to merge the data sources, allowing the volume of the roadside interview data to be used to obtain accuracy in the model estimates, while the home interview could provide information about socio-economic influences on travel.

Acknowledgements

The authors would like to thank Stephane Hess, from the Centre for Transport Studies at Imperial College London, for his comprehensive review of choice-based estimation techniques. This review formed a key input into the selection of the appropriate estimation approach in this work.

Bibliography

Cosslett, S. R. (1981) Efficient estimation of discrete choice models, in Manski, C.F. and McFadden, D. L. (eds.) *Structural Analysis of Discrete Data, with Econometric Applications*, MIT Press.

Gunn, H., M. Ben Akiva and L. Silman (1984) 'Disaggregate trip distribution models', *Proceedings of the Japanese Society of Civil Engineers*, no.347/IV 1.

Gunn, H, J. Fox and P. Mijjer (2001) Extended Tour Representations for Travel Demand Models, *European Transport Conference*, PTRC, Cambridge.