Potential Role of New Technologies in the National Travel Survey

Technical Annex:
Analysis of the accuracy of NTS respondents' estimates of trip distance and duration

by

Alexandra Cronberg,
National Centre for Social Research
&
Peter Bonsall,
University of Leeds

Research Project UG599

Executive summary

Introduction

This paper serves as a technical annex to the report http://www.dft.gov.uk/stellent/groups/dft_transstats/documents/page/dft_transstats_613849.pdf.

This paper reports on an analysis of National Travel Survey (NTS) data which was designed to establish:

- how accurately respondents had been able to estimate journey distances and durations;
- whether different types of respondents produce more accurate estimates than others; and
- whether the accuracy of respondents' estimates varies according to the type of journey

The analysis is based on comparisons of estimates of trip distance and duration attributed to the NTS respondents with various "objective" estimates of distance and duration derived using the reported trip end post codes. Three different routes formulations are explored:

- a Microsoft MapPoint calculation using the quickest-route-by-car;
- a Microsoft MapPoint calculation using the shortest-route-by-car; and
- crow-fly distances (geometric calculation using centroid of postcode zone).

Method

In order to provide an overview of the data, we first present and compare indicative statistics on the values of indicators derived from different sources. We then use linear regression models to explore the data in more detail. Our approach involved building models using data from 2002 and 2003 (combined) and then testing their robustness by comparing them with identical models built using the 2004 data. The prime purpose of our regression models is to try to explain the difference between a respondent's estimate and the MapPoint calculation based on the trip end postcodes. A second set of regression models was then constructed in order to explore the extent to which it is possible to explain (predict) respondent's estimates using the MapPoint calculation and other independent variables.

Results from preliminary inspection of the data

Comparison of the distribution of respondents' estimates with that of the MapPoint estimate (quickest road route) indicates an overall similarity. The means of the two distributions are similar (6.24 and 6.21 respectively) as are the standard deviations (16.75 and 16.36 respectively). However close inspection shows evidence of rounding in the estimates.

Comparison of the mean values of the distances estimated by respondents with the mean values calculated on the basis of the trip end post codes suggest, *inter alia*, that:

- overall, there is a remarkable match between the means of the respondents' estimate and of the MapPoint quickest route estimate;
- the match between respondents' estimate and MapPoint quickest route is better for car trips than for any other mode;
- the match with MapPoint quickest routes is less good for short trips.

- for walk trips, the crow-fly distance gives a better match than any of the road routes.
- for bike trips, the MapPoint shortest route gives the best match;
- the mean of the respondents' estimates are generally higher than those produced by MapPoint and, to an even greater extent, than the crow fly distance. The main exceptions to this rule are walk trips, bike trips, and short car trips (for each of which the mean respondents' estimate is lower than the MapPoint shortest route);
- the respondents' estimates for multi stage trips exceed the MapPoint estimates by an even greater degree, and this is even more marked for 3-stage trips than for 2-stage trips (an unsurprising result because multi stage trips are rarely direct).

Comparison of the mean values of the durations estimated by respondents with the mean values calculated on the basis of the trip end post codes suggest, *inter alia*, that:

- overall, there is a remarkable match between the means of the respondents' estimate and of the MapPoint shortest route estimate (for car trips, however, the MapPoint quickest route provides as close a match as the MapPoint shortest route and, for many car journeys – including those to rural destinations, trips by males or by people under 50, the MapPoint quickest route provides the best match);
- the match between means of the respondents' estimates and the MapPoint estimates seems to vary with the type of trip to a greater extent for durations than for distances;
- the match between the respondents' estimates and the MapPoint estimates is better for car trips than for any other mode;
- for car trips, the match with MapPoint shortest routes is less good for short trips;
- the means of the respondents estimates of duration generally fall between the two means produced by MapPoint (shortest route and quickest route). The exceptions to this rule are walk trips, bike trips and multi-stage trips (for each of which the mean respondents' estimate is considerably above either of the MapPoint estimates);
- among multi-stage trips, the respondents' estimates of duration exceed the MapPoint estimates by particularly large margin for trips involving a bus journey.

Overall, only just over a quarter of respondents' distance estimates fall within 10% of the MapPoint estimate for that trip, and about 45% and fall within 20% of it. The results for car trips are marginally better (at 33% and 53% respectively) while those for trips on foot are worse (at only 13% and 25% respectively).

Unsurprisingly, the overall result for duration estimates is substantially poorer - with 11% of duration estimates falling within 10% of the MapPoint estimate for that trip, and 23% falling within 20% of it. Again, the results for car trips are better (at 14% and 29% respectively) while those for trips by foot only achieve 3% and 6% respectively. The very poor result for walk trips is probably associated with the unsuitability of the MapPoint estimate for that type of trip.

Results from the regression model of difference

None of the models of difference were able to provide much explanation of the difference between the respondent's estimate and the MapPoint calculation for trip distance or duration (R square values were typically less that 0.1) – though the models for duration gave slightly better explanation than those for distance and the models for short trips tended to give slightly more explanation than those for long trips. The only independent variables to add significantly to the explanation were:

- the dummy variables relating to trip length - which suggested, unsurprisingly, that the proportionate difference decreased with trip length;
- the familiarity of the destination (deduced from the journey purpose using a categorisation to indicate whether the trip is likely to be made on a regular basis) – which suggested, again unsurprisingly, that the difference is reduced for familiar destinations;
- gender – with a slight tendency, in some of the distance models for the proportionate difference to be less when the respondent was female (only evident for car trips 0.86 miles or longer and for bike trips under 0.8 miles).

The coefficients which emerge when the best of the models built on the 2002/3 database are applied to the 2004 database are similar, but not identical, to those found in the 2002/3 database. We cannot therefore report that the explanations are absolutely constant over time but can report that they appear to be fairly stable.

Results from the regression models of respondent's estimates

Generally speaking, it proved possible to produce tolerably good explanations of the respondent's estimates using the MapPoint calculation along with variables describing the respondent and their journey.

The degree of fit between distances reported by respondents and distances calculated using the trip-end postcodes varies according to the characteristics of the traveller and the journey they are making.

The models providing the greatest level of explanation (r^2 in the region of 0.95) are those for those for distance of journeys by car and other motorised vehicles. Models of journey duration, and of distances for walk and bike journeys are rather less good (r^2 in the region of 0.3). The variables adding most explanation were the "objective" estimate derived from the trip end post codes, the MapPoint trip length quartile (reflecting the fact that the underlying model is only piecewise linear), the destination purpose (a categorisation to indicate whether the trip is likely to be made on a regular basis) and the respondent's gender. Of the various "objective" estimates, the MapPoint calculation for the quickest road route generally offered most explanation but the crow-fly route was more useful for some walk and bike trips.

Contents

1 INTRODUCTION

1.1 Background

This paper serves as a technical annex to the report *The potential role of new technologies in the National Travel Survey*.

This paper reports on an analysis of National Travel Survey (NTS) data which was designed to establish:

- how accurately respondents had been able to estimate journey distances and durations;
- whether different types of respondents produce more accurate estimates than others; and
- whether the accuracy of respondents' estimates varies according to the type of journey

1.2 Sources of respondents' distance estimates

NTS respondents complete a travel record for 7 days. For the purposes of this analysis we have used respondents' estimates of distance and duration for trips made on Day 7 because it is only on this day that they provide sufficiently detailed origin and destination information (full address including post code) to input into trip information software.

It is important to note that the trip distance and duration information contained in the NTS data files is not necessarily exactly as provided by respondents. This is because, before data are delivered to the Department for Transport, they are first checked by the original interviewer and then by the coders and editors in NatCen's Operations Department. Unlikely or missing values are verified using maps, gazeteers and route planning software before being entered into the data files. Hence the reported trip data are likely to be more similar to objective estimates than would be the case if the raw data were analysed. Unfortunately, the raw data is not available in electronic form because the core NTS contract requires it to be cleaned before delivery. An implication of this is that respondents' estimates of trip distance and duration may appear more accurate than is actually the case. Strictly speaking, therefore, our analysis relates not to respondents' estimates but to "respondents' estimates as delivered through the existing procedures". (A recent analysis revealed that, of 100 distance records, 9 had been revised by the interviewer, 6% by the coder and 23% by the editor allowing for multiple corrections. This suggests that between one quarter and one third of distance records may have been adjusted).

It should also be noted that our analyses necessarily reflect the fact that some travel diaries will have been completed as a result of collaboration between different household members (if not formal proxy-recording). In other words, because respondents within households may have collaborated in completing their travel records, the estimates of distance and duration may not have been provided independently by the traveller in question. In particular, there is likely to be an under-representation of estimates by children.

A final point to make is that our analyses are based on a sample of journeys rather than of people. This is because our inputs comprised information about trips and the accuracy with which they were reported, rather than information about individual respondents. Because some people travel less than others, the implication of this is that estimates by frequent travellers will be over-represented in the models (relative to their occurrence in the population). For the purposes of the current investigation, which focuses on the accuracy of data in the NTS database rather than on individual travellers' ability to estimate distances, we think it was correct to sample journeys rather than people. In fact, very few respondents travelled very frequently and so the difference may not be great anyway.

1.3 Sources of objective trip data

The true distances and durations are, of course, unknown (they could only be known if the "new technologies" had been used). Our proxy for the true distances and durations are "objective" calculations using available software (Microsoft MapPoint version 2004 (11.0)) working with the reported trip end postcodes. It should be noted that MapPoint does not take into account the time of day and the effect that variations in traffic levels may impact on journey times or the routes people may choose.

We calculated trip <u>distance</u> in three ways:

- using the Microsoft MapPoint calculation of journey distance using the quickest-route-by-car;
- using the Microsoft MapPoint calculation of journey distance using the shortest-route-by-car; and
- using crow-fly distances (geometric calculation using centroid of postcode zone).

The last of these is not regarded as a good estimate of the "real" trip distance because physical obstacles usually prevent people following it. Even so, it was thought that it might provide useful information in respect of short walk or cycle trips which might sometimes use more direct routes than would be available to cars.

We calculated trip <u>duration</u> in three ways:

- using the Microsoft MapPoint calculation of journey duration using the quickest-route-by-car;
- using the Microsoft MapPoint calculation of journey duration using the shortest-route-by-car; and
- using crow-fly distances (geometric calculation using centroid of postcode zone) and then applying a generic conversion factor based on average journey speeds.

It was not thought informative to pursue the last of these options since the crow-fly estimate is unlikely to be relevant for most trips. A possible exception would be some journeys by foot or by bike. However, we do not expect to gain much additional understanding from using a crow-fly duration calculation for walk and bike trips because the number of trips where the crow-calculation is the best estimate is likely to be small, especially as most such trips are made in urban areas which, by its very nature, includes many physical impediments.

The decision to explore distances and durations on shortest-route-by-car as well on quickest-routes-by-car was made in recognition of the fact that some drivers, and many users of other modes, would not necessarily use the route which happened to be the quickest route by car.

For most car journeys one might expect the "true" distance to vary around that of the quickest route; some drivers, mindful of distance related costs, will select a route which is somewhat shorter than the quickest, others, whether due to personal preference or ignorance, will select one which is longer than the quickest, but we would not expect many drivers to find a route that is shorter than that which MapPoint regards as the shortest. For public transport modes, the true distance will generally be somewhat longer than the shortest route and is quite likely to be longer than the quickest route. For journeys on foot or by bike the true distance is likely to be close to the shortest route (by car) but may sometimes be longer and sometimes shorter than this.

In terms of true trip duration, the MapPoint calculations are less likely to provide good estimates than they are for true distances. This is primarily because the MapPoint calculations of duration are based on free-flow traffic conditions, and do not allow for variation according to traffic conditions at different times of day. Also, the MapPoint calculations of journey duration are likely to be even less applicable to non-car modes than were the MapPoint calculations of distance (because of the very different speeds typically achieved by different modes[1]).

[1] *pace* the argument championed by Martin Mogridge and others that journey durations by different modes used between a given origin and destination will tend to equal one another, in most circumstances walk trips will take much longer than the equivalent trip by car.

2 METHOD

2.1 Outline

In order to answer the questions posed, we first present and compare indicative statistics on the distributions of the respondents' estimates and the "objective" calculations (Chapter 3). We then use linear regression models to explore the data in more detail (Chapter 4).

The indicative statistics provide an accessible overview of relationships in the data and allow some immediate conclusions to be drawn. The trip length distributions are shown and tabulations are provided to allow comparison of the mean trip lengths and durations estimated by respondents with the mean values calculated from the trip end post codes. Tabulations are also provided to demonstrate what proportion of respondents' estimates lie close to the value calculated on the basis of the trip end post codes reported for that trip.

The modelling work is intended firstly to see whether the differences between respondents' estimates and the post-code-based calculations can be explained and secondly to see whether respondents' estimates can be predicted using the post-code-based calculations and other information relating to the respondent and their journey. Our approach involved building models using data from 2002 and 2003 (combined) and then testing their robustness by comparing them with identical models built using the 2004 data.

The general form of the models is:

$$Y = A0 + (X1*A1) + (X2*A2) + (X3*A3) + \dots (Xn*An) + (X1*X2*A4) + (X1*X3*A5) + (Xn*Xn1*An)$$

where *Y* is the dependent variable
Xn is a characteristic of the journey or respondent (independent variables)
An is a calculated coefficient

We began by including all relevant independent variables and all binary combinations of independent variables. We then removed those which made no significant contribution to the result.

We have built separate regression models for each major mode of transport, i.e. car, walk, bike, and "other motorised vehicles" (the latter mainly comprising motorcycle, taxi, and goods vehicle trips).

It would have been convenient if we could have built one, linear, model for each mode. However, it is clear from Figures 1 - 4 that the relationship between trip distance and proportionate difference is not linear. It is evident that the proportionate difference between the respondents' estimate and the MapPoint mileage decreases sharply at around 0.8 miles[2] regardless of mode of transport (though the critical point is slightly less clear in the case of bike trips). This value approximates to 25th percentile of the MapPoint mileages (0.86 miles). Rather than seek to develop a range of non-linear models, we deduce from Figures 1-4 that we are justified in building two linear models – one for trips whose length fall in the first quartile and one for those whose length exceed that distance.

[2] We interpret this non-linearity for short trips as a being a consequence of the well-known tendency of respondents to provide rounded estimates of distance and duration - which is bound to create a problem when dealing with proportionate differences on very short trips.

Figure 1. Relationship between the proportionate difference and MapPoint mileage for journeys by car (2002-03)

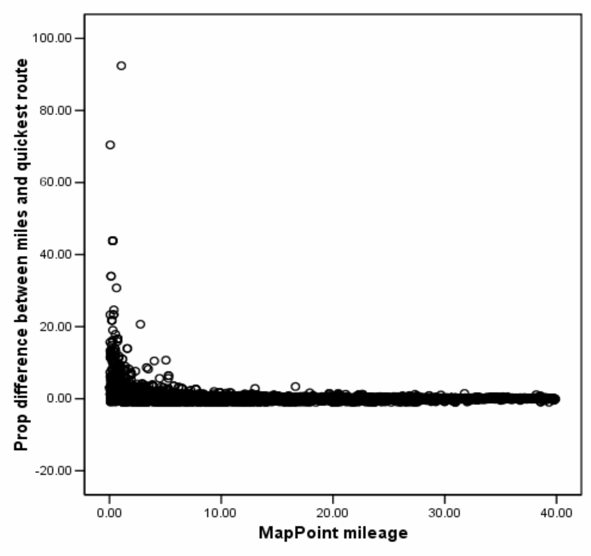


Figure 2. Relationship between the proportionate difference and MapPoint mileage for journeys by bike (2002-03)

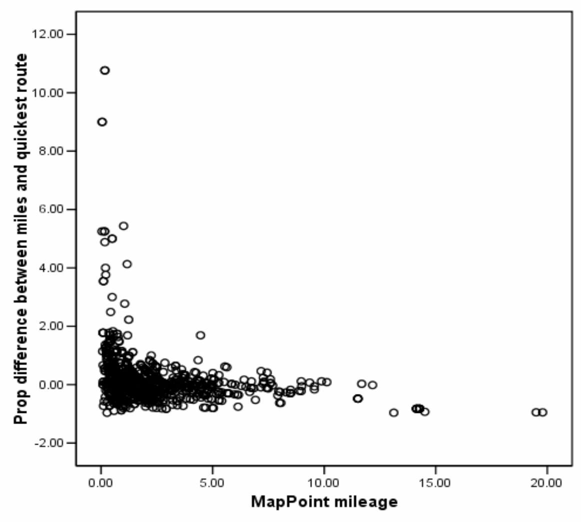


Figure 3. Relationship between the proportionate difference and MapPoint mileage for journeys by foot (2002-03)

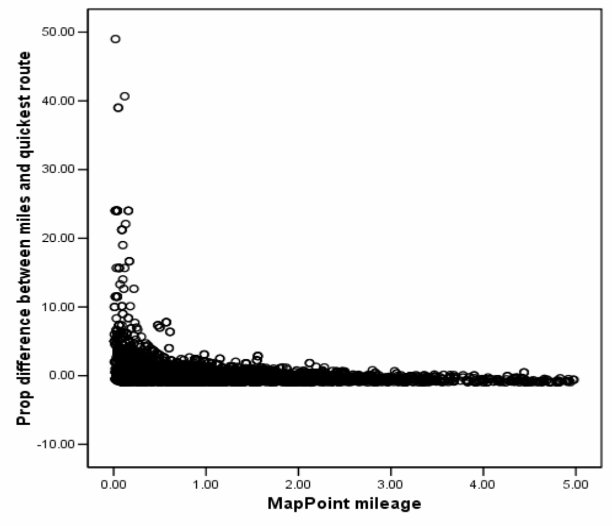
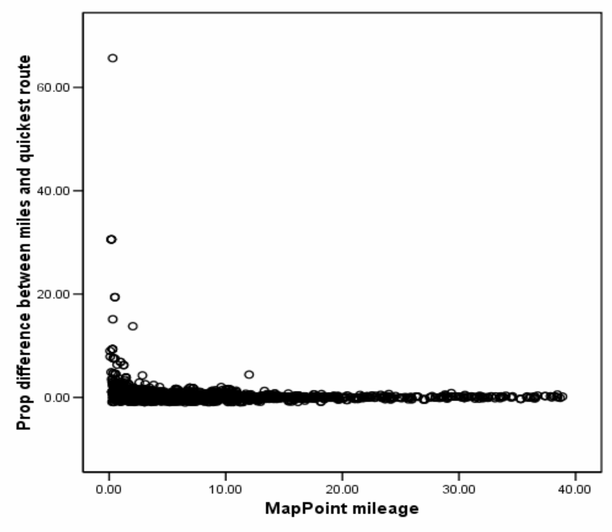


Figure 4. Relationship between the proportionate difference and MapPoint mileage for journeys by other motorised vehicles (2002-03)



2.2 Dependent variables

2.2.1 Models of Difference

The prime purpose of our regression models is to try to explain the difference between a respondent's estimate and the MapPoint calculation based on the origin and destination post codes. The difference is, in modelling terms, considered as an "error" but since we cannot be sure whether the error, if any, is associated with the respondent's estimate or the MapPoint calculation, it is perhaps more appropriate to consider it simply as a difference.

A difference can be expressed in various ways, notably the ratio (a/b), the proportionate difference ((a-b)/b) and the absolute difference (a-b). Our general preference has been to explore the proportionate difference in our modelling work because it provides the most intuitive indication of overall accuracy (Note, however, that, for simplicity, the indicative statistics presented in chapter 3 are based on the ratio).

2.2.2 Models of the Respondent's Estimate

A second set of regression models was constructed in order to explore the extent to which it is possible to explain (predict) respondent's estimates using the MapPoint calculation and other independent variables. This investigation is expected to throw light on the value of asking the respondent for estimates of time and duration in addition to asking for the post codes.

2.3 Independent variables

We examined the following variables:

- **MapPoint estimate** of trip length
  - on quickest route by car
  - on shortest route by car
  - assuming crow-fly route

- **MapPoint estimate** of trip duration
  - on quickest route by car
  - on shortest route by car

- respondent's **gender** (*female*=1 if female, otherwise 0)
- respondent's **age** (10 year age bands)
- respondent's **economic status** (working full-time, working part-time, unemployed, economically inactive)
- **trip length category**[3] (*aboveav*=1 if MapPoint quickest route length is above average (6.23 miles[4]), otherwise 0)
- **trip length category1** (based on the MapPoint quickest route calculations using *quartiles* instead of the mean)
  Q1=1 if length is less than 25th percentile (0.86 miles) otherwise 0
  Q2= 1, if length is between 25th and 50th percentile (2.19 miles) otherwise 0
  Q3=1 if length is between 50th and 75th percentile (5.67 miles) otherwise 0
  Q4=1 if length exceeds 75th percentile (5.67 miles) otherwise 0
- **trip length category3** (*long*=1 if MapPoint quickest route length is above median (2.17 miles), otherwise 0)
- **trip purpose** (*familiar*=1 if the trip can be considered a familiar journey (between home and work, education, or grocery shopping), otherwise 0)

[3] The variables 'trip length category', 'trip length category1', and 'trip length category3' were tested separately in each model.
[4] Based on the average trip length (based on MapPoint quickest-routes-by-car) for trips in our 2002/3 database.

- **urban vs. rural destination** (_rural_ =1 if rural destination, otherwise 0)
- **number of people in the party**[5]
- dummy variable for **car passenger status** (_pass_=1 if passenger, otherwise 0)

2.4 Sample sizes (2002-03)

- Number of day 7 records: **110,563**
- Number of day 7 records having a complete, codeable and apparently accurate postcodes (excluding postcodes that were only proximate to the street level or the local area): **97,379** origins, **97,352** destinations, yielding **85,514** with origin and destination.
- Number of day 7 records having a complete, codeable and apparently accurate postcodes for a single-stage journey: 74,450
- Number of day 7 records having a complete, codeable and apparently accurate postcodes for a single-stage journey by motorised vehicle, walk or cycle **73,977** (i.e. excluding journeys made by train, underground, and light rail).
- Size of sub samples:
  - trips by: car driver =31,811
    car passenger =16,941
    other motorised mode (single stage) = 5584 (of which 2667 are by bus or coach - although such trips will usually involve more than one stage, stages of less than 50 yards are omitted)
    2 stage trips = 5,393
    3 stage trips = 4,149
    walk trips = 19,282,
    bike trips =1,177
  - familiar journeys =40,233, other 28,580
  - female = 39,395, male = 34,582
  - rural destination =11,830, urban destination = 62,147
  - trips below average length (again based on average of 6.23 miles calculated for the 2002/3 database using MapPoint quickest routes) = 54,722, trips above average length = 16,327
  -

2.5 Data from the 2004 dataset

The data for 2004 exhibited similar characteristics to the data for 2002-03 in most respects. The proportions of males and females, trips in rural and urban areas, and for each mode of transport were all similar to the 2002-03 data set. The mean trip length was also similar, even if there was an improvement in the capture of walking stages in the 2004 survey. However, this change did not have a substantial impact on the overall average trip length.

[5] This variable was investigated to test whether, for example, having additional people in the trip would produce more accurate estimates – perhaps utilising more opinions; in fact this variable had no significant impact in any of the models

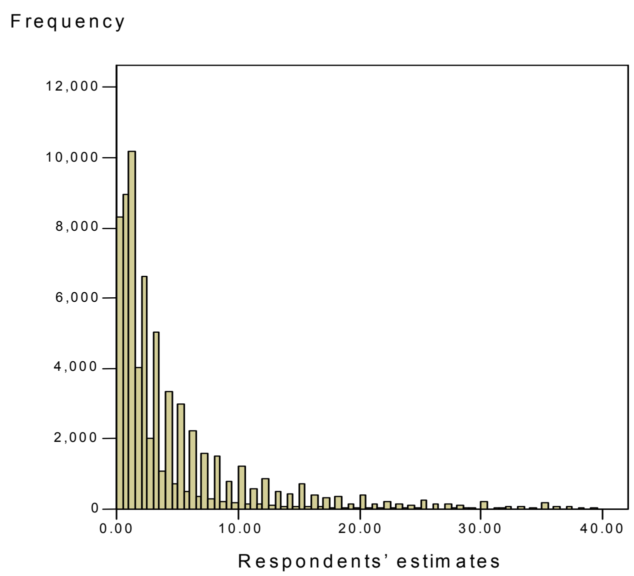
3 RESULTS OF SIMPLE ANALYSIS

Before building complex models (see Chapter 4), we present some simple figures and tables designed to reveal any obvious differences between respondents' estimates of trip distance or duration and the calculated values for these. We present, in turn, the trip length distributions, tabulations relating to mean trip lengths, tabulations relating to mean trip durations and, finally tabulations relating to the closeness of the respondent's estimate to the MapPoint estimate for that same journey.

3.1 Trip Length Distributions

The distribution of journey lengths based on respondents' estimates and on the MapPoint mileage for the quickest and shortest routes can be seen in Figures 5, 6 and 7 respectively. It should be noted that journeys over 40 miles have been excluded in order to allow for a greater level of detail in the diagrams. However, more than 96% of all journeys fell under 40 miles, implying that the results in the diagrams are not unduly affected by this truncation. The diagrams indicate that the distribution of trip distance is fairly similar for all three estimates (i.e. respondent's estimates, MapPoint quickest and shortest route calculations). Yet it is also evident that the variance of trip distance is rather large; the mean distance of MapPoint quickest route calculation (including all eligible journeys) is 6.21 miles, with a standard deviation of 16.36 miles. For journeys calculated on the MapPoint shortest distance the mean distance is 5.61 miles and the standard deviation is 14.86 miles. The equivalent values for journeys based on respondents' estimates is a mean journey length of 6.24 miles with a standard deviation of 16.75 miles.

Figure 5. Distribution of respondents' estimates of journey distance (All eligible journeys, 2002-03)



It is evident that Figure 5 lacks the smoothness of Figures 6 and 7, particularly at short distances; this is almost certainly due to rounding of respondents distance estimates.

Figure 6. Distribution of journey distances, based on MapPoint calculation of quickest route (All eligible journeys, 2002-03)

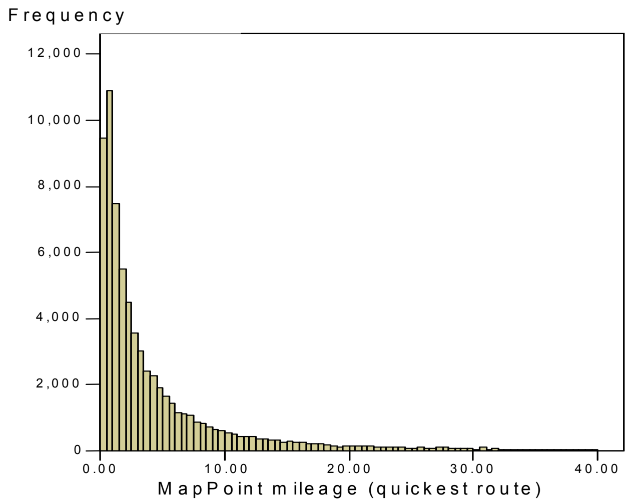
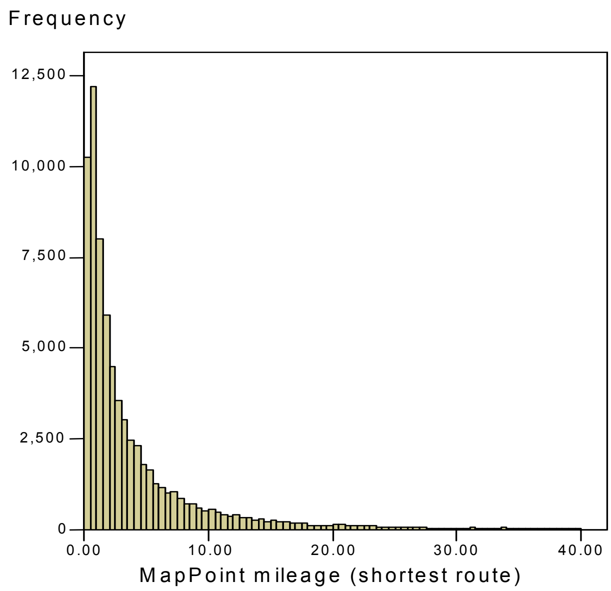


Figure 7. Distribution of journey distances, based on MapPoint calculation of shortest route (All eligible journeys , 2002-03)



3.2 Tabulation of mean trip lengths

Table 3 shows the means of the trip distances estimated by respondents, by MapPoint (quickest route), MapPoint (shortest route) and crow-fly route and the differences between the means of the respondents' estimate and each of the objectively calculated estimates. The table, which draws on data from the 2002 and 2003 NTS data sets, includes results for the main modes and for the most common combinations of modes (combinations such as car-train-tube/bus and walk-train-bus have been excluded from the analysis due to small sample sizes (N<100)). It should be noticed that the order of the modes of transport has not been taken into account.

Separate tabulations were also examined for journeys whose MapPoint calculated distance (on quickest road routes) fell into each of the four distance quartiles (the results of this more detailed analysis are reported here even though the tables are omitted in the interests of avoiding information overload!).

Although the results in Table 3 relate only to mean values and may suffer from unwanted correlation, we are able to draw the following immediate conclusions about the distance estimates:

- overall, there is a remarkable match between the means of the respondents' estimate and of the MapPoint quickest route estimate;
- the match between respondents' estimate and MapPoint quickest route is better for car trips than for any other mode;
- the match with MapPoint quickest routes is less good for short trips;
- for walk trips, the crow-fly distance gives a better match than any of the road routes;
- for bike trips, the MapPoint shortest route gives the best match;
- car passengers' estimates of distance tend to exceed those of car drivers;
- the mean of the respondents' estimates are generally higher than those produced by MapPoint and, to an even greater extent, than the crow fly distance. The main exceptions to this rule are walk trips, bike trips, and short car trips (for each of which the mean respondents' estimate is lower than the MapPoint shortest route);
- the respondents' estimates for multi stage trips exceed the MapPoint estimates by an even greater degree, and this is even more marked for 3-stage trips than for 2-stage trips (an unsurprising result because multi stage trips are rarely direct).

3.3 Tabulation of trip durations

Table 4 shows indicative statistics for trip duration equivalent to those shown in Table 3 for trip distance.

As with Table 3, although these results relate only to mean values and may suffer from unwanted correlation, we are able to draw some immediate conclusions about the estimates. These are:

- overall, there is a remarkable match between the means of the respondents' estimate and of the MapPoint shortest route estimate (for car trips, however, the MapPoint quickest route provides as close a match as the MapPoint shortest route and, for many car journeys – including those to rural destinations, trips by males or by people under 50, the MapPoint quickest route provides the best match);

Table 3: Summary statistics relating to lengths of trips in the 2002/3 data set

	Mean value (miles)				Absolute difference			Ratio		
	1. Respondent's estimate	2. MapPoint time (quickest)	3. MapPoint time (shortest)	4. Crowfly distance	1-2	1-3	1-4	1/2	1/3	1/4
All single stage trips	6.23	6.24	5.62	4.49	-0.01	0.61	1.74	1.00	1.11	1.39
Trips by car	8.31	8.22	7.4	5.96	0.09	0.91	2.35	1.01	1.12	1.39
Trips by bike	1.96	2.32	2.11	1.54	-0.36	-0.15	0.42	0.84	0.93	1.27
Trips by foot	0.65	1	0.91	0.6	-0.35	-0.26	0.05	0.65	0.71	1.08
Other motorised[6]	7.26	6.93	6.23	4.99	0.33	1.03	2.27	1.05	1.17	1.45
Trips as car driver	8.13	8.03	7.21	5.8	0.1	0.92	2.33	1.01	1.13	1.40
Trips as car passenger	8.65	8.59	7.76	6.27	0.06	0.89	2.38	1.01	1.11	1.38
Car trips to familiar destinations	9.56	9.41	8.48	6.89	0.15	1.08	2.67	1.02	1.13	1.39
Car trips to rural destinations	11.15	11.38	10.23	6.24	-0.23	0.92	4.91	0.98	1.09	1.79
Car trips of under 3 miles	1.73	2.15	1.99	1.43	-0.42	-0.26	0.3	0.80	0.87	1.21
Car trips by males	9.35	9.25	8.31	6.73	0.1	1.04	2.62	1.01	1.13	1.39
Car trips by females	7.37	7.3	6.59	5.28	0.07	0.78	2.09	1.01	1.12	1.40
Car trips by people under 30	7.55	7.46	6.73	5.39	0.09	0.82	2.16	1.01	1.12	1.40
Car trips by people 30-50	8.9	8.88	7.98	6.45	0.02	0.92	2.45	1.00	1.12	1.38
Car trips by people over 50	8.36	8.21	7.4	5.96	0.15	0.96	2.4	1.02	1.13	1.40
Car trips by employed people	9.05	8.96	8.05	6.5	0.09	1	2.55	1.01	1.12	1.39
All 2 stage trips	11.64	11.24	10.25	8.47	0.4	1.39	3.17	1.04	1.14	1.37
Walk + bus	4.77	4.78	4.27	3.33	-0.01	0.50	1.44	1.00	1.12	1.43
Walk + car	9.76	9.20	8.42	6.94	0.56	1.34	2.82	1.06	1.16	1.41
Walk + tube	7.77	9.35	8.56	7.32	-1.58	-0.79	0.45	0.83	0.91	1.06
Walk + train	14.98	15.41	14.14	11.62	-0.43	0.84	3.36	0.97	1.06	1.29
Car + bus	11.55	9.98	9.13	7.55	1.57	2.42	4.00	1.16	1.27	1.53
Car + train	34.40	30.82	27.90	23.56	3.58	6.50	10.84	1.12	1.23	1.46
All 3 stage trips	19.77	18.98	17.31	14.63	0.79	2.46	5.14	1.04	1.14	1.35
Walk + car + walk	6.12	5.88	5.07	4.14	0.24	1.05	1.98	1.04	1.21	1.48
Walk + bus + walk	4.22	3.9	3.56	2.78	0.32	0.66	1.44	1.08	1.19	1.52
Walk + train + walk	18.52	17.81	16.41	14.09	0.71	2.11	4.43	1.04	1.13	1.31
Walk + tube + walk	7.11	7.53	6.70	5.59	-0.42	0.41	1.52	0.94	1.06	1.27
Car + train + walk	38.04	36.98	33.71	29.3	1.06	4.33	8.74	1.03	1.13	1.30
Car + bus + walk	10.53	10.35	9.51	7.91	0.18	1.02	2.62	1.02	1.11	1.33
Car + tube + walk	12.61	11.33	10.16	8.65	1.28	2.45	3.96	1.11	1.24	1.46

[6] Single stage by PTW, bus, Coach, lorry, taxi etc – note that the inclusion of some bus trips as "single stage" is perhaps anomalous

Table 4: Summary statistics relating to duration of trips in the 2002/3 data set

	Mean value (minutes)			**Absolute difference**		**Ratio**	
	1. Respondent's estimate	2. MapPoint time (quickest)	3. MapPoint time (shortest)	1-2	1-3	1/2	1/3
All single stage trips	19.16	13.65	19.25	5.51	-0.09	1.40	1.00
Trips by car	19.8	16.94	24.25	2.86	-4.45	1.17	0.82
Trips by bike	15.94	7.93	10.33	8.01	5.61	2.01	1.54
Trips by foot	14.87	4.59	5.65	10.28	9.22	3.24	2.63
Bus and coach trips	27.52	15.71	21.97	11.81	5.55	1.75	1.25
Trips as car driver	19.48	16.73	24.08	2.75	-4.6	1.16	0.81
Trips as car passenger	20.39	17.31	24.53	3.08	-4.14	1.18	0.83
Car trips to familiar destinations	21.27	18.37	26.44	2.9	-5.17	1.16	0.80
Car trips to rural destinations	22.59	21.69	30.76	0.9	-8.17	1.04	0.73
Car trips of under 3 miles	7.39	7.34	9.07	0.05	-1.68	1.01	0.81
Car trips by males	21.14	18.43	26.56	2.71	-5.42	1.15	0.80
Car trips by females	18.6	15.59	22.17	3.01	-3.57	1.19	0.84
Car trips by people under 30	18.66	15.88	22.47	2.78	-3.81	1.18	0.83
Car trips by people 30-50	20.25	17.87	25.61	2.38	-5.36	1.13	0.79
Car trips by people over 50	20.41	16.87	24.34	3.54	-3.93	1.21	0.84
Car trips by employed people	20.58	18.03	26.06	2.55	-5.48	1.14	0.79
All 2 stage trips	44.02	21.00	30.78	23.02	13.24	2.10	1.43
Walk + bus	39.16	12.69	17.96	26.47	21.2	3.09	2.18
Walk + car	33.04	18.04	25.77	15.00	7.27	1.83	1.28
Walk + tube	51.37	22.48	33.97	28.89	17.4	2.29	1.51
Walk + train	54.08	29.71	41.81	24.37	12.27	1.82	1.29
Car + bus	49.58	21.29	31.19	28.29	18.39	2.33	1.59
Car + train	73.02	48.12	73.55	24.90	-0.53	1.52	0.99
All 3 stage trips	64.16	32.48	47.35	31.68	16.81	1.98	1.36
Walk + car + walk	32.16	12.81	17.82	19.35	14.34	2.51	1.80
Walk + bus + walk	39.48	11.16	15.73	28.32	23.75	3.54	2.51
Walk + train + walk	65.16	33.42	48.75	31.74	16.41	1.95	1.34
Walk + tube + walk	49.8	18.79	30.01	31.01	19.79	2.65	1.66
Car + train + walk	87.88	55.32	85.70	32.56	2.18	1.59	1.03
Car + bus + walk	67.09	23.07	32.70	44.02	34.39	2.91	2.05
Car + tube + walk	78.56	26.64	42.70	51.92	35.86	2.95	1.84

- the match between means of the respondents' estimates and the MapPoint estimates seems to vary with the type of trip to a greater extent for durations than for distances (compare tables 3 and 4);
- the match between the respondents' estimates and the MapPoint estimates is better for car trips than for any other mode;
- for car trips, the match with MapPoint shortest routes is less good for short trips
- car passengers' estimates of trip durations tend to exceed those of car drivers;
- the means of the respondents estimates of duration generally fall between the two means produced by MapPoint (shortest route and quickest route). The exceptions to this rule are walk trips, bike trips and multi-stage trips (for each of which the mean respondents' estimate is considerably above either of the MapPoint estimates);
- among multi-stage trips, the respondents' estimates of duration exceed the MapPoint estimates by particularly large margin for trips involving a bus journey.

3.4 Tabulation of closeness of respondents estimate to MapPoint estimate

Tables 3 and 4 suggested that the mean value of the respondents' estimate was generally quite close to that of one or more of the MapPoint calculations (usually the quickest route). But, as is clear from Table 5, this does not mean that, for a given journey, the respondent's estimate is necessarily close to a MapPoint calculation of the same journey.

Table 5 shows the proportion of trips for which the respondent's estimate is within 10% and 20% of the estimate predicted for that trip by MapPoint. It is evident that, overall, slightly more than a quarter (27%) of distance estimates fall within 10% of the MapPoint estimate for that trip, and 45% within 20% of it. The results for car trips are better however (at 33% and 53% respectively) while those for trips on foot are worse (at only 13% and 25% respectively).

Unsurprisingly, the overall result for duration estimates is substantially poorer - with 11% of duration estimates falling within 10% of the MapPoint estimate for that trip, and 23% falling within 20% of it. Again, the results for car trips are better (at 14% and 29% respectively) while those for trips by foot only achieve 3% and 6% respectively. The poor result for walk trips is probably associated with the unsuitability of the MapPoint estimate for that type of trip.

Table 5: Proportion of trips where the respondent's estimate is within 10% and 20% of the estimate predicted for that trip by MapPoint (quickest route except for bike and foot where MP shortest route was used)

	Distance		Duration	
	Within 10% of MP estimate	**Within 20% of MP estimate**	**Within 10% of MP estimate**	**Within 20% of MP estimate**
All trips	26.8	44.9	11.1	22.7
Trips by car	32.5	53.2	14.3	29.3
Trips by bike	18.3	33.3	10.7	20.2
Trips by foot	12.8	24.6	2.9	5.9
Other motorised	24.3	42.1	10.4	21.7
Trips as car driver	33.8	54.7	14.2	29.2
Trips as car passenger	30.3	50.3	14.5	29.6
Car trips to familiar destinations	34.4	55.3	14.5	29.5
Car trips to rural destinations	38.1	59.7	12.5	26.5
Car trips by males	33.5	54.5	14.0	28.8
Car trips by females	31.6	51.9	14.5	29.8
Car trips by people under 30	29.3	49.4	14.1	29.1
Car trips by people 30-50	33.5	54.6	13.7	28.9
Car trips by people over 50	34.1	54.7	15.1	30.0
Car trips by employed people	34.2	55.3	14.0	29.0
Car trips by people without paid employment	30.8	51.0	15.1	30.1

4 RESULTS FROM THE REGRESSION MODELS OF DIFFERENCES

4.1 Outline

As noted in Chapter 2, the dependent variable is the proportionate difference between the respondent's estimate and the MapPoint calculation.

In our initial investigations we tested some all-mode models but since mode was a highly significant variable in these models, and since they also contained significant interaction variables linking mode and other attributes, we decided to treat each mode as a separate model. Separate models are therefore presented for each of the following modes of transport: car, other motorised, walk and cycle (the analysis of indicative statistics suggested that there would be little point in attempting to build models for multistage trips and our initial modelling work suggested that disaggregation of "other motorised modes" into the constituent modes - motorcycle, lorry, taxi, etc - could not be supported by the data).

Separate models are presented, for each mode, for trips below 0.86 miles in length and for trips in excess of this length.

Although we have tested many different models (something in excess of 50), we will, for reasons of clarity and simplicity, present only the best-fitting model for each mode of transport and trip length.

We have run models using MapPoint calculations for the quickest, shortest, and crow-fly routes. The models presented below use the MapPoint calculation which produced the best (highest explanatory power) model for the dependent variable in question. Generally speaking (although perhaps surprisingly), the quickest route calculation produced the best models for trip distances while the shortest route calculations produced the best models for trip durations. The proportionate difference models included in the analysis are thus: the quickest route for both long and short journeys by car, other motorised vehicles and bike. The crow-fly route gave the best models for walk trip distances.

The MapPoint calculation of distance (or duration) was allowed to enter the models as an independent variable but in no case was it significant. This is probably because the effect of distance was represented more effectively through the trip length quartile dummy variables.

As stated in the Introduction, the analysis of multi-stage journeys in this context are associated with a number of problems. In particular, we cannot be certain that the door-to-door distance of journeys with intermediary destinations reflect a 'logical' routing, or to what extent respondents' estimates of access stages are subject to rounding errors. Bearing these limitations in mind, we have not explored the accuracy of respondent's estimates compared with MapPoint calculations with regards to two- and three-stage journeys.

4.2 Models of trip distance

Table 6 shows the summarised results from the regression models of trip distance, which are discussed in greater detail below.

Table 6: Summarised results of distance models

	Short trips less than 0.86 miles			**Long trips 0.86 miles +**		
Trip type	**Predicting proportionate difference between respondent's estimates and MapPoint quickest route (crow-fly route for walk)**			**Predicting proportionate difference between respondent's estimates and MapPoint quickest route (crow-fly route for walk)**		
	Indep. variables accepted	**R2**		**Indep. variables accepted**	**R2**	
Car	None	0.053	Mainly random/unobs. Error falls with length of journey	Female, familiar, 2nd quartile, 3rd quartile	0.012	Mainly random/unobs. Weak effect of female and familiar – variables not sig. for 2004 data
Other motorised	None	0.137	Mainly random/unobs. Error falls with length of journey	2nd quartile	0.018	Mainly random/unobs.
Walk	Familiar	0.070	Mainly random/unobs. Lower error for familiar journeys	Interaction distance/ familiar	0.095	Mainly random/unobs.
Bike	Female	0.174	Mainly random/unobs. Lower error with length and for females	None	0.043	Mainly random/unobs.

Car

Short Journeys

No independent variables were accepted in the model of short journeys by car based on the quickest route (see Appendix I. A1) ($p<0.05$). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.053, $F<0.01$) and most of the proportionate difference is random or associated with unobserved variables. The error is predicted using only the MapPoint mileage, and the model suggests that the error decreases with each additional MapPoint mile.

The error for a typical journey of 0.5 miles can be calculated as:

2.57 - (3.077 * 0.5) = *1.032*

In other words, the model predicts that for a typical journey of half a mile, the respondent's estimate is on average 1.032 miles above the actual distance travelled, i.e. more than three times the trip distance.

Long Journeys

The model of long journeys by car (based on quickest route) (see Appendix I. A3) includes all the statistically significant ($p<0.05$) variables (*familiar journey*, *female, in 2nd Quartile, in 3rd Quartile*). This model does not provide a satisfactory explanation of the dependent variable either (R square = 0.012, $F<0.01$). The model suggests that the proportionate error depends on the quartile in which the trip length falls, and not on the MapPoint mileage on its own. The error is thus likely to be larger for journeys in the 2nd quartile compared with journeys in the 3rd or 4th quartile, and it is somewhat larger for journeys in the 3rd quartile compared with journeys in the 4th quartile. The relatively large predicted errors for journeys in the 2nd quartile may however be due to rounding errors.

The model also predicts that journeys with a familiar origin and destination yield a marginally lower proportionate error than journeys with an unfamiliar origin or destination, controlling for journey distance. Similarly, the proportionate error is slightly lower for females than for males, holding trip length constant.

The error for a typical journey (based on the median distance for car journeys) for males with unfamiliar destination / origin can be calculated as:

0.012 + (0.0 * 4.15) + 0.044 = *0.056*

Application of models on 2004 data

Models with the same specifications as in Appendix I. A1 and A3 were tested on the 2004 data (see Appendix I. A2 and A4). The model for short journeys yielded similar results, even if the explanatory power was slightly higher (R square = 0.077). The coefficient of the MapPoint mileage was of similar magnitude and sign. The model for long car journeys did not fit as well on the 2004 data. In particular the variables *female* and *familiar* were no longer significant. Given the weak effects of these variables in the model based on 2002-03 data however, it is likely that their statistical significance is simply the result of the large sample size.

Other motorised modes (single stage trips only)

Short Journeys

No independent variables were accepted in the model of short journeys by other motorised vehicles, based on quickest route, of the 2002-03 data (see Appendix I. B1) ($p<0.05$). Once more, the error is predicted using only the MapPoint mileage, and is expected to decrease with journey length. This model provides a somewhat better explanation of the dependent variable (R square = 0.137, F<0.01), even if it is still poor.

The error for a typical journey of 0.5 miles can be calculated as:

8.485 - (10.918 * 0.5) = *3.026*

Long Journeys

The final model of long journeys by other motorised vehicles, based on quickest route, (see Appendix I. B3) of the 2002-03 data includes all the statistically significant ($p<0.01$) variables (*in 2nd Quartile*). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.018, F<0.01); most of the proportionate difference is random or associated with unobserved variables. This model also predicts that journeys in the 2nd quartile have larger proportionate errors compared with journeys in the 3rd or 4th quartiles, but that there is no difference in the relative errors between the two upper quartiles.

The error for a typical journey (based on the median distance for other motorised journeys) can be calculated as:

0.026 + (0.0 * 3.64) = *0.026*

Application of models to 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. B1 and B3 were tested on the 2004 data (see Appendix I. B2 and B4). The model for long journeys had similar explanatory power (R square = 0.019), whereas the model for short journeys had a substantially better fit (R square = 0.384). The beta coefficients were of similar magnitude and signs, which the exception of MapPoint mileage in the model for long journeys, which turned negative, even if it was still very close to zero. Given the statistical non-significance of this variable however, this change may be the result of random error.

Walk

Short Journeys

The final model of journeys by foot (based on crow-fly route) (see Appendix I. C1) of the 2002-03 data includes all the statistically significant variables ($p<0.01$) (*familiar*). The model has little explanatory power (R square = 0.070, F<0.01), even if it is greater than most of the models above. The model implies that the proportionate error of journeys with familiar origin and destination is lower compared with journeys with unfamiliar origin and/or destination.

The error for a typical journey of 0.5 miles of unfamiliar origin/destination can be calculated as:

2.433 - (4.508 * 0.5) = *0.179*

Long Journeys

The model of long journeys by foot (based on crow-fly route) (see Appendix I. C2) of the 2002-03 data includes all the statistically significant variables ($p<0.01$) (*familiar, familiar*long)*[7] and also the variable *long (p=0.078*). The variable *long* was included for theoretical reasons, given that we also include the interaction variable *famlong*. It should be noted that journeys by foot over 5.67 miles were excluded from the analysis, due to unreliable data and small sample size; around two-thirds of the journeys by foot in the 4th quartile (n=348) had distance estimates which were less than 0.1 miles. Again, the model had unsatisfactory explanatory power (R square = 0.095, F<0.01). This model suggests that the proportionate error decreases for each crow-fly mile, though this effect is counter-balanced by the positive effect of the dummy variable *long* on the error for journeys in the 3rd quartile. The model further predicts that familiar journeys have lower proportionate errors compared with unfamiliar journeys, but if the trip is both familiar and falls in the 3rd quartile, the negative effect is outweighed by the positive effect.

The error for a typical journey (based on the median distance for walk journeys) with unfamiliar destination/origin can be calculated as:

1.212 - (0.739 * 1.37) = *0.200*

[7] *Long* in this context refers to the 3rd Quartile, since walk journeys in the 4th quartile are excluded due to unreliable data.

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. C1 and C3 were tested on the 2004 data, and had somewhat reduced explanatory power (R square = 0.032 and R square =0.042 for short and long journeys, respectively) (see Appendix I. C2 and C4).

The beta-coefficients only remained significant for *long* and the crow-fly mileage in the model for long journeys ($p<0.01$), and it is also notable that the sign of *long* changes from positive to negative. In the model for short journeys, we note a substantial change of magnitude of the beta-coefficient of the crow-fly. This change may be partially explained by the improvement in the capture of walking stages in the 2004 survey.

Bike

Short Journeys

The final model of journeys by bike (based on quickest route) (see Appendix I. D1) of the 2002-03 data includes all the statistically significant variables ($p<0.05$): (*female*). The model has relatively good fit, compared with the other models above, even if it is still low (R square = 0.174, $F<0.01$). The model predicts that the proportionate error decreases with each mile that is travelled. It also suggests that proportional error is marginally lower for females than for males, holding trip length constant.

The error for a typical journey of 0.5 miles for males can be calculated as:

2.119 - (2.695 * 0.5) = *0.772*

Long Journeys

No independent variables were accepted ($p<0.05$) in the model of long journeys by bike, based on quickest route, of the 2002-03 data (see Appendix I. D3). Thus the error was again predicted using only the MapPoint mileage, and the beta-coefficient is once again negative. This model does not provide a satisfactory explanation of the dependent variable (R square = 0.043, $F<0.01$).

The error for a typical journey (based on the median distance for bike journeys) can be calculated as:

0.058 - (0.04 * 1.96) = *0.020*

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. D1 and D3 were tested on the 2004 data, and had somewhat poorer fit (R square = 0.013 and R square = 0.017 for short and long journeys respectively) (see Appendix I. D2 and D4). The variable *female* was no longer significant in the model for short journeys. Again, given the relatively high, but still significant, p-value in the model on 2002-03 data, it may be surprising that it is no longer significant in the model on 2004 data. Also, the beta-coefficient of MapPoint mileage in the model data for short journeys on 2004 kept the same sign, but was of smaller magnitude. With respect to the model for long journeys, the sign of the MapPoint mileage was the same, and the magnitude of the beta-coefficient was also similar.

4.3 Models of trip duration

Table 7 shows the summarised results of the regression models of trip duration, which are discussed in further detail below.

Table 7: Summarised results of duration models

	Short trips less than 0.86 miles			**Long trips 0.86 miles +**		
Trip type	**Predicting proportionate difference between respondent's estimates and MapPoint shortest route**			**Predicting proportionate difference between respondent's estimates and MapPoint quickest route**		
	Indep. variables accepted	**R2**		**Indep. variables accepted**	**R2**	
Car	None	0.130	Mainly random/unobs. Error falls with journey duration	Rural, 2nd quartile, 3rd quartile	0.065	Mainly random/unobs. Error falls with journey duration, and lower error for rural journeys
Other motorised	None	0.160	Mainly random/unobs. Error falls with journey duration	Rural, 2nd quartile	0.095	Mainly random/unobs. Error falls with journey duration, and lower error for rural journeys
Walk	Familiar	0.020	Mainly random/unobs. Lower error for familiar journeys, and error falls with journey duration	None	0.115	Mainly random/unobs. Error falls with journey duration
Bike	None	0.069	Mainly random/unobs. Error falls with journey duration	None	0.079	Mainly random/unobs. Error falls with journey duration

Car

Short Journeys

No independent variables were accepted ($p<0.05$) in the model of short journeys by car, based on shortest route, of the 2002-03 data (see Appendix I. E1). This model provides a relatively good explanation of the dependent variable (R square = 0.130, $F<0.01$), in comparison with the other models, even if it is still low. The error is predicted using only the MapPoint minutes, and the model implies that the proportional error decreases with journey duration.

The error for a typical short car journey (based on the median distance) for can be calculated as:

2.67 - (0.451 * 3.72) = *0.922*

Long Journeys

The final model of journeys by car (based on shortest route) (see Appendix I. E3) of the 2002-03 data includes all the statistically significant ($p<0.01$) variables (*rural destination, in 2nd Quartile, in 3rd Quartile*) ($p<0.01$). Once more, this model provides a fairly poor explanation of the dependent variable (R square = 0.065, $F<0.01$). The model suggests that the proportionate error decreases marginally with each additional MapPoint minute, which is also implied by the smaller magnitude of the beta-coefficients of *in 3rd quartile* compared with *in 2nd quartile.* Also, journeys in both the 2nd and 3rd quartile are likely to have larger errors than journeys in the 4th quartile. The model further implies that the error is lower for journeys with rural destinations, controlling for journey distance.

The error for a typical journey (based on the median distance of long car journeys) with an urban destination can be calculated as:

-0.083 - (0.002 * 13.18) = *0.023*

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. E1 and E3 were tested on the 2004 data (see Appendix I. A2 and E4). Both the models for short and long journeys had similar explanatory power to those based on the 2002-03 data (R square = 0.177 and 0.066, respectively), and all variables remained statistically significant ($p<0.01$), with the same signs and similar magnitudes.

Other motorised modes (single stage trips only)

Short Journeys

No independent variables were accepted in the model of short journeys by other motorised vehicles based on shortest route of the 2002-03 data ($p<0.05$) (see Appendix I. F1). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.160, $F<0.01$). The error is once more predicted using only the MapPoint minute, and is expected to decrease with journey duration.

The error for a typical short journey (based on the median distance) can be calculated as:

5.075 - (0.694 * 3.95) = *2.33*

Long Journeys

The final model of journeys by other motorised vehicles (based on shortest route) (see Appendix I. F3) of the 2002-03 data includes all the statistically significant ($p<0.01$) variables (*rural destination, in 2nd Quartile*). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.095, $F<0.01$). The model predicts that the proportional error decreases with each additional MapPoint minute, and again the proportional error is predicted to be larger for journeys in the 2nd quartile compared with longer journeys. Trips with rural destinations are also likely to have lower proportionate errors than journeys with urban destinations.

The error for a typical long journey (based on the median distance) with urban destination can be calculated as:

0.721 - (0.006 * 12.23) = *0.648*

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. F1 and F3 were tested on the 2004 data (see Appendix I. F2 and F4). Both the models for short and long journeys had greater explanatory power to those based on the 2002-03 data (R square = 0.306 and 0.103, respectively), and all variables remained statistically significant (p<0.01), with the same signs and similar magnitudes.

Walk

Short Journeys

The final model of journeys by foot (based on shortest route) (see Appendix I. G1) of the 2002-03 data includes all the statistically significant variables (p<0.001) (*familiar*). Again, journeys by foot over 5.67 miles were excluded from the analysis. The model has very little explanatory power (R square = 0.020, F<0.01). Again, this model implies that the proportional error is lower for each additional MapPoint minute and that error is lower for familiar journeys, ceteris paribus.

The error for a typical short journey (based on the median distance) with unfamiliar origin/destination can be calculated as:

10.838 - (1.792 * 2.88) = *5.68*

Long Journeys

No independent variables were accepted in the model of long journeys on foot, based on shortest route, of the 2002-03 data (p<0.05) (see Appendix I. G3). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.115, F<0.01). Again the error is predicted to decrease with each additional MapPoint minute.

The error for a typical long journey (based on the median distance) can be calculated as:

3.612 - (0.206 * 7.13) = *2.143*

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as that in Appendix I. G1 and G3 were tested on the 2004 data (see Appendix I. G2 and G4). The model for short journeys had greater explanatory power than that based on the 2002-03 data (R square = 0.057), whereas the model for long journeys had marginally poorer fit (R square = 0.109). All variables remained statistically significant (p<0.01), with the same signs and similar magnitudes.

Bike

Short Journeys

No independent variables were accepted in the model of short journeys by bike based on quickest route[8] of the 2002-03 data (p<0.05) (see Appendix I. H1). This model does not provide a satisfactory explanation of the dependent variable (R square = 0.069, F<0.01).

[8] The model based on quickest route had marginally better fit for short bike journeys, compared with the model based on shortest route. We thus focus on this model in the analysis…

The error is predicted using only the MapPoint minutes and the model implies that the proportional error decreases with journey duration.

The error for a typical short journey (based on the median distance) can be calculated as:

*2.814 – (0.206*3.31) = 2.13*

Long Journeys

No independent variables were accepted in the model of long journeys by bike based on shortest route of the 2002-03 data (see Appendix I. H3), implying that the error was predicted using only the MapPoint mileage. This model does not provide a satisfactory explanation of the dependent variable (R square = 0.079, F<0.01).

The error for a typical long journey (based on the median distance) can be calculated as:

1.215 - (0.044 * 8.42) = *0.8445*

Application of models on 2004 data

Models with the same specifications (i.e. the same variables) as those in Appendix I. H1 and H3 were tested on the 2004 data (see Appendix I. H2 and H4). Both the models for short and long journeys had somewhat greater explanatory power than the models based on the 2002-03 data (R square = 0.080 and R square = 0.083 for short and long trips, respectively). The independent variables remained statistically significant (p<0.01), keeping the same signs. The model of short journeys had a slightly larger magnitude of the beta-coefficient of MapPoint however compared with the model on 2002-03 data, whereas in the model for long journeys it remained similar.

4.4 Models of absolute difference

The models presented above are all employ the proportionate difference as the dependent variable. We are aware that, given the rounding problem, this formulation is likely to be at a disadvantage when dealing with the very shortest trips (where the rounding error will be proportionately large). An alternative formulation, based on absolute differences might avoid this problem – though might not perform so well for the very longest trips.

To explore this issue further we have run several models based on the absolute difference. Initial indications are that they do not generally perform significantly better than the equivalent model based on the proportionate difference.

5 RESULTS FROM THE REGRESSION MODELS OF THE RESPONDENT'S ESTIMATE

5.1 Overview

The dependent variable for these models is the respondent's estimate (of distance or duration).

The MapPoint calculation of distance/duration via shortest road route or quickest road route (or crow-fly distance) is allowed to be one of the independent variables. As with the models of difference, different MapPoint calculations performed best for different modes. The pattern was again that the quickest route calculation appeared best for most distance estimates – except short walk and bike, while the shortest route calculation appeared best for the duration estimates.

Other potential independent variables included the respondent and journey descriptors listed in section 2.3.

5.2 Results

The results of the individual models are presented in Appendix II. We do not propose to discuss these models in any depth (given that this would be going beyond our specified task and that detailed interpretation of the variables is not a trivial task), but we do note that the models make use of many of the independent variables. We also note that the models providing the greatest level of explanation are those for journeys by car and other motorised vehicles, while those for walk and bike journeys are rather less good.

The MapPoint calculations based on the quickest-routes-by-car generally provide more explanation of respondents' estimates of trip distance than do those based on the shortest-routes-by-car and they in turn offer more explanation than do distances calculated for crow-fly routes.

For respondents' estimates of duration, however, the MapPoint calculations based on the shortest-routes-by-car generally offer more explanation than do those based on the quickest-routes-by-car and they in turn generally offer more explanation than do distances calculated for crow-fly routes. An exception to this rule is provided by some journeys by bike or on foot where the crow-fly distance actually provided more explanation than do either of the MapPoint estimates of duration.

Broadly speaking, the models of respondents' estimates of duration were consistent with the models of their estimates of journey distance, but tended to provide a lower level of explanation.

Overall, the MapPoint calculation of journey duration provides a moderately good indication of the respondent's estimate of car journey duration, in particular for journeys with rural destination (perhaps because they tend to be subject to relatively little congestion?).

As expected, the MapPoint calculations of journey duration provide less explanation of respondents' estimates for journeys by foot and by bike, than they do for car journeys. The walk journey models offered the lowest degree of explanation.

The coefficients which emerge when the best of the models built on the 2002/3 database are applied to the 2004 database are similar, but not identical, to those found in the 2002/3 database. We cannot therefore report that the explanations are absolutely constant over time but can report that they appear to be fairly stable.

6 CONCLUSIONS

Based on this analysis we draw the following conclusions:

6.1 For models of difference:

None of the models of difference were able to provide much explanation of the difference between the respondent's estimate and the MapPoint calculation for trip distance or duration.

The proportionate differences, calculated for typical journeys are summarised in Table 8. From this we note that:

- The differences are all positive – indicating that the respondent's estimate was usually greater than the MapPoint calculation.
- For the most populous category (car trips 0.86 miles or longer) the proportionate difference was about 5%.
- The differences tend, except in the case of walk trips, to be significantly greater for short trips than for long trips (tending to confirm the existence of the rounding problem).
- For motorised trips, the differences were greater for trip distances than for trip durations, whereas for walk and bike trips the opposite tendency was apparent.
- The proportionate difference for duration estimates for short walk trips was significantly greater than for any other mode/length category.

Table 8: model estimates of proportionate difference (for typical trip in the given category)

	Proportionate difference in Distances		Proportionate difference in Durations	
	For trips <0.8 miles	For trips >0.8 miles	For trips <0.8 miles	For trips >0.8 miles
Car	1.03	0.056	0.92	0.02
Other motorised	3.03	0.026	2.33	0.65
Walk	0.18	0.200	5.68	2.14
Bike	0.77	0.020	1.83	0.85

The only independent variables to add significantly to the explanation were:

- the dummy variables relating to trip length - which suggested, unsurprisingly, that the proportionate difference decreased with trip length
- the familiarity of the destination (deduced from the journey purpose) – which suggested, again unsurprisingly, that the difference is reduced for familiar destinations
- whether the destination could be classed as rural – with a tendency, other things being equal, for the proportionate difference to be less when the destination is rural

- gender – with a slight tendency for the proportionate difference to be less when the respondent was female (only evident for car trips 0.86 miles or longer and for bike trips under 0.8 miles)

The coefficients which emerge when the best of the models built on the 2002/3 database are applied to the 2004 database are similar, but not identical, to those found in the 2002/3 database. We cannot therefore report that the explanations are absolutely constant over time but can report that they appear to be fairly stable.

6.2 For models of the respondent's estimate:

Respondents' estimates of trip lengths in a given database can be explained fairly satisfactorily on the basis of distances calculated using the trip end post codes provided that key characteristics of the journey and the respondent are taken into account.

The level of explanation (measured by the R Square) provided by the models of trip length was generally very good. Suggesting that, if one has access to the trip end post codes, one could make a fairly good prediction of what the respondent's estimate of trip length would have been, and so (arguably) one would not need to ask the question.
The models of trip duration are rather less good.

APPENDIX I: MODELS OF DIFFERENCE

I.A1

Best model of proportional difference for car trips 0.86 miles or less (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.231(a)	.053	.053	2.52917

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.570	.110		23.458	.000
	MapPoint mileage	-3.077	.180	-.231	-17.074	.000

a Dependent Variable: Prop difference between miles and quickest route

I.A2

Application of Model A1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.277(a)	.077	.076	2.58188

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.076	.170		18.093	.000
	MapPoint mileage	-3.944	.279	-.277	-14.115	.000

a Dependent Variable: Prop difference between miles and quickest route

I.A3
Best model of proportional difference for car trips over 0.86 miles (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.109(a)	.012	.012	.67204

a Predictors: (Constant), familiar, In 3rd Quartile, female, MapPoint mileage, In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.012	.008		1.591	.112
	MapPoint mileage	.000	.000	-.001	-.143	.886
	In 2nd Quartile	.179	.009	.118	19.921	.000
	In 3rd Quartile	.044	.008	.031	5.345	.000
	female	-.024	.007	-.018	-3.589	.000
	familiar	-.029	.007	-.021	-4.210	.000

a Dependent Variable: Prop difference between miles and quickest route

I.A4
Application of Model A3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.167(a)	.028	.028	.52443

a Predictors: (Constant), female, rural, In 3rd Quartile, purpose of journey, MapPoint mileage, In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.020	.009		-2.227	.026
	MapPoint mileage	.000	.000	.007	.929	.353
	Type of area of destination	-.018	.010	-.012	-1.746	.081
	In 2nd Quartile	.216	.010	.181	21.398	.000
	In 3rd Quartile	.045	.009	.041	4.871	.000
	familiar	-.006	.008	-.006	-.840	.401
	female	-.013	.007	-.012	-1.781	.075

a Dependent Variable: Prop difference between miles and quickest route

I.B1
Best model of proportional difference for other motorised trips 0.86 miles or less (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.370(a)	.137	.134	5.06115

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	8.485	.999		8.496	.000
	MapPoint mileage	-10.918	1.547	-.370	-7.059	.000

a Dependent Variable: Prop difference between miles and quickest route

I.B2

Application of Model B1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.620(a)	.384	.380	3.91341

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.354	.964		11.783	.000
	MapPoint mileage	-15.307	1.509	-.620	-10.142	.000

a Dependent Variable: Prop difference between miles and quickest route

I.B3

Best model of proportional difference for other motorised trips over 0.86 miles (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.135(a)	.018	.018	.53100

a Predictors: (Constant), In 2nd Quartile, MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.026	.009		2.799	.005
	MapPoint mileage	.000	.000	.003	.219	.827
	In 2nd Quartile	.162	.017	.136	9.763	.000

a Dependent Variable: Prop difference between miles and quickest route

I.B4
Application of Model B3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.137(a)	.019	.018	.51554

a Predictors: (Constant), In 2nd Quartile, MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.020	.014		1.436	.151
	MapPoint mileage	-.001	.001	-.020	-.969	.333
	In 2nd Quartile	.150	.024	.131	6.225	.000

a Dependent Variable: Prop difference between miles and quickest route

I.C1
Best model of proportional difference for walk trips 0.86 miles or less (given MapPoint crow-fly distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.265(a)	.070	.070	2.48303

a Predictors: (Constant), familiar, Crow-fly mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.433	.048		50.867	.000
	Crow-fly mileage	-4.508	.155	-.257	-29.060	.000
	familiar	-.233	.046	-.044	-5.022	.000

a Dependent Variable: Prop difference between miles and crowfly route

I.C2
Application of Model C1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.180(a)	.032	.032	2.15286

a Predictors: (Constant), familiar, Crow-fly mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.439	.038		37.688	.000
	Crow-fly mileage	-.007	.001	-.104	-7.965	.000
	familiar	-.672	.058	-.152	-11.656	.000

a Dependent Variable: Prop difference between miles and crowfly route

I.C3
Best model of proportional difference for walk trips over 0.86 miles (and less than 5.67 miles) (given MapPoint crow-fly route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.309(a)	.095	.095	1.23438

a Predictors: (Constant), Longer than median, familiar, famlong, Crow-fly mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.212	.037		33.049	.000
	Crow-fly mileage	-.739	.039	-.340	-18.915	.000
	familiar	-.198	.037	-.076	-5.280	.000
	famlong	.244	.100	.042	2.426	.015
	Longer than median	.136	.077	.038	1.764	.078

a Dependent Variable: Prop difference between miles and crowfly route

I.C4
Application of Model C3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.282(a)	.080	.078	.92526

a Predictors: (Constant), Longer than median, familiar, Crow-fly mileage, famlong

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.499	.026		18.887	.000
	Crow-fly mileage	-.004	.000	-.145	-7.906	.000
	familiar	-.040	.039	-.021	-1.049	.294
	famlong	.143	.097	.037	1.475	.140
	Longer than median	-.675	.064	-.257	-10.564	.000

a Dependent Variable: Prop difference between miles and crowfly route

I.D1
Best model of proportional difference for bike trips 0.86 miles or less (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.418(a)	.174	.169	1.34819

a Predictors: (Constant), female, MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.119	.209		10.154	.000
	MapPoint mileage	-2.695	.356	-.398	-7.575	.000
	female	-.334	.157	-.112	-2.125	.034

a Dependent Variable: Prop difference between miles and quickest route

I.D2

Application of Model D2 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.129(a)	.017	.005	1.00431

a Predictors: (Constant), female, MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.742	.249		2.978	.003
	MapPoint mileage	-.703	.413	-.135	-1.700	.091
	female	-.123	.163	-.060	-.753	.452

a Dependent Variable: Prop difference between miles and quickest route

I.D3

Best model of proportional difference for bike trips over 0.86 miles (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.206(a)	.043	.041	.45132

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.058	.024		2.402	.017
	MapPoint mileage	-.040	.007	-.206	-6.046	.000

a Dependent Variable: Prop difference between miles and quickest route

I.D4

Application of Model D3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.132(a)	.017	.015	.42836

a Predictors: (Constant), MapPoint mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.011	.036		.295	.768
	MapPoint mileage	-.031	.012	-.132	-2.689	.007

a Dependent Variable: Prop difference between miles and quickest route

I.E1

Best model of proportional difference for car trips 0.86 miles or less (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.360(a)	.130	.130	1.85358

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.670	.074		36.177	.000
	MapPoint time in minutes (shortest route)	-.451	.016	-.360	-27.713	.000

a Dependent Variable: Prop diff duration (shortest)

I.E2
Application of Model E1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.421(a)	.177	.177	1.68876

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.844	.101		28.241	.000
	MapPoint time in minutes (shortest route)	-.497	.022	-.421	-22.651	.000

a Dependent Variable: Prop diff duration (shortest)

I.E3
Best model of proportional difference for car trips over 0.86 (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.254(a)	.065	.064	.64527

a Predictors: (Constant), In 3rd Quartile, Rural, MapPoint time in minutes (shortest route), In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.083	.008		-10.830	.000
	MapPoint time in minutes (shortest route)	-.002	.000	-.105	-18.814	.000
	Rural	-.071	.008	-.042	-8.766	.000
	In 2nd Quartile	.297	.009	.198	32.302	.000
	In 3rd Quartile	.132	.008	.094	15.831	.000

a Dependent Variable: Prop diff duration (shortest)

I.E4
Application of Model E3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.257(a)	.066	.066	.65501

a Predictors: (Constant), In 3rd Quartile, Rural, MapPoint time in minutes (shortest route), In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.069	.011		-6.357	.000
	MapPoint time in minutes (shortest route)	-.002	.000	-.109	-13.737	.000
	Type of area of destination	-.093	.013	-.051	-7.283	.000
	In 2nd Quartile	.295	.013	.194	22.172	.000
	In 3rd Quartile	.114	.012	.080	9.421	.000

a Dependent Variable: Prop diff duration (shortest)

I.F1
Best model of proportional difference for other motorised trips 0.86 miles or less (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.400(a)	.160	.157	2.78206

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.075	.445		11.401	.000
	MapPoint time in minutes (shortest route)	-.694	.090	-.400	-7.697	.000

a Dependent Variable: Prop diff duration (shortest)

I.F2
Application of Model F1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.553(a)	.306	.302	2.32656

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.083	.492		12.372	.000
	MapPoint time in minutes (shortest route)	-.890	.104	-.553	-8.531	.000

a Dependent Variable: Prop diff duration (shortest)

I.F3
Best model of proportional difference for other motorised trips over 0.86 miles (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.309(a)	.095	.095	1.05825

a Predictors: (Constant), Rural, MapPoint time in minutes (shortest route), In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.721	.023		31.790	.000
	MapPoint time in minutes (shortest route)	-.006	.000	-.186	-13.610	.000
	In 2nd Quartile	.334	.034	.135	9.807	.000
	Rural	-.471	.043	-.145	-10.963	.000

a Dependent Variable: Prop diff duration (shortest)

I.F4
Application of Model F3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.321(a)	.103	.102	1.11190

a Predictors: (Constant), Rural, MapPoint time in minutes (shortest route), In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.859	.038		22.462	.000
	MapPoint time in minutes (shortest route)	-.010	.001	-.223	-10.668	.000
	In 2nd Quartile	.378	.054	.146	6.970	.000
	Rural	-.272	.079	-.069	-3.439	.001

a Dependent Variable: Prop diff duration (shortest)

I.G1
Best model of proportional difference for walk trips 0.86 miles or less (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.141(a)	.020	.020	22.98164

a Predictors: (Constant), familiar, MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10.838	.484		22.404	.000
	MapPoint time in minutes (shortest route)	-1.792	.123	-.133	-14.537	.000
	familiar	-1.688	.431	-.036	-3.912	.000

a Dependent Variable: Prop diff duration (shortest)

I.G2
Application of Model G1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.239(a)	.057	.057	11.91493

a Predictors: (Constant), familiar, MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.814	.362		27.139	.000
	MapPoint time in minutes (shortest route)	-1.548	.092	-.222	-16.891	.000
	familiar	-1.429	.325	-.058	-4.396	.000

a Dependent Variable: Prop diff duration (shortest)

I.G3
Best model of proportional difference for walk trips over 0.86 miles (and less than 5.67 miles) (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.338(a)	.115	.114	1.96658

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.612	.075		47.864	.000
	MapPoint time in minutes (shortest route)	-.206	.008	-.338	-25.733	.000

a Dependent Variable: Prop diff duration (shortest)

I.G4

Application of Model G3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.330(a)	.109	.108	1.85563

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.374	.098		34.405	.000
	MapPoint time in minutes (shortest route)	-.189	.010	-.330	-18.240	.000

a Dependent Variable: Prop diff duration (shortest)

I.H1

Best model of proportional difference for bike trips 0.86 miles or less (given MapPoint quickest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.263(a)	.069	.066	2.13787

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.814	.313		8.987	.000
	MapPoint time in minutes (shortest route)	-.347	.074	-.263	-4.661	.000

a Dependent Variable: Prop diff duration (shortest)

I.H2

Application of Model H1 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.282(a)	.080	.074	5.17568

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.510	1.096		5.026	.000
	MapPoint time in minutes (shortest route)	-1.051	.271	-.282	-3.878	.000

a Dependent Variable: Prop diff duration (shortest)

I.H3

Best model of proportional difference for bike trips over 0.86 miles (given MapPoint shortest route distance) (using 2002-03 data)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.281(a)	.079	.078	1.03368

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.215	.073		16.636	.000
	MapPoint time in minutes (shortest route)	-.044	.005	-.281	-8.387	.000

a Dependent Variable: Prop diff duration (shortest)

I.H4
Application of Model H3 on 2004 data

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.288(a)	.083	.081	1.16406

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.341	.119		11.254	.000
	MapPoint time in minutes (shortest route)	-.053	.009	-.288	-6.051	.000

a Dependent Variable: Prop diff duration (shortest)

APPENDIX II: MODELS OF RESPONDENT'S ESTIMATES

Models with respondent's estimate of trip distance as dependent variable, 2002-03

II.1 Car

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.968(a)	.936	.936	4.79107

a Predictors: (Constant), female, familiar, In 3rd Quartile, MapPoint mileage, In 2nd Quartile

Coefficients(a)

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	.465	.048		9.703	.000
	MapPoint mileage	.977	.001	.966	776.507	.000
	In 2nd Quartile	-.070	.059	-.002	-1.194	.233
	In 3rd Quartile	-.145	.054	-.004	-2.695	.007
	familiar	-.241	.046	-.006	-5.225	.000
	female	-.077	.045	-.002	-1.728	.084

a Dependent Variable: Miles travelled

II.2 Other Motorised Vehicles

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975(a)	.952	.952	4.36504

a Predictors: (Constant), In 2nd Quartile, MapPoint mileage

Coefficients(a)

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.157	.074		-2.133	.033
	MapPoint mileage	1.056	.003	.977	326.597	.000
	In 2nd Quartile	.336	.135	.007	2.491	.013

a Dependent Variable: Miles travelled

II.3 Walk

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.516(a)	.266	.266	.49688

a Predictors: (Constant), Longer than median, familiar, famlong, Crow-fly mileage

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.339	.007		51.838	.000
	Crow-fly mileage	.738	.012	.568	63.960	.000
	familiar	-.014	.008	-.012	-1.746	.081
	famlong	.395	.038	.086	10.338	.000
	Longer than median	-.365	.029	-.131	-12.643	.000

a Dependent Variable: Miles travelled

II.4 Bike

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.614(a)	.377	.376	1.45847

a Predictors: (Constant), female, MapPoint mileage (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.417	.061		23.245	.000
	MapPoint mileage (shortest route)	.327	.013	.594	25.194	.000
	female	-.451	.092	-.115	-4.885	.000

a Dependent Variable: Miles travelled

Models with respondent's estimate of trip duration as dependent variable, 2002-03

II.5 Car

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.876(a)	.768	.768	12.496

a Predictors: (Constant), Rural, In 3rd Quartile, MapPoint time in minutes (shortest route), In 2nd Quartile

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.319	.116		37.246	.000
	MapPoint time in minutes (shortest route)	.652	.002	.878	360.219	.000
	In 2nd Quartile	-.426	.158	-.007	-2.704	.007
	In 3rd Quartile	.449	.142	.008	3.150	.002
	Rural	-1.795	.150	-.027	-11.952	.000

a Dependent Variable: Journey duration in minutes (computed from dep/arr times)

II.6 Other Motorised Vehicles

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.813(a)	.660	.660	18.432

a Predictors: (Constant), In 2nd Quartile, MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10.117	.360		28.122	.000
	MapPoint time in minutes (shortest route)	.808	.008	.808	99.942	.000
	In 2nd Quartile	-1.284	.581	-.018	-2.210	.027

a Dependent Variable: Journey duration in minutes (computed from dep/arr times)

II.7 Walk

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.360(a)	.129	.129	13.154

a Predictors: (Constant), familiar, MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.962	.196		40.686	.000
	MapPoint time in minutes (shortest route)	1.476	.029	.359	50.065	.000
	familiar	-1.151	.205	-.040	-5.619	.000

a Dependent Variable: Journey duration in minutes (computed from dep/arr times)

II.8 Bike

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.623(a)	.388	.387	9.962

a Predictors: (Constant), MapPoint time in minutes (shortest route)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.449	.409		20.637	.000
	MapPoint time in minutes (shortest route)	.722	.027	.623	26.617	.000

a Dependent Variable: Journey duration in minutes (computed from dep/arr times)