

Distance in the Travel Diary – comparisons with GIS distances

1. Introduction

1.1 The Scottish Household Survey (SHS) collects 'Travel Diary' information from a randomly chosen adult (aged 16+) in the household. This data provides essential information about travel patterns and behaviours of Scottish residents. It is used extensively both internally (i.e. by Transport Scotland and Scottish Government) and externally to assist the making and monitoring of policy, resource allocation, research and marketing campaigns.

1.2 As with all survey information, the precision of the data is largely dependent on the accuracy of respondents' answers, the data cleaning process and the methodology behind the derivation of outputs.

1.3 The current "straight line" method of calculating distances in the Travel Diary is one such area which has been suggested for improvement as it is likely to underestimate actual distance travelled.

1.4 This paper attempts to provide an approximate value for the extent of underestimation of distance, explore which journeys are most affected by this and determine the overall impact of using this method for distance calculations. This will also help inform future work around the distance calculation within the Travel Diary.

2. Background to the Travel Diary

2.1 The Travel Diary is a series of interviewer administered questions which asks respondents about their travel on the previous day. Information is collected about each journey (or stage of the journey) such as mode, purpose, start time and duration. The origin and destination of each journey (or stage) is collected to allow the distance travelled to be calculated at a later stage. Associated information such as experience of congestion, parking payment and number of occupants per car journey is also asked.

2.2 Since its introduction in 1999, the Travel Diary has undergone some changes to help enhance the quality of the data it generates. These include:

- improvements to question wording and the necessity to report journeys less than ¼ mile or 5 minutes on foot to reduce underreporting of journeys;
- an improved weighting system to better account for non-response bias;
- changes in the way information is collected and recorded, e.g. duration asked instead of calculated, return journey question introduced.

2.3 Most of these took effect from 2007. In 2009, details were collected for a total of 18,679 journeys and 18,934 stages via the Travel Diary. This paper focuses on a sample¹ of 2009 records taken from the stage file.

¹ Further information about the sample profile is available in Annex A.

3. Address information and distance calculation in the Travel Diary

3.1 As mentioned above, the origin and destination of each journey (or stage) is collected during the interview. The respondent is asked to provide as much address information as possible (street name, postcode etc.) for all origins and destinations.

3.2 Where a postcode has been given, this is checked during data processing - any erroneous ones are amended. Where no postcode has been supplied, whatever address information has been provided is used to determine one - so the more information available, the more accurate the postcode e.g. a street name will lead to a more precise postcode than an approximate location such as 'city centre'. For some records, the address information will not be accurate or detailed enough to determine a full postcode. It is sometimes possible to assign a postal district (e.g. EH10), but in other cases, no postcode information can be allocated at all.

3.3 As it is the postcodes which are used to calculate travel distances, it is important that accurate postcode information is available. Distances are calculated by using the grid co-ordinates of the origin and destination postcodes to find the "straight line" distance between these two points².

3.4 However, as most journeys are not made in a straight line - road networks built around housing estates, bus routes devised to provide the most efficient service etc. - distances calculated in this way are likely to underestimate the actual distance travelled.

3.5 Where sufficient detail of the origin/destination could not be collected to assign a postcode, or the origin and destination information is the same³ (i.e. a "circular" journey), the distance is calculated as:

$$\text{Average mode speed} \times \text{Time taken for journey}^4$$

3.6 Circular journeys are particularly problematic as there is no way of knowing if the total time includes stopping off somewhere. For example, a respondent may state that a circular walking journey lasted 1 hour, however, this may actually only have involved a 15 minute walk each way with a half hour stop in between. As this information is not known, the calculation will not take account of this and journey distances may be overestimated for these particular cases.

3.7 The distance of a multi-stage journey is calculated by adding up the distances of each of its component stages. For journeys involving a series of calls (e.g. work to visit friends to shops to home) respondents are asked to estimate the total distance travelled.

3.8 During data processing, the distances calculated (via the straight line method) are checked against other information, such as journey time, to ensure they do not

² Straight line distance is calculated using Pythagoras Theorem $a^2 + b^2 = c^2$

³ In 2009, approximately 7% of records were not assigned either an origin and/or destination postcode and a further 8% had origin and destination postcodes that were the same.

⁴ Average mode speed is calculated from other Travel Diary records. Time taken is reported by the respondent.

result in unrealistic speeds e.g. a car travelling at 150km/hr. Where this is found to be the case, the distances are recalculated using the $d = s \times t$ formula above⁵.

3.9 It is therefore evident that the precision of results is largely dependent on the quality of respondent information (e.g. address, journey length). This issue is one which is more difficult to resolve or remove completely, especially with questions which rely on interviewees having to recall information.

4. GIS distance calculation

4.1 Geographic Information System (GIS) software can be used to obtain a more realistic distance calculation by mapping the origin and destination data to a route following the road or public transport network as appropriate for the mode⁶ of transport.

4.2 A sample of records from the 2009 Travel Diary were selected and ran through the GIS software to obtain a comparison between distances calculated this way and those using the current straight line method. This would verify whether distances are underestimated in the Travel Diary by using the straight line method and provide an indication of the extent of this.

4.3 The sample records were selected to ensure that all modes and geographical areas were adequately represented as well as providing large enough sample sizes to obtain breakdowns by journey characteristic e.g. purpose, length. From this, it can be determined if particular journeys were more likely to be affected by the underestimation of distance than others.

4.4 The GIS software allowed for two different calculations of distance:

- **Minpath** – this is the length of the journey which covered the least physical distance
- **Mintime** – this is the length of the journey which took the shortest amount of time

4.5 It is difficult to say with any certainty which calculation is most accurate. In reality, journeys will vary between the minpath and mintime routes, with people likely to switch between, or use a combination of, the two (if either). Choice of travel route will also depend on external factors such as time of travel (e.g. peak vs non-peak), road works, weather, preferential route (e.g. don't like motorways, more familiar with route) or other tasks (e.g. drop by a post box en route).

4.6 It was not always possible for the GIS software to calculate a value for minpath. This may be due to a shortest path being available but only one which followed a private or restricted road and therefore a value would not be generated. However, a mintime would still be possible by following roads where access was permitted. For records with no minpath value, this was set equal to mintime.

⁵ Approximately 3% of journeys in 2009.

⁶ Car journeys were mapped to the road network, bus journeys to bus routes and timetables, train journeys to the rail network and timetables etc.

4.7 For circular journeys (those with the same origin and destination postcode), the minpath and mintime values would provide a value of zero. In these cases, the distance would have to be calculated using an alternative method which did not rely on address information, perhaps along the same lines as that currently used in the Travel Diary (distance = speed x time).

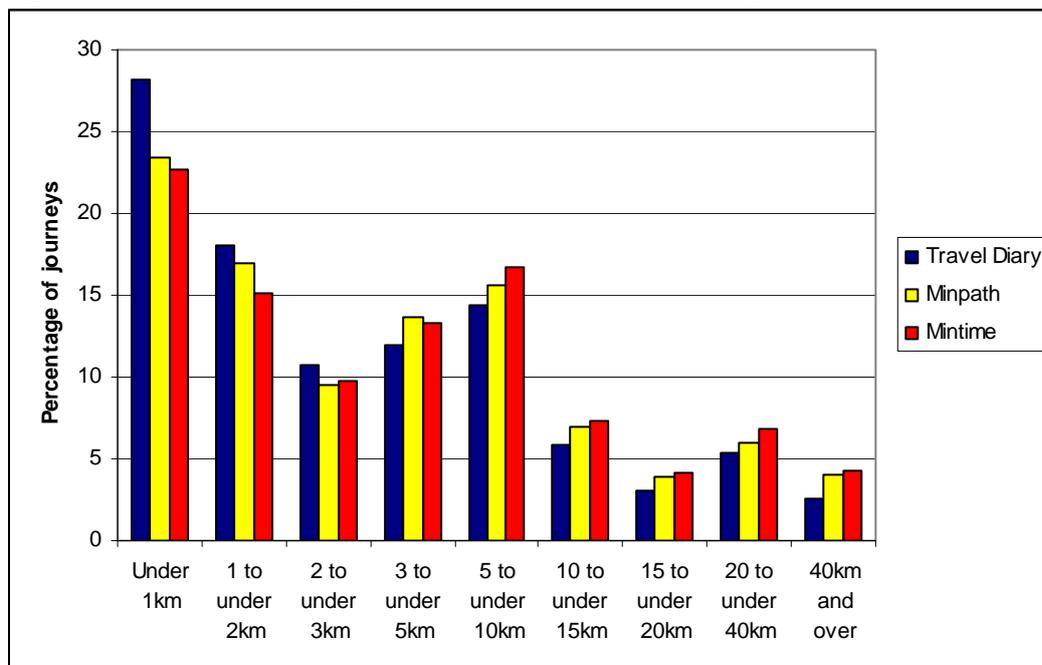
5. Comparison of distance calculations

5.1 As zero distances are meaningless, and given that the focus of the paper is to look at differences in the values generated by the straight line method and the GIS software, circular journeys have been removed. So, considering only journeys with a positive mintime or minpath value, the comparison of different distance calculations are shown graphically in Figure 1 with the corresponding values given in Table 1.

5.2 As all data comes from a sample survey, and the records used for this analysis were not selected using a statistical sampling technique, all results should be treated with caution as they will be subject to sampling variability. The lower the sample size, the higher the variability will be around the corresponding estimate.

5.3 The SHS data is weighted to take account of sampling bias and the Travel Diary data presented in this paper will take account of this weighting. However, data for minpath and mintime calculations have not been weighted as the sample was not selected to be representative across the range of travel characteristics. Annex A shows the sample is in fact broadly similar to the Travel Diary for all characteristics with the exception of modal split. For reference, modal share was used to calculate a weighting for minpath and mintime mean distances – this comparison is also shown in Annex A.

Figure 1: Distance comparisons for non-zero minpath/mintime



5.4 Figure 1 shows that the proportions for journeys less than 3km are greatest in the Travel Diary but all others are greater for the GIS mintime and minpath calculations. This verifies that distances are underestimated when calculated using the straight line method.

Table 1: Distance (km) calculation comparisons for non-zero minpath/mintime

	Travel Diary	Minpath	Mintime	Difference from Travel Diary	
				Minpath	Mintime
	<i>column percentages</i>				
Under 1 km	26.2	21.6	20.9	-18%	-20%
1 to under 2km	17.2	16.6	14.9	-3%	-13%
2 to under 3km	11.3	9.4	9.3	-17%	-18%
3 to under 5km	12.5	14.5	14.6	16%	17%
5 to under 10km	15.5	16.6	17.5	7%	13%
10 to under 15km	5.8	7.1	7.6	22%	31%
15 to 20km	2.7	3.6	3.5	33%	30%
20 to 40km	5.7	6.2	7.1	9%	25%
40km and over	3.0	4.5	4.7	50%	57%
<i>Sample size</i>					5,717

5.5 It can be seen from Table 1 that, in general, the minpath calculation generates values that are closer to the straight line method used currently. Of course, this makes sense since the Travel Diary calculation is a particular type of minimum path calculation, just not one that takes account of transport networks. Mintime calculations are also more likely to include roads such as motorways which result in a shorter journey time but cover a greater distance.

5.6 Looking at the summary statistics (Table 2) gives a better indication of the spread of the data.

Table 2: Distance (km) summary statistics for non-zero minpath/mintime

	Travel Diary	Minpath	Mintime	Difference from Travel Diary	
				Minpath	Mintime
Lower decile	0.4	0.5	0.5	25%	25%
Lower quartile	0.9	1.2	1.2	33%	33%
Median	2.5	3.2	3.6	28%	44%
Upper quartile	7.1	8.6	9.2	21%	30%
Upper decile	16.9	21.6	23.1	28%	37%
Mean	7.0	8.7	9.3	24%	33%
<i>Sample size</i>					5,717

5.7 Table 2 shows that the minpath and mintime distances are 24%-33% longer than those reported currently. Distances are more similar at the lower end of the data (it should be noted that smaller values will be more susceptible to fluctuations, with small changes in values being likely to result in large percentage changes). At the upper end of the data, the difference between minpath and mintime calculations are more apparent – the minpath calculation for the upper decile is 28% higher than the Travel Diary compared to 37% for mintime. As journey length increases, more

route options will become available which may further explain these larger differences.

5.8 Looking at the median (or middle) value of each of the calculations, minpath shows a 28% increase and mintime a 44%. As the median is much lower than the mean, this suggests that most journeys are likely to cover a shorter distance. The mean calculation shows a 24% difference for minpath and 33% difference for mintime. Due to the fact that both the mean and median values show an underestimation of distance, for ease of reference, only one will be used for comparison throughout the paper. As the mean takes in to consideration all journeys, and the median record is likely to be different across all three data sets, the mean will be used. However, it should be noted that mean values may be distorted by longer journeys which, as seen above, are more likely to show greater differences between the calculation methods.

6. Journey characteristics

6.1 It is important to establish whether this underestimating of distances is more apparent in particular types of journeys or if it is uniform across them all. Looking at journey characteristics – purpose, duration etc. – enables journeys which are more susceptible to distance underestimation to be identified.

Mode

Table 3: (Mean) Average distance (km) by mode

	Travel Diary	Minpath	Mintime	Difference from Travel Diary		Sample size
				Minpath	Mintime	
Walking	1.0	1.8	1.8	80%	80%	1,574
Driver car/van	10.0	12.7	13.4	27%	34%	1,758
Passenger car/van	9.5	11.8	12.5	24%	32%	462
Bicycle	3.2	3.8	3.8	19%	19%	134
Bus	7.1	8.3	9.1	17%	28%	1,475
Rail	18.9	20.7	22.6	10%	20%	236

6.2 Table 3 shows that walking journeys are underestimated by nearly half – the highest proportionate difference (although small changes will have a greater impact on small numbers). Bicycle and rail journeys are shown to experience the least amount of underestimation and car journeys are in line with the overall average.

6.3 Bus journeys appear to suffer less from an underestimating of distance than car journeys which is slightly unexpected given that bus routes are more likely to be less direct than those taken if travelling by car. This may be partly explained by journey length - car journeys, on average, are longer than those undertaken by bus, and since longer journeys are more likely to underestimate distance, car journey figures may be slightly distorted by long journeys. The quality of address information is also likely to vary between modes e.g. bus users have less need to know exact street names, postcodes etc. than car drivers. Buses will also be restricted to a particular route or type of road (e.g. non motorways), limiting the number of possible routes, although equally there will be some bus routes that cars are not permitted to

use. The GIS software allocated the nearest bus stop to the origin/destination postcode, however there is no way of knowing if this is the one that was in fact used.

6.4 Caution should be taken when analysing the mode figures. Annex A shows that the sample of records are not representative of travel modes across the survey as a whole. This is due to the fact the sample was selected to provide a large enough sample size for each mode (which would not necessarily have been the case if this had been proportionate to actual travel diary information) and to ensure a range of geographical coverage.

Purpose

Table 4: (Mean) Distance (km) by purpose

	Travel Diary	Minpath	Mintime	Difference from Travel Diary		Sample size
				Minpath	Mintime	
Commuting	9.3	11.4	12.3	23%	32%	1,313
Education	6.3	7.6	8.2	21%	30%	282
Shopping	3.7	4.5	4.9	22%	32%	1,397
Visit hospital/ other health	6.2	7.2	7.7	16%	24%	228
Other personal business	8.6	11.3	12.1	31%	41%	453
Visiting friends/ relatives	6.5	8.0	8.5	23%	31%	582
Eating/drinking	7.3	8.2	8.6	12%	18%	253
Sport/entertainment	6.0	7.6	8.2	27%	37%	470
Escort	4.9	5.8	6.1	18%	24%	311
Go home	14.4	17.5	18.9	22%	31%	124
Just go for a walk	3.9	3.9	4.2	0%	8%	161

6.5 For most purposes, the extent of distance underestimation was comparable to the overall average. Only one journey purpose, ‘other personal business’, showed a greater underestimation, however, as this is a “catch-all” category, it is difficult to determine why this might be. Journeys with a purpose of ‘escort’, ‘visit hospital or health’ and ‘eating/drinking’ have a lesser degree of underestimation with journeys ‘to go for a walk’ being closest to the calculated travel diary distance. This may be due to the fact that journeys to go for a walk will not be as restricted – there will be less need to follow roads, shortcuts can be taken across parks etc. They are also more likely to cover a shorter distance.

6.6 The purpose ‘just go for a walk’ should not be confused with walking as a mode of travel. Although the vast majority of ‘just go for a walk’ journeys were undertaken on foot, the majority of journeys carried out on foot were for the purpose of shopping.

Urban/Rural

Table 5: (Mean) Distance (km) by urban/rural classification

	Travel Diary	Minpath	Mintime	Difference from Travel Diary		Sample size
				Minpath	Mintime	
Large urban	5.0	5.9	6.4	18%	28%	2,501
Other urban	9.0	11.1	11.9	23%	32%	1,532
Small accessible towns	6.9	8.1	8.7	17%	26%	541
Small remote towns	5.5	6.6	6.9	20%	25%	270
Accessible rural	9.7	11.7	12.5	21%	29%	408
Remote rural	13.2	19.3	20.1	46%	52%	465

6.7 Remote rural areas experienced the highest proportionate amount of underestimation of distance. This is probably due to the nature of the roads in rural areas – they are likely to be longer and windier than city centre roads and hence further away from a straight line distance. The larger difference in rural areas may also be caused by the relatively high proportion of trips for ‘other personal business’ and relatively low proportions for ‘eating/drinking’ and bus travel.

6.8 Large urban areas and small accessible towns had minpath and mintime distances which were closest to the Travel Diary. This is in line with other findings since average distance increases with rurality and journeys with longer distances experience greater differences between methods of calculation.

6.9 It should be noted that the urban/rural classification relates to the home address of the respondent and not necessarily where the journey took place.

Duration

Table 6: (Mean) Distance (km) by duration

	Travel Diary	Minpath	Mintime	Difference from Travel Diary		Sample size
				Minpath	Mintime	
Less than 5 mins	0.8	1.8	1.9	125%	138%	276
5 to 10 mins	2.0	2.9	3.0	45%	50%	2,176
11 to 20 mins	4.5	5.6	6.0	24%	33%	1,583
21 to 30 mins	8.4	10.5	11.4	25%	36%	743
31 to 60 mins	17.5	20.2	21.6	15%	23%	179
61 to 120 mins	31.9	37.9	40.6	19%	27%	706

6.10 As journey distance and duration are likely to be highly related, i.e. the shorter the distance covered, the less time the journey will take and vice versa, it is unsurprising to see these findings coming through in the table above. The shorter the journey, the more likely the distance is to be underestimated, which again may be symptomatic of the smaller numbers involved at the lower end.

6.11 Journeys less than 5 minutes long appear to have distances reported in the Travel Diary which are less than half of the GIS values. A higher proportion of these journeys are undertaken on foot than any other journey length which may explain the difference.

6.12 Journeys 61-120 minutes in length have the biggest percentage difference but the lowest proportionate difference. This may be partly explained by the relatively high percentage of journeys for the purpose of 'go home' and the relatively low percentage of journeys for 'other personal business'. These journeys are also more likely to be carried out by bus than journeys with shorter durations and by those in urban (particularly 'other' urban) areas.

Start time

Table 7: (Mean) Distance (km) by start time of journey

	Travel Diary	Minpath	Mintime	Difference from Travel Diary		Sample size
				Minpath	Mintime	
Before 7am	12.5	14.8	15.9	18%	27%	170
7am to 9.30 am	7.6	9.4	10.1	24%	33%	1,035
After 9.30am to before 12pm	6.1	7.3	7.8	20%	28%	963
12 noon to 2pm	5.6	7.3	7.8	30%	39%	984
After 2pm to before 4.30pm	6.8	8.3	9.0	22%	32%	957
4.30pm to 6.30 pm	7.9	9.7	10.3	23%	30%	814
6.30 pm onwards	6.8	8.3	8.8	22%	29%	794

6.13 Journeys which started before 7am were least likely to be underestimated in distance. This may be because of the type of travel which is undertaken at this time. The majority of travel before 7am is for commuting purposes although it has been seen that the extent of underestimation for commuting journeys is in line with the overall average. The mode of travel may provide a greater understanding – there are less walking journeys at this time and a higher proportion of rail journeys.

6.14 Journeys starting between 12 noon and 2pm were most likely to be underestimated. Journeys starting at this time had relatively high proportions of journeys undertaken by walking and for the purpose of 'other personal business' which may explain this finding.

6.15 Journeys starting at all other times were roughly in line with the overall average.

7. Circular journeys

7.1 Circular journeys (origin=destination) were excluded from the above analysis due to the fact that they generated zero distances which were difficult to interpret. However, it is worthwhile bearing in mind that the exclusion of circular journeys will skew the figures. Given the complications around circular journeys as already described, it is difficult to know how these would be calculated more accurately unless additional information was available. Since this is not the case for the 2009 data, it is likely that distances would be calculated as they are currently ($d = s \times t$). This would mean the distances would remain the same for these journeys and the overall extent of underestimation will be less than that stated above.

7.2 Tables 8 and 9 show the distance comparisons and summary statistics for all journeys. Zero minpath and mintime distance have been substituted with the current Travel Diary values.

Table 8: Distance (km) calculation comparisons for all journeys

	Travel Diary	Minpath	Mintime	Difference from Travel Diary	
	<i>column percentages</i>			Minpath	Mintime
Under 1 km	26.8	22.6	21.9	-16%	-18%
1 to under 2km	17.2	16.6	15.0	-3%	-13%
2 to under 3km	11.3	9.5	9.4	-16%	-17%
3 to under 5km	12.5	14.3	14.4	14%	15%
5 to under 10km	15.5	16.5	17.4	6%	12%
10 to under 15km	5.6	6.8	7.2	21%	29%
15 to 20km	2.6	3.5	3.4	35%	31%
20 to 40km	5.4	5.8	6.7	7%	24%
40km and over	3.1	4.5	4.7	45%	52%
<i>Sample size</i>	6,239				

Table 9: Distance (km) summary statistics for all journeys

	Travel Diary	Minpath	Mintime	Difference from Travel Diary	
				Minpath	Mintime
Lower decile	0.4	0.5	0.5	25%	25%
Lower quartile	0.9	1.1	1.2	22%	33%
Median	2.5	3.2	3.5	28%	40%
Upper quartile	6.9	8.2	8.9	19%	29%
Upper decile	16.7	20.6	22.5	23%	35%
Mean	7.0	8.5	9.1	21%	30%
<i>Sample size</i>	6,239				

7.3 As expected, the data shows a similar pattern as the values in Tables 1 and 2, but with lesser differences between the methods of calculation. The mean distance comparisons now become a difference of 21% for minpath and 30% for mintime (previously 24% and 33% respectively).

7.4 Again, this will affect different types of journeys to varying extents. It is worthwhile noting that a large proportion of circular journeys are undertaken on foot.

8. Conclusion

8.1 Comparing the straight line distance calculations currently used in the Travel Diary with those calculated using the GIS software confirms that the distances reported through Travel Diary are underestimates of actual distance travelled.

8.2 It is difficult to provide an exact figure for underestimation due to unknown routes and the external factors that influence this choice. Comparing GIS and straight line calculation methods shows the extent of the underestimation will be between 24% and 33% depending on the route taken (that of least physical distance or that of least time). Differences are more apparent at the upper end of the distribution. As these figures are derived by concentrating on non-circular journeys only, in reality the true extent of underestimation will be slightly less than this.

8.3 The characteristics of the journey itself will impact upon the degree to which distance is underestimated. It is inevitable that journey characteristics will overlap e.g. a certain mode will be more likely to be used at a particular time of the day, and it is often difficult to separate these out to be able to provide an explanation as to why the distances are so different.

8.4 Looking at individual characteristics, journeys with a greater proportion of underestimation are those: undertaken on foot, for the purpose of 'other personal business', taken by residents of remote rural areas, which are shorter in duration and which start between 12 noon and 2pm. Some apparent large differences may be due to the fact that small numbers are involved.

8.5 Journeys which were least likely to be underestimated were: undertaken by bicycle or rail, for the purposes of 'escort', 'visit hospital or health' and 'eating/drinking' and particularly 'to go for a walk', by residents of large urban areas or small accessible towns, for journeys of greater durations and those which started before 7am.

9. Next steps

9.1 A 'new' SHS started in the field in January 2012. This includes revisions to the Travel Diary such as amended questions, layout and structure to improve data quality. Discussions are yet to be held regarding the distance calculation methodology but it is hoped that it will be possible to incorporate a GIS element in to this to further enhance quality.

9.2 The work outlined in this paper will contribute to these discussions and help inform decisions when agreeing upon an appropriate methodology. The distance calculations and comparisons will also be used when quality assuring the 'new' Travel Diary data.

9.3 The purpose of this paper is to provide an estimation of the extent of underestimation for distance calculations used in the current Travel Diary. It is to be used as a guide and reference document alongside the published distance data. There are no plans to calculate or publish a 'correction factor' for the historical data given the break in time series caused by the new contract and potential differences in calculation methodology. However, this may be reconsidered if there proves to be necessary demand.

9.4 Should the calculation methodology differ substantially from the current straight line method, the possibility of producing two different distance calculations – one using the 'new' method and the other using the 'old' straight line method – would be investigated. This would allow users to obtain a consistent time series as well as highlighting the differences between the calculations.

10. Annex A

Selection of records for use in GIS calculations

10.1 The sample was selected by choosing records which provided a large enough sample size for each mode of travel so that these could be broken down by journey characteristic. A statistical sampling technique was not used as this would be unlikely to provide a large enough sample sizes for particular modes.

10.2 It was important to ensure the data covered a range of geographic areas and the urban/rural classification was used as a guide.

10.3 The stage file was used to select the sample from (as opposed to the journey file) as this provided a greater level of detail. The journey file condenses journeys with more than one stage in to a single journey and assigns a 'main' mode (that used for the greatest distance) etc. so some information is lost. As it is the basic information that is required for distances, and some modes will be more likely than others to be missed off the journey file, the stage file was the most appropriate for this purpose.

10.4 The tables below show how the sample composition compares to that of the 2009 Travel Diary information as a whole. With the exception of mode (for reasons explained above), the sample is suitably representative of the 2009 data.

DISTANCE

	Sample	2009
Under 1 km	26.2	25.7
1 to under 2km	17.2	15.6
2 to under 3km	11.3	10.0
3 to under 5km	12.5	12.0
5 to under 10km	15.5	15.2
10 to under 15km	5.8	7.0
15 to 20km	2.7	3.7
20 to 40km	5.7	6.3
40km and over	3.0	4.5
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

MODE

	Sample	2009
Walking	27.1	22.3
Driver Car/Van	29.2	51.1
Passenger Car/Van	9.1	12.6
Other	0.7	1.4
Bicycle	2.3	0.9
Bus	25.1	8.9
Taxi/Minicab	0.9	1.3
Rail	5.7	1.5
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

PURPOSE

	Sample	2009
Commuting	25.1	23.2
Business	0.7	1.4
Education	6.7	2.9
Shopping	22.3	23.2
Visit hospital or other health	3.2	3.1
Other personal business	7.4	7.5
Visiting friends or relatives	8.9	11.6
Eating/drinking	5.4	3.7
Sport / Entertainment	8.3	7.4
Holiday / day trip	1.6	2.2
other not coded	0.3	0.6
Escort	5.2	7.0
Go home	2.3	3.2
Just go for a walk	2.8	3.2
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

URBAN/RURAL

	Sample	2009
Large urban areas	50.3	36.8
Other urban	26.4	28.3
Small accessible towns	7.6	9.2
Small remote towns	2.8	5.3
Accessible rural	7.1	10.6
Remote rural	5.7	9.8
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

DURATION

	Sample	2009
Less than 5 min	4.5	6.6
5 to 10 min	36.7	39.3
11 to 20 min	28.1	25.9
21 to 30 min	13.6	12.5
31 to 60 min	12.7	10.6
61 to 120 min	3.5	3.3
121 to 179 min	0.3	0.6
180 min and over	0.6	1.2
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

START TIME

	Sample	2009
Before 7am	3.7	3.2
7am to 9:30am	18.7	17.2
After 9:30am to before 12noon	14.3	16.1
12noon to 2 pm	16.6	17.3
After 2pm to before 4:30pm	16.4	16.6
4:30pm to before 6:30pm	15.5	15.7
6:30pm onwards	14.9	14.9
<i>Sample size</i>	<i>5,717</i>	<i>18,934</i>

Weighting minpath and mintime data

10.5 The SHS is weighted to take account of sampling bias and all Travel Diary data presented in this paper has taken account of this weighting. Due to the fact that the sample of records selected for this report were not chosen to be representative across the range of travel characteristics, no weighting has been applied to the minpath or mintime values.

10.6 As can be seen above, the sample selected is broadly similar to the Travel Diary for all characteristics with the exception of modal split. The difference in modal split figures was used to calculate a weighting factor for the minpath and mintime data. This was applied to the data to produce weighted mean distance values for minpath and mintime – the comparison between these and the non-weighted data is given in the table below. This is to be used as a broad indication only.

Mean distance (km) for non-zero values

	Minpath	Mintime
Non-Weighted	8.7	9.3
Weighted	9.7	10.3