

Scottish
Household
Survey

TRAVEL DIARY

User Guide

Contents

1.	Introduction.....	<i>Page 3</i>
2.	Background to the SHS Travel Diary.....	<i>Page 3</i>
3.	SHS Travel Diary datasets and variables.....	<i>Page 6</i>
4.	The collection and imputation of information about journeys.....	<i>Page 14</i>
5.	Origins and destinations, and distances travelled.....	<i>Page 16</i>
6.	“Home to Work” and “Home to School” distances.....	<i>Page 17</i>
7.	Enquiries and further information.....	<i>Page 18</i>

Annex

A.	Details of the problems found and the imputation methods used.....	<i>Page 19</i>
B.	Numbers of records before and after imputation.....	<i>Page 27</i>

1. Introduction

1.1 The Travel Diary collects information on personal travel, on the day prior to the interview, from randomly chosen adults (aged 16+) in Scotland. It is part of the Scottish Household Survey (SHS), which is a major cross-sectional survey, commissioned by the Scottish Government in order to provide reliable and up-to-date information on the composition, characteristics and behaviour of Scottish households, both nationally and at a sub-Scotland level. Please see the separate user guidance, “*Scottish Household Survey: User Guide*” for more information on the “main”, (i.e. “non-travel diary”) SHS data. Technical reports on the SHS can be purchased from the Stationery Office bookshop and found on the following website www.scotland.gov.uk/shs¹

1.2 From the outset, it was intended that the main SHS data and the SHS Travel Diary should be made publicly available for analysis. The aim of this document is to provide potential users and other interested parties with a description of the structure and content of the Travel Diary datasets.

1.3 This document has the following structure.

- A brief account of the background and a description of the coverage of the Travel Diary appear in *section 2*.
- The main definitions used for the Travel Diary data and a list of variables is given in *section 3*.
- Brief overview of collection and imputation of information about journeys and comparisons with the GB National Travel Survey (NTS) are covered in *section 4*.
- Information on how distance is calculated is provided in *section 5*.
- This document also describes how distances between “home and work” and “home and school” are estimated (they are provided in files which are separate to the Travel Diary), *section 6*.
- The *Annex* at the end of this document contains more details of the problems found in the data and imputation methods used.

2. Background to the SHS Travel Diary

2.1 Background

2.1.1 The SHS Travel Diary is part of the Scottish Household Survey (SHS). The SHS covers a wide range of topics to allow links to be made between different policy areas. There is a particular focus on information on transport, communities and local government. The SHS is financed by the Scottish Government and undertaken by a partnership of TNS-BMRB Scotland and MORI Scotland. It started in February 1999.

2.1.2 The sample for the survey is designed to meet a number of criteria. It is designed to provide nationally representative samples of private households and of the adult population in private households. This is achieved by splitting the interview between the Highest Income

¹ The website includes a full set of documentation including the Scottish Household Survey User Guide, Technical Reports, Questionnaire and Topic lists. This documentation is provided under the ‘Publications’ and ‘Survey details’ section of the website.

Householder or their spouse / partner and an adult selected at random from the permanent residents of the household (these may be the same person). The Travel Diary is asked of 75% of randomly chosen adults.

2.1.3 In order to meet the reporting requirements, the sample is structured to be nationally representative each quarter and to provide a representative sample for larger local authorities each year (those with over 120,000 households). Furthermore, it is also designed to provide data for each local authority over a two-year period. This is achieved by disproportionately sampling to achieve a minimum sample equivalent to a simple random sample of 500 in each local authority in each two-year period. Reweighting factors “correct” for the disproportionate sampling.

2.1.4 A combination of clustered and unclustered sampling is used. The Scottish Government’s urban rural classification is used to identify areas where sample should be clustered or unclustered. The general approach was that areas classified as ‘large urban areas’ or ‘other urban areas’ would use unclustered sampling while areas in the other four categories (accessible small towns, remote small towns, accessible rural and remote rural) would use clustered sampling. This approach is modified in two ways:

- Where more than 80% of households in a local authority fall into the ‘urban’ or ‘non-urban’ category, the whole local authority is treated as that category
- The three island local authorities (Eilean Siar, Orkney Islands and Shetland Islands) use wholly unclustered sampling even though their urban rural classification suggests that they should use wholly clustered sampling. In these areas, the sampling interval is between 1 and 6 households and 1 in 8 households, which means that clustered sampling would be no more efficient than unclustered sampling.

2.1.5 The sample is selected from the Small User File of the Postal Address File (PAF) for Scotland, expanded to take account of addresses which might only be listed once but actually contain multiple dwellings, such as tenement blocks and multi-storey flats. Although the small user PAF excludes many institutional addresses such as student halls of residence or nurses’ homes, there are no geographical exclusions from the survey.

2.1.6 The survey questionnaire is in two parts. The Highest Income Householder or his/her spouse/partner completes Part 1 (household level questions) of the interview. Once the composition of the household has been established, one of the adults in the household is randomly selected to complete Part 2 (individual questions) – 75% of these interviews will include the Travel Diary. In all households with a single adult the same person completes both parts but as the number of adults in the household increases, the probability of the random adult being the same as the household respondent declines.

2.1.7 Part 1 (i.e. the Highest Income Holder, or his/her spouse/partner section of the interview) deals with topics such as household composition, housing and tenure, health, the vehicles available to the household, the occupation and industry of the highest income householder, household income, financial services and housing costs. Part 2 (i.e. the random adult section) deals with individuals’ housing change, neighbourhood problems, community safety and anti-social behaviour, internet use, transport and use of public transport, public services, volunteering, culture and sport, income and employment as well as travel made on the previous day.

2.1.8 Interviewing was conducted using Computer Assisted Personal Interviewing (CAPI).

2.1.9 In 2009 and 2010, a total of 28,404 valid interviews were carried out with householders. Of these 28,404 households, interviews were achieved with 24,982 random adults (aged 16+) (88%). Of the 24,982 random adults interviews, 17,896 included the Travel Diary; 75% of these respondents made a journey on the previous day.

2.1.10 In summary, the SHS Travel Diary is asked of a random sample of adults (aged 16+) living in private households across Scotland about the travel they made on the day prior to the interview. This information is obtained from 75% of interviews with one randomly chosen adult per household in the sample. Each adult in the household has an equal chance of selection for these questions. Reweighting factors “correct” for the lower selection probability of individual adults in larger households.

2.2 The coverage of the SHS Travel Diary

2.2.1 The SHS Travel Diary collects information about travel for private purposes or for work or education, provided the main reason for the journey is for the traveller himself or herself to reach the destination. It includes the following types of travel:

- personal travel for domestic, social or recreational purposes - e.g. driving to and from work, travelling into town to go shopping, and going to see friends;
- journeys in the course of work, provided the purpose of the journey is for the traveller to reach a destination - e.g. taking a bus into town to attend a meeting, flying down to London on business;
- journeys made to take or accompany someone else - e.g. taking the children to school, walking someone home.

2.2.2 Journeys made by land, air or water within the United Kingdom are included. Journeys which start or end outwith the UK (e.g. a holiday flight from Spain) are excluded. However, if a respondent were to say that he or she had flown back from a holiday abroad on the previous day, the interviewer should record details of the journey home from the airport (but *not* record details of the flight to the UK).

2.2.3 The SHS Travel Diary does *not* cover:

- journeys which are made in the course of work by people who are employed (e.g.) as drivers or crew of public transport vehicles; to drive lorries; to deliver letters, parcels, leaflets or goods; as police officers or traffic wardens; etc. However, it does cover their journeys to and from their places of work;
- travel away from public roads or highways, such as walking or climbing in the hills, and yachting or flying for pleasure;

2.2.4 The Travel Diary collects information about three different types of journey:

- “single stage” journeys (95.39% of the journeys recorded in 2009/2010) - such as driving to an out-of-town shopping centre, or going from home to work by bus;
- “multi-stage” journeys (1.19%) - such as going by bus to the station (the first stage, assuming that the walk to the bus stop is below the threshold) and then catching a train to - say - Manchester (the second stage);
- “series of calls” journeys (3.42%) - in order to reduce the burden on respondents, travel involving a number of stops for the same main purpose and using the same form of transport is treated as one continuous “series of calls” journey from the first such call to

the last one. Only shopping and travel in course of work can be treated in this way. Information is collected about the part of the journey to the first call, the total number of calls, and the part of the journey after the last call. In the case of a doctor's round, for example, the interviewer might ask about the travel from (say) the surgery to the first patient, the total number of calls, and the travel from the final patient back to the surgery (or wherever the doctor went after the last call).

The three types are treated differently to facilitate the interviewing process. There are slight differences in the detailed kinds of information collected about each type of journey, which have implications for the processing and analysis of the data but do not affect the main conclusions that can be drawn.

3. SHS Travel Diary datasets and variables

3.1 The data for the 2009/2010 Travel Diary has been supplied in two files: a "Journey" file containing one record per journey (for multi-stage journeys, the record contains details of the origin and start time of the first stage and the destination and end time of the last stage; similarly, for series of call journeys, the record contains the information on the origin and start time of the first call and the destination and end time of the last call); and a "Stage" file containing one record for each stage of each journey, i.e. one record for each single stage journey, one record for each stage of a multi-stage journey, and, in the case of a series of call journey, one record for the first call and one record for the last call.

3.2 The basic definitions used in the Travel Diary

3.2.1 *Journeys:* the basic unit of travel, a journey, is defined as a one-way course of travel having a single main purpose. Outward and return halves of a return journey are treated as two separate journeys. A journey cannot have two separate purposes, and if a single course of travel involves a mid-way change of purpose then it, too, is split into two journeys. However, trivial subsidiary purposes (e.g. a stop to buy a newspaper) are disregarded.

3.2.2 *Stages:* a journey consists of one or more stages. A new stage is defined when there is a change in the form of transport or when there is a change of vehicle requiring a separate ticket.

3.2.3 *Journey purpose:* the purpose of a journey is normally taken to be the activity at the destination, unless that destination is "home" in which case the purpose is defined by the origin of the journey. A number of purposes are distinguished, such as "place of work", "in the course of work", "educational establishment", "shopping", and so on. Separate categories are used when the traveller has no purpose of his or her own, other than to escort or accompany another person; for example, taking a child to school.

3.2.4 *Mode of transport:* in the code-lists which are used by the interviewers to record details of the journeys, vans are counted with cars; taxis and minicabs are in a separate category from ordinary cars; and there are separate categories for (i) rail and (ii) underground, and for (a) school bus, (b) works bus and (c) ordinary (service) bus.

3.2.5 *Main mode of transport:* where a journey involves more than one mode of transport (e.g. first a bus and then a train), the main mode of a journey is defined, as in the GB National

Travel Survey, as that used for the longest (in distance) stage of the journey. For example, if a journey involved a two mile bus ride followed by a 50 mile train trip, the main mode for the journey would be "rail". It should be noted that this definition does *not* use the total of the distances travelled by each of the different modes to determine the main mode - for example, a journey involving a 1 mile walk to a bus stop, a 1½ mile bus ride and a 1 mile walk to the ultimate destination would be classified as "main mode = bus", because bus was the mode of transport used for the longest stage of the journey, even though more than half the total distance was covered on foot. If there is no single longest stage, and the two (or more) longest stages do not involve the same mode of transport, the "main" mode of the journey is defined (as in the GB National Travel Survey (NTS)) as the mode which was used for the last of the longest stages. In practice, because of the way that the distances are calculated (which produces results which appear to be accurate to about the nearest metre, although they will not be as precise as this - see *section 5*), it is very unlikely that there will be many journeys which have two stages which involve *exactly* the same distance.

3.3 List of variables on the Travel Diary Datasets

3.3.1 The following table shows the names of each of the travel diary variables, an explanation of what they are, and a list of possible values where applicable.

(J) = on journey dataset only, (S) = on stage dataset only.

<u>VARIABLE NAME</u>	<u>LABEL</u>	<u>VALUE / CODE</u>
UNIQID	Unique household identifier	Any numeric
DYEAR	The year's data the household is included in for the annual reports	2009, 2010 etc.
QUARTER	The quarter in which the interview took place	1,2,3,4
TRIPNO	Number of the journey – derived in chronological order, derived from the reported start times of each journey	Numeric
STAGE (S)	Number of stage of multi-stage journey	Numeric: <ul style="list-style-type: none"> • 1-n if multi-stage journey • “missing” if single stage journey • 1 if a “series of calls” journey

DD	Day of travel	1 - 31
MM	Month of travel	1 - 12
YY	Year of travel	e.g. 2009, 2010, etc.
TRAVDAY	Day of travel	i.e. 1 = Monday, 2 = Tuesday etc.
RE16	Type of journey	1 = Single stage journey, 2 = Multi-stage journey, 3 = Series of calls journey
TRAV_WT	Travel Diary weight: weighting factor for Travel diary data: “corrects” for differences in selection probabilities between local authorities, between adults in different sizes of household and between days on which people are available to be interviewed. See <i>section 4.2</i> for more details.	Numeric
MODE (S)	Mode of transport used for the stage of the journey	1 = Walking 2 = Driver Car/Van 3 = Passenger Car/Van 4 = Motorcycle/ Moped 5 = Bicycle 6 = School Bus 7 = Works Bus 8 = Ordinary (Service) Bus 9 = Taxi/Minicab 10 = Rail 11 = Underground 12 = Ferry 13 = Aeroplane 14 = Horse-riding 15 = Other 16 = Not stated
MAINMODE (J)	Main mode of journey. For multi-stage journeys, this is the mode from the stage with the greatest distance. See <i>section 3.2.5</i> for more details.	Same as MODE.

PURPOSE	Purpose of the journey	0 = not stated 1 = place of work 2 = in course of work 3 = educational establishment 4 = shopping 5 = visit hospital or other health 6 = other personal business 7 = visiting friends or relatives 8 = eating/drinking alone or at work 9 = eating/drinking other occasions 10 = entertainment/other public activities 11 = participating in sport 12 = coming/going on holiday 13 = day trip 14 = other not coded 21 = escort - home 22 = escort - work 23 = escort - at work 24 = escort - education 25 = escort - shops 26 = escort - personal 27 = escort - other 28 = go home 29 = just go for a walk
NUMOCC (S)	Number of occupants (if mode for the stage is by "car/van")	Numeric: "missing" if mode not "car/van"
STARTHR	Start hour of stage / journey (as appropriate)	0 – 24 (NB: Midnight's hour = 24)
STARTMIN	Start min of stage / journey (as appropriate)	0 - 59
ENDHR	End hour of stage / journey (as appropriate)	0 – 24
ENDMIN	End min of stage / journey (as appropriate)	0 - 59
DURATION	Duration in minutes – derived from start time and end time	Minutes
ORIGPLC	Numeric variable showing whether the origin is Home, Work or Other	1 = Home, 2 = Work, 3 = Other
DESTPLC	Numeric variable showing whether the destination is Home, Work or Other	1 = Home, 2 = Work, 3 = Other

ORIGC	Council area of origin	100 Aberdeen City 110 Aberdeenshire 120 Angus 130 Argyll & Bute 150 Clackmannanshire 170 Dumfries & Galloway 180 Dundee City 190 East Ayrshire 200 East Dunbartonshire 210 East Lothian 220 East Renfrewshire 230 Edinburgh, City of 235 Eilean Siar 240 Falkirk 250 Fife 260 Glasgow City 270 Highland 280 Inverclyde 290 Midlothian 300 Moray 310 North Ayrshire 320 North Lanarkshire 330 Orkney Islands 340 Perth & Kinross 350 Renfrewshire 355 Scottish Borders 360 Shetland Islands 370 South Ayrshire 380 South Lanarkshire 390 Stirling 395 West Dunbartonshire 400 West Lothian 500 Other UK “Missing” or “0” means the council area could not be derived.
DESTC	Council area of destination	Same as ORIGC.
RTPORIG	Regional Transport Partnership area of origin	10 Highlands and Islands 20 North East 30 Shetland 40 South East Scotland 50 South West Scotland 60 Strathclyde 70 Tayside and Central Scotland
RTPDEST	Regional Transport Partnership area of destination	Same as RTPORIG

OCODE	Indicator of quality of origin postcode	h = Home postcode, w = Work postcode, d = Definite postcode, n = Notional postcode See <i>section A.1.9</i>
DCODE	Indicator of quality of destination postcode	Same as OCODE.
CONGA (S)	Whether part of trip was delayed due to traffic congestion (if mode for the trip was by car/van)	1 = Yes 2 = No
CONGB (S)	Estimated time lost due to traffic congestion (if mode for the trip was by car/van)	Mins
CONGC (S)	Whether part of trip was delayed (if mode for trip was by bus/train)	1 = Yes 2 = No
CONGDA (S)	Whether (part of) bus/train journey was delayed by bus/train arriving late	1 = Yes 2 = No
CONGDB (S)	Whether (part of) bus/train journey was delayed by bus/train not turning up	1 = Yes 2 = No
CONGDC (S)	Whether (part of) bus/train journey was delayed by bus/train breaking down	1 = Yes 2 = No
CONGDD (S)	Whether (part of) bus/train journey was delayed by bus/train being involved in or being delayed by accident	1 = Yes 2 = No
CONGDE (S)	Whether (part of) bus/train journey was delayed by bus lane being blocked	1 = Yes 2 = No
CONGDF (S)	Whether (part of) bus/train journey was delayed by congestion on the roads	1 = Yes 2 = No
CONGDG (S)	Whether (part of) bus/train journey was delayed by assault on bus/train crew	1 = Yes 2 = No
CONGDH (S)	Whether (part of) bus/train journey was delayed by vandals damaging vehicle or track	1 = Yes 2 = No
CONGDI (S)	Whether (part of) bus/train journey was delayed by traffic lights/signals not working	1 = Yes 2 = No
CONGDJ (S)	Whether (part of) bus/train journey was delayed for other reasons	1 = Yes 2 = No
CONGDK (S)	Whether (part of) bus/train journey was delayed by bad weather	1 = Yes 2 = No
CONGDL (S)	Whether (part of) bus/train journey was delayed by a large number of passengers getting on or off	1 = Yes 2 = No

CONGDM (S)	Whether (part of) bus/train journey was delayed as bus/train full so took a long time to get people on or off	1 = Yes 2 = No
CONGDN (S)	Whether (part of) bus/train journey was delayed by bus passengers asking for directions about the route	1 = Yes 2 = No
CONGDO (S)	Whether (part of) bus/train journey was delayed as bus passengers needed change/without correct fare	1 = Yes 2 = No
CONGDP (S)	Whether (part of) bus/train journey was delayed by other reasons	1 = Yes 2 = No
CONGDQ (S)	Don't know why journey was delayed	1 = Yes 2 = No
CONGE (S)	Estimated time thought lost in bus/train journey	Mins
PAYA (S)	Whether paid for parking at the end of (part of) the journey	1 = Yes 2 = No
PAYB (S)	Where vehicle was parked	1 = In a commercial car park 3 = On the street in a space you pay for 4 = Paid for in a car park provided by employer 8 = Residential parking permit 9 = Other
PAYC (S)	How much paid for parking	Pence
PAYD_HR (S)	Length of time parked at parking place	Hours
PAYD_MN (S)	Length of time parked at parking place	Mins
DISTANCE (S)	Stage Distance (as the crow flies)	Km (with decimal places)
JOURDIST (J)	Journey Distance (as the crow flies)	Km (with decimal places)
SERDIST	Estimated distance travelled on all calls	Miles (with decimal places)
IMPUTED	Flag for imputed stages / journeys	Please see <i>section A</i> in the <i>Annex</i> for more details of the values of this variable.
IMPDIST	Flag for imputed distance	Please see <i>section A</i> in the <i>Annex</i> for more details of the values of this variable.

3.4 Number of records on the Travel Diary datasets

The following datasets are available from the Data Archive, with the number of records and number of variables (shown in brackets) given for each dataset:

	Journey	Stage	Home-Work Distances	Home-School Distances
1999/2000	57,045 (27)	61,913 (29)	13,497 (4)	5,737 (3)
2001	28,519 (27)	30,406 (29)	7,260 (4)	2,977 (3)
2002	26,944 (27)	28,812 (29)	6,799 (4)	3,110 (3)
2003	26,790 (29)	28,412 (51)	6,877 (4)	3,015 (3)
2004	27,122 (29)	28,881 (51)	7,334 (4)	3,083 (3)
2005	24,658 (31)	26,387 (58)	7,094 (5)	2,995 (4)
2006	25,215 (31)	27,177 (59)	7,157 (5)	2,966 (4)
2007	20,606 (29)	20,814 (57)	7,519 (5)	2,094 (4)
2008	20,449 (29)	20,637 (57)	7,054 (5)	2,179 (4)
2009	18,679 (29)	18,934 (57)	6,217 (5)	2,290 (4)
2010	16,296 (29)	16,552 (57)	6,147 (5)	2,154 (4)

3.5 Linking the Travel Diary to the “main” SHS data

Each household has a unique identifier called “UNIQID”. A combination of “UNIQID” and “TRIPNO” will identify any journey on the “Journey” dataset, and a combination of “UNIQID”, “TRIPNO” and “STAGE” will identify any stage on the “Stage” dataset. If the “main” SHS data has to be linked to the Travel Diary, the variable “UNIQID” should be used. Make sure the random adult’s information is selected for any analysis by personal characteristics (e.g. age, sex etc). The person number of the random adult is contained in the variable “RANDPEO” which is on the “main” SHS dataset. The User Guide for the “main” SHS dataset describes some variables which contain particular pieces of information about the random adult (e.g. the random adult’s age “RANDAGE”, sex “RANDSEX”, current situation “RANDECON”, etc).

3.6 Requesting specialised datasets

The postcodes of the origin and destination of most stages are recorded in the SHS Travel Diary. For reasons of confidentiality, postcodes are *not* available in the version of the Travel Diary at the UK Data Archive. Two types of dataset with more detailed geographical information may be made available in a limited number of special cases by the Scottish Government, they are:

- Postcode Sector information for origins and destinations, and;
- Self specified zones i.e. one would have to specify zones using a “shape” file which could be sent to the Scottish Government GIS to be matched to the Travel Diary. In order to ensure that the zones are of a sufficient size, each zone must contain a minimum number of sampled households, which the Scottish Government will calculate taking account of the sampling fraction(s) for the local authority(ies) concerned, and the number of years data to be provided.

In both cases the need for “anonymisation” of data means that only a few variables relating to the household and the random adult will be provided, and that there will be (e.g.) banding of age and income. If you wish to apply for a specialised dataset please contact the SHS Project Manager (*see section 7.1*) to discuss what might be available.

4. The collection and imputation of information about journeys

4.1 The Travel Diary collects information about journeys that were made on the day *before* (defined as midnight to midnight) the interview: so, someone interviewed on Sunday will be asked about the journeys he or she made on Saturday. Journeys which start on one day and finish on another should be counted on the basis of the day on which they *started*: so, if a person interviewed on Sunday went out on Friday evening and returned home in the early hours of Saturday, the journey home on Saturday should be recorded only if it started after midnight (because otherwise that would mean that it started on the day before the interview); and if the person went out on Saturday evening, the journey home should be reported provided that it started before midnight.

4.2 Interviews are *not* spread evenly across the week, because some types of people are more likely to be found at home, available for interview, on certain days. Therefore, the results need to be re-weighted using factors, which depend upon the day of the week and the adult's current situation (or economic status), so that, within each category of "current situation", the reweighted number of interviews is spread evenly across the days of the week. The reweighting process covers *all* interviews, including those with people who had *not* made any journeys on the day before the interview. Therefore, the reweighted numbers of people who said that they had made journeys, and the reweighted numbers of journeys themselves, are *not* necessarily evenly spread over the days of the week.

4.3 It may be felt that the SHS's results will be biased, tending to over-estimate the number of journeys, because the interviewer asks only about travel on the previous day: for example, people may be more likely to be interviewed on the days on which they made no journeys than on the days on which they made many journeys, since they are more likely to be available for interview on days on which they have not made any journeys. Therefore, the probability of being interviewed on a particular day depends, to some extent, upon the amount of travel on that day. It follows that the day for which the information about journeys is collected (the day before the interview) does not represent a "completely random" choice of day, and therefore that the Travel Diary results may not be properly representative.

4.4 In cases where it is obvious that journeys are missing, these have been imputed – e.g. if the only journey recorded for the day was to work at 8.00 a.m., a return journey was imputed, using the same mode of transport and with the same duration. The imputation process uses information about the time spent at the destination by other people with the same current situation (economic status) who had reported making both an outward journey and a return journey for the same purpose. The average times spent at the destination, and the distributions of such times, are used to impute the times at which the return journeys would start (unless the imputed time would be after midnight, in which case a return journey is not imputed). The Scottish Government also split what were recorded as "circular"/"round trip" journeys (such as a two-stage journey from A to B and then back to A) into separate outward and return journeys. Imputation in these, and other, cases has increased the total number of journeys in the database by about 4.8 %. However, it *cannot* compensate for the full extent of under-reporting of journeys, because there will be many "overlooked" journeys

that cannot possibly be imputed from the information that was recorded (e.g. in a case where two journeys were recorded - to work and back - one would not know if the person had forgotten to mention - say - mid-day journeys from work to some shops and back). More information on the methods of imputation used can be found in the *Annex* at the end of this document.

4.5 Because the imputation process does not compensate fully for the under-reporting of journeys in the SHS, and because the one day nature of the Travel Diary may over-estimate the number of journeys, the published results so far have not included any estimates of the averages of the total numbers of journeys, or of the total distance travelled, per person per year. Instead, information has been provided about the characteristics of the journeys which were recorded (such as the percentage made using each mode of transport), which should not be affected greatly by under-reporting (unless, for some reason, the journeys which were reported were markedly atypical of all journeys made by adults).

4.6 Initial examination of the data collected in the SHS's first few months of interviews identified a number of problems, as a result of which several improvements were made to the interviewers' computer systems, in order to collect better quality data. Further improvements have been made periodically (notably before the 2007 survey). However, the nature of the SHS is such that it cannot collect travel data of the same quality as are obtained by the GB National Travel Survey (NTS). There are a number of reasons for this:

- the SHS is a multi-purpose survey, which has transport as just one of its (from 1999 to 2006) three priority topic areas. The NTS is a specialised survey, designed to collect information about travel;
- the SHS collects information about one day's travel by one randomly-chosen adult member of the household. The NTS collects information about seven days' travel by every member of the household, including children;
- the SHS interviewee has to recall, off the top of the head, with no prior warning, details of journeys made on the previous day, during an interview on many topics. In the NTS, each member of the household is asked to record, using a special form, information about journeys which are made in a subsequent period of seven days;
- the SHS interviewer normally has only one contact with the randomly-chosen adult member of the household. The NTS interviewer makes contact on several occasions (to make an appointment, if necessary; before the seven day period for the travel diaries starts, to explain the procedures to the household; sometimes, a mid-"week" visit to remind and help the household; following the travel diary period, to pick up the completed forms; and, rarely, a telephone call or a further visit if there are outstanding queries);
- the SHS interviewers' computers carry out some simple checks on the information as it is collected (e.g. that the time recorded for the end of the journey is after the time at which it was recorded as starting). The NTS also has computer checks. In addition, at the "pick up" visits, the NTS interviewers check the credibility and completeness of the journeys recorded in the travel diaries, sometimes discovering that some journeys had not been recorded, and obtaining details of them. (Inevitably, such scrutiny of the data, and discussions with respondents about what they did and how they recorded it in the travel diary, will identify errors that could not be detected by any computer check on the details that are keyed in.).

Overall, therefore, the NTS data about travel are undoubtedly more comprehensive and of better quality than those of the SHS. However, the NTS's small sample size in Scotland

(1,620 households in 2009 and 2010 combined) means that it is not suitable for detailed analyses relating to Scotland, such as looking at the differences in travel patterns between different sub-groups of the population.

5. Origins and destinations, and distances travelled

5.1 *Origin and destination:* for each stage the interviewer asks where the person started from (origin), and where he or she went to (destination). "Home" and "work" can be recorded easily as people are more likely to know detailed address information of these locations (they are captured elsewhere in the survey); for other origins and destinations, the interviewer types in as much detail as possible of the address (e.g. the name of a shop, the street and the town). When appropriate, the interviewer can specify that the previous destination is the origin of the current stage / journey. The contractors determine the relevant postcodes – where they have not been provided or are incorrect - at a later stage in the processing of the data from the survey. In cases where only an approximate location is recorded (e.g. "centre of Edinburgh"), an arbitrary “notional” postcode (such as that of the main post office) is assigned. In some cases, the contractors may be unable to allocate a postcode, and can only provide a less "precise" indication of the location, such as a postal district (e.g. "EH10"). Inevitably, there are occasions on which the contractors cannot provide any indication of the location of the origin or the destination of a journey. Since the survey started, the interviewers' computer systems have been improved, to collect better "address" data.

5.2 *Duration:* The interviewer also records the times at which each stage of each journey started and the duration, from which the end time can be calculated. The recording process will only be accurate to - at best - say the nearest five minutes, for example because many people will not remember precisely, or correctly, the times at which some (or all) of their journeys on the previous day started and finished. Therefore, the estimated durations of some journeys will be subject to possibly large percentage errors.

5.3 *Distance travelled:* the length of any journey stage is the estimated distance “as the crow flies”, based upon the grid co-ordinates of the "centres" of the postcodes (or the most accurate area recorded) of the origin and destination of that stage of the journey. Therefore, the estimated distance would be zero in the case of a stage for which exactly the same postcode (or other type of area) was recorded (or derived) for both the origin and the destination. For example, if it was known (or deduced) only that the stage involved travel from (say) "EH10" to "EH10", the estimated distance would be zero. However, if it was known (or deduced) that the journey was from "EH10 6UD" to "EH10 6XE", the "crow flies" distance between the "centres" of the two postcodes would be calculated. Clearly, the percentage error in the estimation of distances will tend to be smaller for longer journeys - such as for a journey from "EH1" to "G1" compared with one from "EH10 6UD" to "EH6 6QQ". (Because the distances are estimated using grid co-ordinates recorded to the nearest metre, they may appear to be accurate to about a metre. However, it must be remembered that the grid co-ordinates relate to the "centres" of the areas concerned, and a journey may start or finish some distance from the "centre" of the recorded area, so the estimated distances are not as precise as it may appear from the number of decimal places).

5.4 In cases where the interviewer could not obtain sufficient details of the origin and/or destination to enable the contractors to assign a postcode (or other type of area), the distance travelled has been imputed. The imputation process uses information about the recorded time

taken for the trip and an imputed speed for the trip. The imputed speed is based upon information about the speeds of other trips made by the same mode of transport for the same purpose by people who live in the same type of area (in terms of the "urban"/"rural" category), distinguishing between journeys which start in what may be described as "peak" periods (for this purpose, taken as 7.00 a.m. to 9.30 a.m. and 4.30 p.m. to 6.30 p.m.) and at other times of the day. Both the average speed and the distribution of speeds for each mode / purpose / area / time category were used to impute the speed for each journey for which this was required. The imputed speed was then multiplied by the recorded journey time to calculate the imputed distance. Overall for 2009/2010, the distance was imputed for about 17.8% of all journey stages.

5.5 The distance of a multi-stage journey was calculated by adding up the distances of each of its component stages. For series of calls journeys, as information is only collected about the first and last calls, there are no estimates of distance for all the intermediate stages, and therefore the total distance is underestimated. In addition, as most journeys are not made in a straight line, the "crow flies" distance will underestimate the actual distance travelled.

6. “Home to Work” and “Home to School” distances

6.1 The SHS interviewer asks the randomly chosen adult for their work address and work postcode (or for as much of it as they know), and asks the Highest Income Householder or his/her spouse for the name and LA of the randomly chosen school child’s school. This information is then used by the contractors to derive the work postcodes where they were not complete (i.e. only the address was provided), and the Scottish Government uses information collected about the randomly chosen school child’s school, to obtain the schools postcode. The straight line “as the crow files” distance between “home and work” and “home and school” is calculated by the Scottish Government using grid references. Please see *section 5* above for more details on how distance is calculated.

6.2 Imputation was not carried out for the small percentages of cases where distance was “missing” for “home to work” or “home to school”, because it was not felt worthwhile to develop imputation methods for small numbers of “missing” cases. Therefore, a “Home to Work” distance record has only been provided where there were “Home” and “Work” postcodes (and the random adult was not “unemployed” or did *not* “work from home”). Similarly, “Home to School” distance records have only been provided where there were both “Home” and “School” postcodes.

6.3 The “home and work” and “home and school” distances are available as separate files called “dswk09_10da” (containing the unique identifier “UNIQID”, and the distance between home and work) and “homesc09_10da” (containing “UNIQID”, and the distance between home and school) respectively.

7. Enquiries and further information

7.1 General enquiries about the SHS should be addressed to the survey's Project Manager:

SHS Project Manager
 Scottish Government
 Area 1-F (Dockside)
 Victoria Quay
 Edinburgh, EH6 6QQ

Tel: 0131 244 0824
 Fax: 0131 244 7573
 E-mail: shs@scotland.gsi.gov.uk

7.2 Further information about the survey can be found on the SHS website at:
<http://www.scotland.gov.uk/shs>

This website provides some background to the survey, information about the progress of the survey, and the published results.

7.3 Enquiries about the SHS Travel Diary data should be addressed to:

Transport Statistics
 Transport Scotland
 Victoria Quay
 Edinburgh, EH6 6QQ

Tel: 0131 244 1457
 Fax: 0131 244 0871
 E-mail: transtat@transportscotland.gsi.gov.uk

Further information on Transport Statistics can be found on the Transport Scotland website at: <http://www.transportscotland.gov.uk/analysis/statistics>

7.4 Published results from the SHS Travel Diary, are available in the Scottish Government Statistical bulletins They are available at the Stationery Office bookshop, or at the following website www.scotland.gov.uk under the topics of 'Statistics' and 'Transport and Travel' (Please note the Scottish Government has carried out subsequent data cleaning since the first bulletin was published in February 2002. Therefore, results using the data from the Data Archive might differ slightly from those published).

7.5 Published results from the SHS travel diary for the 2009/2010 SHS are available from Transport Scotland's Statistical Bulletin "Scottish Household Survey: Travel Diary 2009/2010", published November 2011 at the following website: <http://www.transportscotland.gov.uk/analysis/statistics/publications/shs-travel-diary-results-previous-editions>

7.6 Please contact the Project Manager if you wish to be added to an *e-mail mailing list* to be kept informed of any significant updates to the information on the SHS website. The Project Manager will also, on request, distribute paper copies of information about the survey, and about significant developments when they occur, to people who are unable to access the website.

ANNEX

A. Details of the problems found and the imputation methods used

A.1 Investigation of the Travel Diary revealed the following problems:

A.1.1 Mode Duration Hours and Minutes are equal

Journeys with the same Hours and Minutes for the duration of their Mode of transport. (eg. Hours = 15 and Minutes =15) - this has been recorded as the time of day instead of the duration of 15 minutes only.

A flowchart was developed by the Scottish Government to get around this problem, based on various scenarios within the data. These are detailed as follows (in the order they should be applied):

- if Purpose is 'Place of Work' or 'In Course of Work' and Start time > 9.am then use Minutes
- if Mode is 'Aeroplane' then use Hours
- if Mode is 'Taxi' then use Minutes
- if Mode is 'Driver Car/Van' and hours > 6 then use Minutes
- if Mode is 'Walking' and Purpose is 'Participating in Sport' or 'Day Trip' or 'Just go for a Walk' and hours > 6 then use Minutes
- if Mode is 'Walking' or 'Bicycle' and Purpose is 'Place of Work' or 'In Course of Work' and hours > 1 then use Minutes
- if Mode is 'Walking' or 'Bicycle' and Purpose is 'Place of Work' or 'In Course of Work' and hours = 1 then use Hours
- if Mode is 'Walking' and hours > 2 then use Minutes
- if Mode is 'Walking' and hours are 1 or 2 then use Hours
- if Purpose is 'Shopping' and hours > 2 then use Minutes
- if a Return Journey and the Mode is the same as the Outward Journey and Mode then use Minutes from Outward Journey
- if hours = 0 and minutes =0 and the Destination is known then Impute using the Imputation programs

A.1.2 Pre-Imputation Data Edits

As agreed with The Scottish Government, a Pre-Imputed Data file is produced as a 'check' of the data prior to the Imputation stage of the analysis. Following a change to the questionnaire from Q1 2007 with regards to allowing Journeys of less than 15 minutes Duration and Interviewer error, further editing has been introduced.

Data Edits were developed based on different scenarios. Once the Edits are applied, it is 'flagged' for The Scottish Government to check that the 'corrections' were applied correctly and either accept or reject them. If they rejected any cases, the data is modified, or if the data was erroneously deleted, added back in.

Scenarios detailed as follows (in the order they are applied):-

- **No Information To Process, Then Delete** – there is no Start or End Time even though there is Journey data, so the Journey would be deleted
- **Zero Durations** – if the Duration was zero, then the Journey would be deleted
- **Stages Recorded As Journeys** – firstly get a chronological chain of at least two Journeys.

If the Duration between the two Journeys is less than or equal to 5 minutes and the 1st Journey Origin Address was not equal to the 1st Journey Destination address and the 2nd Journey Origin Address was not equal to the 2nd Journey Destination address and the Purpose of the two Journeys was equal, then delete one and flag' the data for checking.

If the Duration between the two Journeys is less than or equal to 5 minutes and the 1st Journey Origin Address was not equal to the 1st Journey Destination address and the 2nd Journey Origin Address was not equal to the 2nd Journey Destination address and the Purpose of either Journey was 'Just Go For A Walk', then delete one and 'flag' the data for checking.

There is also a check that none of the Multi-Stage Journeys have Walking Stages within them.

- **Precise Duplicates** – within the same Uniqid (Respondent no.), two Journeys have the same Mode, Purpose, Start Time and Duration. If this is the case, delete one. If they have a Return Journey, delete the Return Journey as well.
- **Imprecise Duplicates** - within the same Uniqid (Respondent no.), two Journeys have the same Start Time, regardless of the Mode, Purpose or Duration. If this is the case, delete one, If they have a Return Journey, delete the Return Journey as well.
- **Accidental Returns** – if the Start Time of a Return Journey equals the Start Time of the Outward Journey then do:-
If the Mode is the same, then drop the Single Journey and keep the Return.
If the Mode is different, then drop the Return Journey and replace the Purpose of the Original Journey with the Return Journey Purpose.

From Q3 2009 onwards, some minor edits were carried out on the pre-imputation data to prevent erroneous imputations from occurring.

A.1.3 Non chronological order of journeys

In some cases, respondents do not report journeys in chronological order. This was corrected by sorting the data in order of the start time of each journey and deriving a new variable called "TRIPNO".

A.1.4 Similar consecutive single stage journeys at different times

A number of similar consecutive journeys at different times were found (*about 0.1% of journeys*). In such cases, both journeys had the same origin, and both had the same destination e.g. "Home to Work" in the morning and "Home to Work" in the evening with nothing recorded in between.

It was decided to "correct" only consecutive "single stage" journeys. The correction involved swapping the origin and destination of a journey in cases where the origin of the first journey matched the origin of the second journey and the destination of the first journey matched the destination of the second. But which journey should be "swapped" round?

Example A

Journey 1 Home to work from 8 am to 8.30am

Journey 2 Home to work from 5pm to 5.30pm

It would be reasonable to “correct” the second journey to become “Work” to “Home” from 5pm to 5.30pm.

But in the following case:

Example B

Journey 1 Work to Home from 8am to 8.30am

Journey 2 Work to Home from 5pm to 5.30pm

It would perhaps make more sense to “swap” the origin and destination of the first journey.

What should be done could also depend on what the person’s current situation is. The above “corrections” make sense if the person works in a “normal” full time job, but, it might not be the case if they are, say, a part time cleaner. Therefore, it was decided to look at “swapping” only in cases where the “swapped” details would be more likely to be correct in general. This meant that only those cases for which the origin was “Home” or the destination was “Home”, as these were the only ones for which one can easily apply a “general rule” – e.g. in general one might expect the first journey of the day to start at home. So, if the first journey of the pair is the first journey for the person and it starts at “Home” the origin and destination of the second journey of the pair were “swapped”. Or, if the last journey of the pair is the last for the person, and the journey ends at “Home” then the origin and destination of the previous journey of the pair were “swapped”.

The “IMPUTED” flag is set to “10” for these cases.

A.1.5 Only one journey made

A large number of people were found to have made only one journey (e.g.) a journey in the morning from “Home” to “Work”, with nothing else recorded for the rest of the day (*about 0.7% excluding series of calls journeys and those journeys for the purpose of coming back from / going on holiday*). Series of calls journeys were excluded from this method of imputation, as they could be genuine. Circular journeys were also excluded as these were dealt with separately.

It was decided to only impute a reverse journey in cases where either the origin or destination of the only recorded journey was “Home”, as imputing a reverse journey in other cases would probably not be correct e.g. suppose the only recorded journey was a single stage journey from the “shops” to “a friend’s house”. If a return journey was imputed then this person would appear to spend the rest of the day (until after 12 midnight) at the shops!

Examples of the method used are as follows:

Case 1) **Two journeys both** with origin = “Home” and destination = “anywhere else”

Case 2) **Two journeys both** with origin = “anywhere else” and destination = “Home”

In case 1) a reverse journey would be generated x hours later, only if the start time was *not* after 23.59.

In case 2) a previous journey would be generated x hours earlier, only if the start time was *after* 00.00.

For example, for only one “single” stage journey with origin “Home”, the imputation method works as follows:

- a reverse journey would be generated
- the “purpose”, “mode” and “distance” would be the same
- the origin and destination information would be “swapped”
- start time would depend on the end time of the journey from “Home”, the “purpose” of the journey and the person’s “current situation” (e.g. journey “to work” and the person was in “full-time employment”)
- the “end time” would be calculated by working out the duration of the first journey and adding it to the “start time”
- the journey would not be imputed if it started after 23.59

In order to impute the start time of the reverse journey, an estimate of how long the person might have spent at “the place” was derived. This was done by working out the mean and standard deviation of the time spent at “the place” for all the combinations of “purpose” and “current situation” for those people who had made *two* journeys in the day: one from “Home” and one to “Home”. The imputed value was then taken as a random variate from a normal distribution with the relevant mean and standard deviation. However, if there were fewer than ten “donor cases” valid journeys for that combination of “purpose” / “current situation” (e.g. “purpose” of “to work” and “current situation” of “full-time employment”) then the relevant mean was used as the imputed time spent at “the place”.

The “IMPUTED” flag is set to “21” where the reverse journey of a single stage journey has been imputed and “IMPUTED” is set to “22” where the reverse journey of a multi stage journey has been imputed.

A.1.6 Circular journeys

A “circular” journey is one for which the origin and the destination are the same e.g. from “Work” to “Work”. A number of circular journeys were found to have been recorded (*about* 7%). They may be the result of (e.g.) two journeys mistakenly entered as one journey, or perhaps the person took a dog for a walk (i.e. the journey had no “obvious” real destination). (Circular series of calls journeys were excluded from this method of imputation, as they may well be genuine e.g. it would be perfectly valid to record a “series of calls” journey from “Home” to calls at several shopping centres and then back to “Home”).

The imputation method works as follows:

- Return Circular journeys are selected from everybody, regardless of the number of stages or Purpose. Then the outward journey is deleted. (“IMPUTED” is set to “31.5” in these cases)
- Circular journeys with only one stage were split into two journeys: one to an unknown destination, and the other the trip back. (It is assumed that the person spends a minute at the unknown destination). (“IMPUTED” is set to “31” in these cases)
- Circular journeys with two stages were split into two journeys with one stage each. (“IMPUTED” is set to “32” in these cases)

- Circular journeys with three stages were split into two journeys. How the stages were split between the two new journeys depended on the “time gaps” between the stages. If the gap between stages 1 and 2 was longer than the gap between stages 2 and 3 the journey was split so that stage 1 became a single journey and stages 2 and 3 became stages 1 and 2 of a new journey. Otherwise they were split so that stages 1 and 2 form the first journey and stage 3 became a new single stage journey. (“IMPUTED” is set to “33” in these cases).
- Circular journeys with 4 or more stages were not included in the imputation process as the number of these journeys was small.

A.1.7 Apparently incomplete sequences of journeys

Investigation also revealed some apparently “incomplete” sequences of journeys, e.g. in cases where a journey 1 was from A to B and journey 2 was from A to C so it appears that there is a missing journey from B to A, (*about 1.4%*).

It is reasonable to impute journeys to fill the gaps where the origin of the previous journey matches the origin of the next journey e.g.

Journey 1 A to B

Journey 2 A to C

And where the destination of the previous journey is the same as the destination of the next journey e.g.

Journey 1 B to A

Journey 2 C to A

Nothing was done in cases where journey 1 was from A to B, and journey 2 from C to D with a “gap” between B and C, because it would have been too complicated to try to impute a journey to fill the “gap” as there is no information in either of the recorded journeys with which to impute the apparently unreported journey (and B and C might be very close together).

For simplicity, journeys were only imputed where there were single stage journeys on either side of the gap, neither of which was non-circular, and they had either

- (a) the same origin e.g. A to B then A to C – when “B to A” was imputed to fill the gap, or
- (b) the same destination e.g. A to B then C to B – when “B to C” was imputed to fill the gap.

The methodology used was like that described in *section A.1.5* above, except that the mean and standard deviation of the combinations of “purpose” and “current situations” are used for *all* valid journeys.

The “IMPUTED” flag is set to “40” for these cases.

A.1.8 No return journey to “Home” at the end of the day

A large number of respondents’ journeys were found not to end up at home at the end of the day (*about 5.8%*). Therefore, a return journey has been imputed to home where the last stage is “Home” to “somewhere” – excluding those cases where the final recorded journey had the “purpose” of “coming back from/going on holiday”, series of calls journeys, and cases where

the imputed journey would start after mid-night (e.g. a night shift worker's return journey to home would start after midnight, so would not be imputed). For simplicity, return journeys to home were only imputed for those who made single stage journeys (*about 1.1%*).

The "IMPUTED" flag is set to "50" for these cases.

A.1.9 Missing or suspect distances

As explained in *section 5* above, distance is calculated from grid references obtained from the postcodes of the origins and destinations of the stages. Missing or incomplete origin and destination postcodes led to missing distance information for about 2.1% of stages. The following tables give an indication of the quality of the postcodes recorded. (Note the results are from the "Stage" file for 2009 and 2010 *prior* to imputation and that some of the postcodes that were recorded or derived were not complete)

Origin postcode	Number of cases	Percentage of cases
Home	17,393	52.0 %
Work	3,536	10.6 %
Definite (not "Home" or "Work")	7,449	22.3 %
Notional (usually the nearest post office)	5,047	15.1 %
"Missing"	392	1.16 %

Destination postcode	Number of cases	Percentage of cases
Home	15,898	47.5 %
Work	4,210	12.6 %
Definite (not "Home" or "Work")	7,939	23.7 %
Notional (usually the nearest post office)	5,448	16.3 %
"Missing"	322	0.95 %

The methodology for imputing distances was similar to that described in *sections A.1.5 and A.1.7* above. A speed was calculated for each stage, for which a distance had been estimated. It will only be approximate because of the imprecision of the estimates of the distance and the journey time (calculated as the difference between the "start time" and the "end time"). The mean and standard deviation of the speeds were then calculated for the Main Mode of Transport. In cases where the distance was "missing", or the calculated speed was "suspect" the speed for the stage was imputed. The following "suspect" speeds were replaced by imputed values:

- Walking speeds greater than 20km/hr
- Driver/passenger car/van, motorcycle/moped, taxi/minicab speeds greater than 150km/hr
- Bus, underground, horse riding, ferry, other speeds greater than 100km/hr
- Aeroplane speeds greater than 1000km/hr.

NB: these thresholds are "high" because of the imprecision of the estimated distance and the estimated time spent travelling (since both the "start time" and "end time" may be rough estimates).

Imputed distance was then calculated as $\text{Distance} = \text{Imputed Speed} * \text{Travelling Time}$. Journey distance (as opposed to stage distance) for multi-stage journeys is the sum of the distances for the individual stages.

If distance has been imputed the "IMPDIST" flag is set to "10".

A2. Flags for imputed stages and journeys

The table in *section A.3* below summaries the “IMPUTED” and “IMPDIST” flags used in the Travel Diary “Stage” and “Journey” datasets. The coding used for multi-stage journeys can be complicated. If there is an entry in the “IMPUTED” or “IMPDIST” variables for a multi stage journey on the “Journey” dataset this has to indicate whether information has been imputed for the first and/or last stage of the journey. When a four digit code is used, the first two digits relate to the first stage, and the last two digits relate to the last stage. Some examples of this are:

- Suppose the only journey recorded for the day was a multi stage: stage 1: A to B and stage 2: B to C.
 - A reverse journey would be imputed, Journey 2: stage 1 C to B, stage 2 B to A.
 - The “IMPUTED” flag on the “Stage” dataset would be set to “22” for both stages of the imputed reverse journey, and on the “Journey” dataset “IMPUTED” would be set to “2222”.
- Suppose the only journey recorded for the day was a multi-stage journey which had three or more stages then the “IMPUTED” flag on the “Journey” dataset would only indicate if the first or last stage of the journey had been imputed, (e.g.) stage 1 A to B, stage 2 B to C, and stage 3 C to D.
 - A reverse journey would be imputed, Journey 2 stage 1 D to C , stage 2 C to B, and stage 3 B to A.
 - The “IMPUTED” flag on the “Stage” dataset would be set to “22” for all stages of the imputed reverse journey, and on the “Journey” dataset “IMPUTED” would be set to “2222” indicating that the first and last stages of the journey had been imputed.
- If the distance was imputed for the first stage and the last stage of a multi-stage journey the “IMPDIST” flag on the “Journey” dataset would be set to “1010”.

A.3. Number of imputed records generated

The following tables show the number of imputed journeys and stages generated by the imputation process for each “problem” mentioned above. They also show the resulting values of the “IMPUTED” and “IMPDIST” flags.

“Flags” on the “Stage” dataset for 2009 and 2010

Name of problem	Name of “flag”	Value of “flag”	Number of stages in 2009/2010	Percentage of stages in 2009/2010
Similar consecutive single stage journeys at different times	IMPUTED	10	51	0.14
Only one single stage journey made	IMPUTED	21	627	1.77
Only one multi stage journey made	IMPUTED	22	18	0.05
Circular journey: delete return	IMPUTED	31.5	172	0.48
Circular journey: one stage	IMPUTED	31	1,815	5.11
Circular journey: two stages	IMPUTED	32	43	0.12
Circular journey: three stages	IMPUTED	33	0	0
Apparently incomplete sequences	IMPUTED	40	492	1.39

of journeys				
No return journey to “Home” at the end of the day	IMPUTED	50	373	1.05
Missing or suspect distances	IMPDIST	10	6,334	17.85

“Flags” on the “Journey” dataset for 2009 and 2010

Name of problem	Name of “flag”	Value of “flag”	Number of journeys in 2009/2010	Percentage of journeys in 2009/2010
Similar consecutive single stage journeys at different times	IMPUTED	10	51	0.16
Only one single stage journey made	IMPUTED	21	627	1.79
Only one multi stage journey made	IMPUTED	22	1	0.002
Only one multi stage journey made (first and last stage imputed)	IMPUTED	2222	15	0.04
Circular journey: delete return	IMPUTED	31.5	172	0.21
Circular journey: one stage	IMPUTED	31	1,815	5.19
Circular journey: two stages	IMPUTED	32	40	0.11
Circular journey: two stages (first and last stage imputed)	IMPUTED	3232	1	0.002
Circular journey: three stages	IMPUTED	33	0	0
Circular journey: three stages (first and last stage imputed)	IMPUTED	3333	0	0
Apparently incomplete sequences of journeys	IMPUTED	40	492	1.41
No return journey to “Home” at the end of the day	IMPUTED	50	373	1.07
Missing or suspect distances	IMPDIST	10	6,124	17.51
Missing or suspect distances (first stage distance was imputed)	IMPDIST	1000	53	0.15
Missing or suspect distances (first and last stage distance was imputed)	IMPDIST	1010	58	0.17

B. Numbers of records before and after imputation*Based on 2009/2010 data (unweighted numbers)***B.1 Number of records**

	before	after	change
Stage Dataset	33,817	35,486	1669
Journey Dataset	33,281	34,975	1694

B.2 Number of stages by type of journey

Type of journey	before	after	change
Single stage	31,661	33,380	1719
Multi-stage	959	909	-50
Series of calls	1,197	1,197	0

B.3 Unweighted frequencies of “mode” and “purpose”

Mode	Stages			% of all stages		
	Before	after	change	before	after	diff
Not Recorded	0	0	0	0	0	0.00
Walking	7,647	7,953	306	22.61	22.41	-0.20
Driver Car/Van	17,153	18,011	858	50.72	50.76	0.04
Passenger Car/Van	4,389	4,622	233	12.98	13.02	0.04
Motorcycle/ Moped	57	62	5	0.17	0.17	0.00
Bicycle	280	301	21	0.83	0.85	0.02
School Bus	56	63	7	0.17	0.18	0.01
Works Bus	52	53	1	0.15	0.15	0.00
Ordinary (Service) Bus	2,889	3,053	164	8.54	8.60	0.06
Taxi/Minicab	406	429	23	1.20	1.21	0.01
Rail	479	502	23	1.42	1.41	-0.01
Underground	58	62	4	0.17	0.17	0.00
Ferry	70	73	3	0.21	0.21	0.00
Aeroplane	59	64	5	0.17	0.18	0.01
Horse-riding	0	0	0	0.00	0.00	0.00
Other	222	238	16	0.66	0.67	0.01

Purpose	stages			% of all stages		
	before	after	change	before	after	diff
not stated	0	0	0	0.00	0.00	0.00
place of work	8,006	8,354	348	23.67	23.54	-0.13
in course of work	426	439	13	1.26	1.24	-0.02
educational establishment	892	944	52	2.64	2.66	0.02
shopping	8,112	8,398	286	23.99	23.67	-0.32
visit hospital or other health	1,048	1,095	47	3.10	3.09	-0.01
other personal business	2,479	2,612	133	7.33	7.36	0.03
visiting friends or relatives	3,854	4,067	213	11.40	11.46	0.06
eating/drinking alone or at work	155	168	13	0.46	0.47	0.01
eating/drinking other occasions	1,070	1,132	62	3.16	3.19	0.03
entertainment/other public activities	1,034	1,103	69	3.06	3.11	0.05
participating in sport	1,315	1,401	86	3.89	3.95	0.06
coming/going on holiday	46	46	0	0.14	0.13	-0.01
day trip	698	748	50	2.06	2.11	0.05
other not coded	167	182	15	0.49	0.51	0.02
escort - home	240	254	14	0.71	0.72	0.01
escort – work	223	236	13	0.66	0.67	0.01
escort – at work	24	26	2	0.07	0.07	0.00
escort – education	1,067	1,125	58	3.16	3.17	0.01
escort – shops	57	60	3	0.17	0.17	0.00
escort – personal	362	375	13	1.07	1.06	-0.01
escort – other	380	400	20	1.12	1.13	0.01
go home	919	1,092	173	2.72	3.08	0.36
just go for a walk	1,243	1,229	-14	3.68	3.46	-0.22