Department for Transport

National Travel Survey
Data Extract User Guide, 1995-2014

Contents

Data Update

2002-2014 dataset supersedes the 2002-2012 (SN: 5340 and SN: 7553) files:

Data for 2013 and 2014 have been processed and added to the previous data extract to allow users to analyse NTS data for all available years.

1995-2001 dataset supersedes the 1995-2001 (SN: 6108) files:

The National Travel Survey (NTS) team in DfT recently completed the redevelopment of the database in which all NTS data is held. All historic data (1995-2001) was copied from a Quantum Quanvert system into a new SQL database. This has over written previous data for years 1995 to 2001 which was previously processed in Quantum and supplied to the UK Data Service.

This change means that variable names in both the 1995-2001 and 2002-2014 files are corresponding and as such it is possible for users to join the datasets. (Note that the files contain different variables in some cases and as such consistent analysis is not always possible – see Table Structures document for more information on the availability of variables) All files can now be joined using the unique ID fields.

In addition, as part of the database redevelopment all processing methods were reviewed, including imputation of missing data for some variables. Improvements and corrections have been made which has resulted in some minor revisions to all data published from 1995. See Revision section of this document for details of the main changes made which have affected time series results.

Coverage:

In 2013, the coverage of the NTS changed to sample residents of England only. This change was agreed following a public consultation in 2011. Details of the consultation outcome can be found at:
https://www.gov.uk/government/uploads/system/uploads/attachment data/file/230560/NTSconsultationSummaryofresponses.pdf

Due to this change analysis for Great Britain is not possible from 2013 onwards. It is recommended that analysis is undertaken for households/individuals who are English residents only. For more information on other travel surveys within Great Britain see the Background information in the 2014 National Travel Survey publication:
https://www.gov.uk/government/statistics/national-travel-survey-2014

All variable names have changed:

For users who are familiar with the old Quantum variable names, the Table Structures spreadsheet includes a mapping of old to new variable names. The data file containing all trip records is now called Trip.sav; in previous versions it was named Journey.sav.

A significant number of extra variables have been supplied to the UK Data Service for the End User Licence dataset. Further additional variables are available to users via Special Licence and the Secure Access Portal. The Table Structures spreadsheet indicates which variables are available and in which survey years – in this document 'Full' represents DfT's own version of the database.

Key issues to consider when analysing NTS data

- The NTS is primarily designed to measure long-term trends in travel and is **not suitable for monitoring short-term trends** or year-on-year changes.

- **Careful attention should be paid to sample sizes** for all analyses. (See note on sample sizes in the 2014 NTS technical report)

- **Appropriate 'short walk' weights should be applied to all trip/stage-level analysis** to account for the fact that short walks are only recorded on the seventh day of the travel week. (See section on 'Short walk weighting')

- In addition, there is a weighting strategy to adjust for non-response and 'drop-off', **appropriate weights should be applied to all analyses of data**. (See section on 'Non-response and "drop-off" weighting')

- **All analyses of data from the travel diary should be based on fully co-operating households only.** Analysis at the individual, household and vehicle level should be based on the interview sample which includes data from additional 'partially co-operating' households. (See section on 'Weighting')

- Before carrying out any analysis, all users should **first try to reproduce the examples in the annexes of this document and some published results tables** to ensure they are using the data and applying the weights correctly. The 2014 NTS publication is available at: https://www.gov.uk/government/publications/national-travel-survey-2014

- In addition, **users should be familiar with the survey methodology** to ensure data is analysed and interpreted correctly. See details below and the NTS technical information, available at: https://www.gov.uk/government/collections/national-travel-survey-statistics

- When analysing the data users should be aware that, although the core survey has remained consistent over time, **some relatively minor changes are made to the survey each year**, e.g. addition/removal of questions, changes to filters and variable categories. The appendix of the annual Technical Reports (see link above) give details of questionnaire changes since 2002.

For advice on using and interpreting the NTS data, please contact: national.travelsurvey@dft.gsi.gov.uk

Survey Overview

Survey background

The NTS is carried out in order to monitor long-term changes in personal travel and provide a better understanding of the use of transport facilities made by different sectors of the population. Individuals in sampled households are interviewed face-to-face to collect personal information, such as age, gender, working status, car access and driving licence holding. They are also asked to complete a seven day travel diary and provide details of trips undertaken, including purpose, method of travel, time of day and trip length.

The first NTS survey was commissioned by the Ministry of Transport in 1965 with further periodic surveys being carried out in 1972/73, 1975/76, 1978/79 and 1985/86. There are a number of differences between the current survey and surveys up to 1978/79. These are detailed in Chapter 7 of the National Travel Survey 2008 Technical Report, which is available on the archived NTS website at:
http://webarchive.nationalarchives.gov.uk/20091203140650/http://www.dft.gov.uk/pgr/statistics/datatablespublications/personal/methodology/ntstechreports/ntstechrep2008
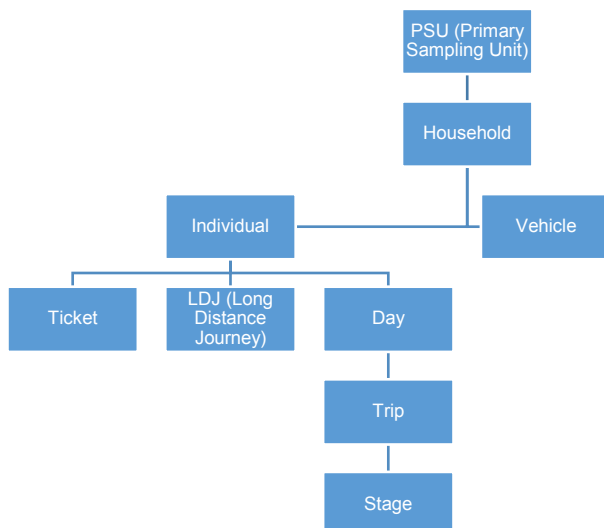
Conducting the survey

Prior to the interviewer's first call, letters are sent out to the sampled addresses. These introduce the survey and explain that an interviewer will call. The interviewer then arranges an appointment for a 'placement interview'. During this visit, the interviewer obtains information on the household, each household member and on all vehicles to which the household has access, via a computer assisted personal interview (CAPI). The procedures for the seven day travel diary record are then explained. Each household is given a randomly assigned start date for the seven day travel diary.

The placement call is generally followed by a reminder call, just before the start of the travel week, to remind the household to begin their travel records, and by a midweek call during the travel week to check that records are being completed correctly. Within six days of the end of the travel week the interviewer will make a 'pick-up call' to collect the travel records and to check the information recorded with the informants.

Data types

The data is hierarchical in nature and consists of several record types or 'levels'. Most significantly, data is held about households, vehicles, individuals, trips and stages. Data from each level are linked to make cross-level analysis possible. NTS methodology determines that vehicles belong to households and not individuals. Vehicle data, therefore, can only be analysed by household or vehicle attributes.

Levels in the NTS database

PSU (Primary Sampling Unit)

Household

Individual

Vehicle

Ticket

LDJ (Long Distance Journey)

Day

Trip

Stage

PSU – Primary Sampling Units (PSU) are used to provide a list of areas from which to select a sample of addresses from. In this case each PSU is a postcode sector.

Key identifier variables

To enable the linking of data from each level of the dataset hierarchy, a set of identifier variables are attached to each data file.

The combination of identifier variables that uniquely identifies a record at each level of the database are:

PSU	-	PSUID
Household	-	HouseholdID
Vehicle	-	VehicleID
Individual	-	IndividualID
LDJ	-	LDJID
Day	-	DayID
Trip	-	TripID
Stage	-	StageID
Ticket	-	IndTicketID

To link levels together the identifier variables from the highest analysis level should match those at the lower level. These identifier variables are new unique IDs which allow for linkage between data levels (without requiring linkage on multiple variables, as in previous versions of the datasets held by the UK Data Service).

Example:

To link an individual to the trips they made, merge the individual and trip level files and ensure that the individual level identifier, IndividualID, matches on both files:

Individual.IndividualID=Trip.IndividualID

If merging datasets together, users should remove unwanted variables to reduce the file sizes and length of time it takes to merge.

See examples in Annex A and B for detailed instructions of how to merge data files.

Using NTS Data

Distance

The length of any trip stage is the distance actually covered, as reported by the traveller, and not the distance as the crow flies. Distance is measured in miles (previous versions of the datasets held by the UK Data Service used tenth of mile).

Distance by mode

Average distance travelled results are calculated based on the *stage distance* for each mode and are presented as 'average distance travelled by 'mode' – see table NTS0305. Trip rates are usually presented as 'main mode', i.e. based on the mode that is the longest stage distance – see table NTS0303.

Time

Unit = minutes.

Denominator

It is standard NTS methodology to include all respondents in the denominator not just those making trips. For example children should be included when calculating driver trip rates.

Number of records

There are less records in 2011 due to the GPS experiment which was conducted using a sub sample and data collected was not incorporated into the main database.

Mode: Taxi/minicab

When using main mode or mode variables the split between taxi and minicab should not be used as it is not accurate. Results should be presented combined like published tables, e.g. NTS0303

Mode: Walking

Short walks (those under one mile) are only collected on day 7 of the travel diary. 'Short walk weights' account for this and should be applied when producing analysis – see 'Weighting' section below. For detailed walking trip analysis it may be necessary to limit the base sample to day 7 only data, e.g. NTS9909.

Key Definitions

Trip

A **trip** is defined as a one-way course of travel having a single main purpose, e.g. a walk to school or a trip to work without any break in travel.

Usually the respondent is clear what the single main purpose of a particular trip is. However, sometimes people go out for a number of reasons, or go out for one main reason but carry out a number of different activities, perhaps at different places. Complex travel like this is broken into separate trips so that the data can be analysed. Where a stop is entirely secondary to the main purpose (such as a stop to buy a newspaper on the way to work), the stop is disregarded.

Stage

A trip 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.

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 are treated as one continuous 'series of calls' trip from the first such call to the last one. Only shopping and 'in course of work' travel can be treated in this way. A doctor's round could therefore consist of one trip to the first patient, one series of calls trip to the other patients and one trip from the last call back to the surgery or home.

A full list of definitions is available in the Notes and Definitions document available at:
https://www.gov.uk/government/collections/national-travel-survey-statistics

Weighting

Short walk weighting

Because walking trips of less than one mile in distance are recorded only on the seventh day of the travel week, these trips must be weighted by a factor of seven when analysed. Also, for consistency with earlier surveys, 'series of calls' trips are excluded from analysis of stage and trip counts and time. Therefore, one of the following 'short walk weights' must be applied to any tabulations using trip or stage counts, distance or time:

SSXSC	Number of stages to be counted, grossed for short walks and excluding 'series of calls' trips.
SD	Stage distance travelled, grossed for short walks.
STTXSC	Stage travelling time grossed for short walks and excluding 'series of calls' trips.
JJXSC	Number of trips to be counted, grossed for short walks and excluding 'series of calls' trips.
JD	Trip distance travelled, grossed for short walks.
JOTXSC	Overall trip time (includes travelling and waiting time), grossed for short walks and excluding 'series of calls' trips.
JTTXSC	Trip travelling time, grossed for short walks and excluding 'series of calls'.

These weighted variables have been constructed as follows:

SSXSC

If 'series of calls'	SSXSC = 0
If not 'series of calls' and 'short walk stage'	SSXSC = 7
If not 'series of calls' and not 'short walk stage'	SSXSC = 1

STTXSC

If 'series of calls'	STTXSC = 0 * StageTime
If not 'series of calls' and 'short walk stage'	STTXSC = 7 * StageTime
If not 'series of calls' and not 'short walk stage'	STTXSC = 1 * StageTime

SD

If 'short walk stage'	SD = 7 * StageDistance
If not 'short walk stage'	SD = 1 * StageDistance

JJXSC

If 'series of calls'	JJXSC = 0
If not 'series of calls' and 'short walk trip'	JJXSC = 7
If not 'series of calls' and not 'short walk trip'	JJXSC = 1

JOTXSC

If 'series of calls'	JOTXSC = 0 * TripTotalTime
If not 'series of calls' and 'short walk trip'	JOTXSC = 7 * TripTotalTime
If not 'series of calls' and not 'short walk trip'	JOTXSC = 1 * TripTotalTime

JTTXSC

If 'series of calls'	JTTXSC = 0 * TripTravTime
If not 'series of calls' and 'short walk trip'	JTTXSC = 7 * TripTravTime
If not 'series of calls' and not 'short walk trip'	JTTXSC = 1 * TripTravTime

JD

JD = Sum(SD) Stage distances

Non-response and "drop-off" weighting

In addition to the short walk weights described above, a weighting strategy for the NTS was developed following a recommendation in the 2000 National Statistics Quality Review of the NTS. For the first time, the 2005 NTS results were based on weighted data. The weighting methodology was applied to data back to 1995 and all NTS figures for 1995 onwards which are published or released are now based on weighted data.

As well as adjusting for non-response bias, the weighting strategy for the NTS also adjusts for the drop-off in the number of trips recorded by respondents during the course of the travel week; for uneven recording of short walks by day of the week and for the short-fall in reporting of long distance trips.

Further information on the weighting methodology is available in the Technical Report at the link below. Please note that "W" notation in the Technical Report is not the same as the "W" notation in the NTS datasets as described in the 'Weighting matrices' section below). https://www.gov.uk/government/collections/national-travel-survey-statistics

Therefore, there are several sets of weights which apply to different levels of the database; household, trip and long distance journey. It is important to select the correct weights for each analysis. Initial results should be checked against published data to ensure weights are being applied correctly.

There are two samples which can be used for analysis:

Diary sample

Analysis of travel data is based on the diary sample. This comprises all 'fully co-operating households', defined as households for which the following information is available: a household interview, an individual interview for each household member, a seven day travel diary for each individual and, where applicable, at least one completed vehicle section. Weights were produced to adjust for non-response and, at the trip-level, they were also produced for drop-off in recording observed during the seven day travel week.

Interview sample

Analyses at household, individual and vehicle level are based on the interview sample. This sample comprises all fully co-operating households included in the diary sample, together with some additional 'partially co-operating households'. Generally these partially co-operating households had co-operated fully with the various interviews but not all household members had completed the travel diary. Prior to the introduction of the weighting methodology, data from partially co-operating households was not included in NTS analyses but this expanded dataset is now used for analyses which do not require data from the seven day travel record.

Fully co-operating households (Interview data + travel data)	} Diary sample
Partially co-operating households (Interview data only)	
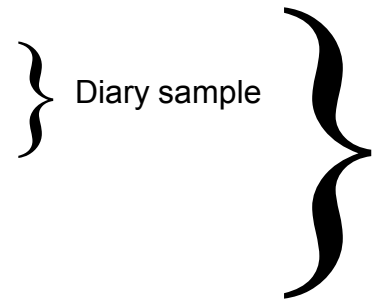
It is important to use the correct sample for all analysis.

The weighting methodology produces weights at the household, LDJ and trip level. The household weights apply to all individuals and vehicles within the household, and they have therefore been attached to the individual and vehicle files for ease of use. Similarly, the trip level weights apply to all stages within trips and have therefore been attached to the stage-level records for ease of use.

The variable ' OutCom_B02ID' (1=fully productive, 0=partially productive) is included on the household file to enable users to select the diary sample. This variable can be merged into other files if necessary, e.g. individual file.

Weighting matrices

The following weighting matrices are available and should be used on the appropriate level data for the analysis being produced:

W1

Unweighted diary sample - this gives unweighted results for the diary sample only. (This is equivalent to the results produced before the weighting strategy was introduced and can be used to generate unweighted sample sizes for analysis of the diary sample. It is effectively the same as the OutCom_B02ID variable mentioned above – formerly named the 'status' variable).

W2

Diary sample household weight – (adjusts for non-response) apply to all analysis of the diary sample at **household**, **individual** and **vehicle** level.

W3

Interview sample household weight - apply to all analysis of the interview sample at **household**, **individual** and **vehicle** level.

W4

LDJ weight incorporating household weight - apply to all analysis at **LDJ** level

W4xhh

LDJ weight excluding household weight – apply only if also applying a household weight (i.e. W2 or W3)

W5

Trip/stage weight - apply to all analysis of **trip** and **stage** data

W5xhh

Trip/stage weight excluding household weight - apply only if also applying a household weight (i.e. W2 or W3)

No weighting matrix or W0

If no weighting matrix is applied, this gives unweighted results for the interview sample.

- For most analyses at household, individual and vehicle level, w3 should be applied.

- For most analyses of travel patterns, w5 should be applied to trip/stage data and w2 should be applied at the individual level in order to calculate rates. In addition, the short walk weights should be applied (see 'Short walk weighting' above).

Examples of applying weights:
- To generate trip rates - apply w5 to trip data and apply w2 to individual data (i.e. Diary sample)
- To calculate household car ownership - apply w3 to the household data (Interview sample)
- To calculate the proportion of driving licence holders - apply w3 to the individual data (Interview sample)
- To determine the unweighted sample size for trip rate analysis - apply w1 to the trip data and w1 to the individual data (Diary sample)
- To determine the unweighted sample size for household car ownership or driving licence figures - apply no weights or w0 (Interview sample)

Long distance journey (LDJ) weights
Long distance journeys (trips over 50 miles) are collected in two ways in the NTS – via the travel diary and in the interview.

From 2002-2005 long distance trips were collected during the interview for an additional period of 3 weeks prior to the start of the travel diary. From 2006 onwards, they have been collected during the single week prior to the placement interview.

Analysis of the LDJ data a few years ago highlighted concerns over the quality of the interview recall data compared to the data collected during the travel week via the diary. Therefore the LDJ weighting methodology was revised back to 2006 to take account of the

systematic under-reporting of shorter long distance journeys. The re-weighting had little impact upon the overall LDJ trip rates. However, it resulted in a downward adjustment to distance estimates.

Revisions

The following are the main changes made to the 2002-2012 datasets which account for the difference in the datasets for 2002-2010 previously available from the UK Data Service.

SQL processing of 2008-2012 raw data
Due to changed imputation methodologies for key variables, e.g. trip mode and purpose, the trip data processed in SQL for 2008-2010 needed to be re-weighted. In most cases the overall results have changed very little from those processed in the old database.

Geographic data
A full recode of all geographic variables at all levels was conducted so that there was a consistent time series available at household level. This included creating a new Area Type variable which now correctly allocates households in South Yorkshire to the Metropolitan built-up areas category. This recoding also now makes the allocation of Scottish postcodes to an area type prior to 2007 consistent with 2007 onwards. More information on Area type can be found in the 2012 Notes & Definitions document. All published results tables by region and area type are now all based on household geographies rather than at the PSU level so there are some minor differences.

Ticket data
The data on tickets used and the associated costs for each stage have been re-processed in a consistent way for all years from 2002. These data are now available in a separate ticket file – previously they were within the individual data file.

Corrections
Whilst working on the database redevelopment project some errors were found in the processing of some NTS data in Quantum Quanvert. These errors have been corrected in the new SQL database and therefore the raw datasets. The allocation of 'missing' (NA) and 'does not apply' (DNA) cases has also been made consistent over all years.

Imputation indicators
The NTS now includes imputation indicators in the data files for Household, Vehicle, Individual, Trip and Stage for some key variables from 2008 onwards. These indicators inform users if the value was originally missing and hence imputed.

Weighted and unweighted sample sizes, 1995-2014

Sample numbers on which analyses are based: England, 1995 to 2014[1]

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Household sample																				
Unweighted diary	2,733	2,747	2,711	2,503	2,559	2,957	2,990	6,301	7,052	6,951	7,166	7,056	7,145	6,965	7,225	6,970	6,668	7,078	6,830	6,900
Unweighted interview	2,982	3,000	3,000	2,819	2,879	3,255	3,232	7,537	7,853	7,692	8,065	7,884	7,879	7,665	7,859	7,534	7,290	7,725	7,822	7,439
Weighted diary	2,749	2,753	2,698	2,510	2,585	2,945	3,002	6,388	7,100	6,980	7,251	7,124	7,227	6,983	7,186	6,945	6,624	7,038	6,830	6,900
Weighted interview	2,987	3,005	2,977	2,816	2,897	3,248	3,254	7,593	7,898	7,738	8,137	7,950	7,957	7,695	7,822	7,520	7,257	7,701	7,822	7,439
Individual sample																				
Unweighted diary	6,580	6,587	6,454	5,876	5,927	6,936	6,883	14,369	16,685	16,487	16,956	16,648	16,858	16,360	17,299	16,553	15,730	16,670	16,192	16,491
Unweighted interview	7,296	7,307	7,325	6,839	6,846	7,783	7,581	17,851	18,876	18,506	19,402	18,945	18,763	18,218	18,956	18,061	17,366	18,441	18,765	17,909
Weighted diary	6,581	6,631	6,417	6,003	6,122	6,964	7,022	15,068	16,868	16,630	17,321	17,063	17,201	16,608	17,180	16,582	15,850	16,900	16,411	16,624
Weighted interview	7,156	7,235	7,089	6,743	6,850	7,669	7,600	17,906	18,778	18,412	19,420	19,049	18,942	18,303	18,724	17,974	17,337	18,486	18,786	17,913
Child (<16 yrs) sample																				
Unweighted diary	1,483	1,439	1,418	1,244	1,268	1,494	1,424	2,917	3,615	3,566	3,540	3,386	3,426	3,292	3,566	3,417	3,128	3,242	3,291	3,398
Unweighted interview	1,665	1,609	1,658	1,492	1,474	1,698	1,590	3,792	4,066	4,031	4,079	3,870	3,798	3,686	3,892	3,710	3,442	3,626	3,831	3,699
Weighted diary	1,373	1,411	1,311	1,257	1,283	1,434	1,388	3,042	3,385	3,292	3,405	3,334	3,259	3,180	3,313	3,208	3,083	3,266	3,171	3,224
Weighted interview	1,490	1,535	1,451	1,417	1,421	1,574	1,512	3,606	3,751	3,630	3,813	3,718	3,582	3,502	3,596	3,479	3,353	3,555	3,623	3,471
Adult (16+) sample																				
Unweighted diary	5,097	5,148	5,036	4,632	4,659	5,442	5,459	11,452	13,070	12,921	13,416	13,262	13,432	13,068	13,733	13,136	12,602	13,428	12,901	13,093
Unweighted interview	5,631	5,698	5,667	5,347	5,372	6,085	5,991	14,059	14,810	14,475	15,323	15,075	14,965	14,532	15,064	14,351	13,924	14,815	14,934	14,210
Weighted diary	5,208	5,220	5,106	4,746	4,839	5,530	5,634	12,026	13,483	13,338	13,916	13,729	13,942	13,428	13,867	13,375	12,767	13,634	13,240	13,400
Weighted interview	5,665	5,700	5,637	5,326	5,430	6,096	6,088	14,301	15,027	14,782	15,607	15,331	15,361	14,801	15,129	14,495	13,984	14,931	15,163	14,442
Motor vehicle sample																				
Unweighted diary	2,863	2,916	2,828	2,715	2,755	3,256	3,254	7,069	8,047	7,840	8,467	8,391	8,556	8,264	8,660	8,456	7,954	8,429	8,129	8,245
Unweighted interview	3,175	3,256	3,210	3,136	3,171	3,652	3,562	8,622	9,082	8,802	9,677	9,577	9,489	9,175	9,449	9,235	8,764	9,251	9,358	8,957
Weighted diary	2,917	2,995	2,886	2,789	2,829	3,306	3,325	7,328	8,209	8,019	8,728	8,566	8,740	8,405	8,574	8,442	7,916	8,324	8,125	8,262
Weighted interview	3,169	3,273	3,202	3,125	3,176	3,644	3,588	8,679	9,130	8,911	9,781	9,579	9,592	9,245	9,329	9,132	8,674	9,106	9,333	8,949
Trip sample																				
Unweighted diary	107,869	108,633	106,338	97,026	98,013	112,564	112,123	239,426	270,853	266,915	275,222	268,250	261,533	254,889	267,214	251,798	235,327	251,555	238,293	243,126
Unweighted diary (short walks grossed up & excluding series of calls)	133,740	132,968	131,116	118,550	117,684	134,345	133,434	279,407	317,647	313,666	324,116	316,721	303,101	294,605	312,263	292,332	273,411	290,893	273,751	280,052
Unweighted interview	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Weighted diary	111,382	116,970	109,804	103,956	106,520	118,718	120,202	262,947	288,405	281,918	298,268	291,846	281,277	276,013	276,832	265,409	249,691	267,853	254,687	257,362
Weighted diary (short walks grossed up & excluding series of calls)	136,676	141,174	134,058	125,554	126,883	140,226	141,687	303,712	335,048	328,318	347,240	341,057	322,150	316,186	320,929	305,486	288,749	307,974	290,413	293,801
Weighted interview	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Stage sample																				
Unweighted diary	112,759	112,932	110,652	100,946	102,051	118,097	115,696	248,526	281,917	278,077	287,259	280,786	274,002	267,357	280,135	264,653	246,611	263,966	250,381	255,520
Unweighted diary (short walks grossed up & excluding series of calls)	149,298	146,885	145,096	129,826	129,690	149,022	143,907	299,727	341,545	338,340	350,595	344,479	330,936	321,557	341,528	321,633	299,059	319,090	301,637	308,089
Unweighted interview	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Weighted diary	116,469	121,784	114,408	108,276	111,131	124,763	124,392	273,832	301,184	294,926	312,314	306,221	295,456	290,352	290,908	279,585	262,209	281,950	268,793	271,448
Weighted diary (short walks grossed up & excluding series of calls)	152,744	156,460	149,077	137,992	140,785	156,449	154,231	328,775	363,303	357,514	378,974	374,015	354,371	347,884	354,204	338,824	318,407	341,180	323,877	326,370
Weighted interview	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

1 Data for 1995 to 2001 is based on calendars years. Data for 2002 to 2014 is based on survey years which run mid-January to mid-January.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/457695/nts2014-sample.xls

Annex A: Example tabulation 1

NTS data users should replicate analysis produced by DfT to ensure that they understand the methodology of weighting and the grossing up of short walks.

Trips per person per year by gender and survey year

Load the journey level SPSS data file 'Trip.sav'

Load the individual level SPSS data file 'Individual.sav'

Before performing the analysis you should ensure that both files are sorted based on the key identifier variables at each level.

Select the **Trip** file.

From the menu select **Data -> Sort Cases**

Highlight **TripID** and move over to the 'Sort by' list. Ensure that **ascending** is checked.

The 'Sort by' list should contain:

TripID (A)

Select OK.

Select the **Individual** file.

From the menu select **Data -> Sort Cases**

Highlighting **IndividualID** and move over to the 'Sort by' list. Ensure that ascending is checked.

The 'Sort by' list should contain

IndividualID (A)

Select OK.

To disaggregate the data by the gender of the individual making the trip, individual level data needs to be attached to the trip file.

Select the **Trip** file.

From the menu **select Data -> Merge files -> Add variables**

Select '**An open dataset**' and then select file '**Individual.sav**'

Select continue.

Data should be linked using the identifier variable at individual level – **IndividualID**.

Check '**Match cases on key variables**' and '**Cases are sorted in order of key variables in both datasets**'

Check '**Non-active dataset is keyed table**'

From the excluded variables list highlight IndividualID and move to the key variables box

Select OK.

A message will appear 'Warning: Keyed match will fail if data are not sorted in ascending order of key variables'

This can be ignored as we have already ensured the correct sorting order.

The individual characteristics of the person making each trip will now be attached to the journey data.

Before producing the tabulation of gender by survey year, the weighting must be specified:

From the menu select Data -> Weight cases

Check 'Weight cases by' and move across W5 (trip weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **Analyze -> Compare Means -> Means**

To gross up short walks we must use variable **JJXSC** as the dependent list variable.

Move JJXSC across to 'Dependent List'

Select **Gender** as the layer 1 variable by moving **Sex_B01ID** across to the independent list selection

Select 'Next' to choose the 2nd layer variable

Select **Survey Year** as the layer 2 variable by moving **SurveyYear** across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the 'Cell statistics' box. Remove all other statistics.

Select continue.

Select OK.

The following SPSS output should be produced:

Report

Sum

Sex of person	Survey year - actual year	Number of trips - grossing up short walks and excluding series of calls
Male	2002	170087
	2003	188735
	2004	183612
	2005	194585
	2006	188663
	2007	177980
	2008	175774
	2009	177365
	2010	169684
	2011	161512
	2012	172216
	Total	1960213
Female	2002	181036
	2003	199417
	2004	196275
	2005	207770
	2006	205150
	2007	193844
	2008	190552
	2009	194107
	2010	185215
	2011	175516
	2012	185400
	Total	2114284
Total	2002	351123
	2003	388152
	2004	379887
	2005	402356
	2006	393812
	2007	371823
	2008	366326
	2009	371473
	2010	354900
	2011	337029
	2012	357616
	Total	4074497

These are the weighted and grossed up number of trips *per week* for each gender/survey year combination.

For each gender/survey year the weighted number of individuals also needs to be known. Note – it is standard NTS methodology to include all respondents in the denominator not just those making a trip.

Select the '**Individual.sav**' file.

Before producing the tabulation of gender by survey year, the weighting must be specified

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W2** (household non-response weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **'Analyze' -> 'Descriptive Statistics' -> 'Crosstabs'**

Move **Survey Year** across to the row specification box

Move **Sex_B01ID** across to the column specification box

Select **Cells** and in the **Non-integer weights** section select **'No adjustments'**

Select OK.

The following SPSS output should be produced:

Survey year - actual year * Sex of person Crosstabulation

Count

		Sex of person		Total
		Male	Female	
Survey year - actual year	2002	8549.647	8944.077	17493.723
	2003	9578.418	9999.596	19578.014
	2004	9451.343	9850.284	19301.626
	2005	9843.876	10259.522	20103.397
	2006	9699.779	10093.774	19793.553
	2007	9772.677	10167.305	19939.982
	2008	9454.678	9798.554	19253.233
	2009	9770.212	10147.339	19917.551
	2010	9435.616	9774.031	19209.647
	2011	9008.588	9330.581	18339.169
	2012	9604.327	9942.373	19546.700
Total		104169.160	108307.436	212476.596

The trip counts can now be divided by the individual sample. It should be noted that the trip counts are for a single week so therefore need to be multiplied by 52.14 (number of weeks in a year).

e.g. Trips per male in 2002 $= \frac{170,087}{8,549.647} \times 52.14$

$$= 1,037$$

The following results should be calculated showing trips per person per year by gender:

	Year	Trip rate
Male	2002	1037
	2003	1027
	2004	1013
	2005	1031
	2006	1014
	2007	950
	2008	969
	2009	947
	2010	938
	2011	935
	2012	935
	2002/12 average	981
Female	2002	1055
	2003	1040
	2004	1039
	2005	1056
	2006	1060
	2007	994
	2008	1014
	2009	997
	2010	988
	2011	981
	2012	972
	2002/12 average	1018
Total	2002	1047
	2003	1034
	2004	1026
	2005	1044
	2006	1037
	2007	972
	2008	992
	2009	972
	2010	963
	2011	958
	2012	954
	2002/12 average	1000

Finally, repeat the above step using W1 to get the unweighted number of individuals. This needs to be known to ensure that the base sample sizes are sufficient.

Select the '**Individual.sav**' file.

Before producing the tabulation of gender by survey year, the weighting must be specified

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W1** (unweighted diary weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **'Analyze' -> 'Descriptive Statistics' -> 'Crosstabs'**

Move **Survey Year** across to the row specification box

Move **Sex_B01ID** across to the column specification box

Select OK.

The following SPSS output should be produced:

Survey year - actual year * Sex of person Crosstabulation

Count

		Sex of person		Total
		Male	Female	
Survey year - actual year	2002	8085.000	8801.000	16886.000
	2003	9360.000	10107.000	19467.000
	2004	9219.000	9980.000	19199.000
	2005	9560.000	10344.000	19904.000
	2006	9396.000	10094.000	19490.000
	2007	9490.000	10245.000	19735.000
	2008	9127.000	9856.000	18983.000
	2009	9583.000	10331.000	19914.000
	2010	9232.000	9840.000	19072.000
	2011	8718.000	9351.000	18069.000
	2012	9244.000	9910.000	19154.000
Total		101014.000	108859.000	209873.000

Annex B: Example tabulation 2

Miles per person per year, by mode (at stage level)

Load the stage level data file 'Stage.sav'

Load the individual level data file 'Individual.sav'

Before performing the analysis you should ensure that both files are sorted based on the key identifier variables at each level.

Select the **Stage** file.

From the menu select **Data -> Sort Cases**

Highlight **StageID** and move over to the 'Sort by' list. Ensure that ascending is checked.

Select OK.

Select the **Individual** file.

From the menu select **Data -> Sort Cases**

Highlighting **IndividualID** and move over to the 'Sort by' list. Ensure that ascending is checked.

Select OK.

Before producing the tabulation of stage distance by survey year and mode, the weighting must be specified.

Select the **Stage** file

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W5** (stage weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **Analyze -> Compare Means -> Means**

To gross up short walks we must use variable **SD** as the dependent list variable.

Move **SD** across to 'Dependent List'

Select **SurveyYear** as the layer 1 variable by moving it across to the independent list selection

Select 'Next' to choose the 2nd layer variable

Select **Stage Mode** as the layer 2 variable by moving **StageMode_B01ID** across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the '**Cell statistics**' box. Remove all other statistics.

Select continue.

Select OK.

The following output should be produced (for all years down to 2012):

		Report		
Sum		
Survey year - actual year	Stage mode of travel - as recorded in diary - 18 categories	Stage distance - grossing up short walks
2002	Walk	66278.50
	Bicycle	12132.08
	Private (hire) bus	41637.52
	Car	1845539.78
	Motorcycle, scooter, moped	11629.80
	Van/lorry	92132.81
	Other private	6914.02
	London stage bus	18742.35
	Other stage bus	75319.03
	Coach/Express bus	10923.44
	Excursion/Tour bus	9005.67
	London Underground	27065.54
	Surface Rail	138536.02
	Light rail	2536.75
	Air	14752.25
	Taxi	13266.63
	Minicab	6594.07
	Other public	1073.24
	Total	2394079.50

These are the weighted and grossed up number of stage distances *per week* for each survey year/mode combination.

For each mode/survey year the number of individuals also needs to be known. Note – it is standard NTS methodology to include all respondents in the denominator not just those making a trip.

Select the '**Individual.sav**' file.

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W2** (household weight) as the frequency variable.

Select OK.

Before producing the tabulation of mode by survey year, the weighting must be specified.

From the menu select '**Analyze**' -> '**Descriptive Statistics**' -> '**Frequencies**'

Move **Survey Year** across to the Column box

Select OK.

The tabulation can now be specified.

The following output should be produced:

Survey year - actual year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2002	17494	8.2	8.2	8.2
	2003	19578	9.2	9.2	17.4
	2004	19302	9.1	9.1	26.5
	2005	20103	9.5	9.5	36.0
	2006	19794	9.3	9.3	45.3
	2007	19940	9.4	9.4	54.7
	2008	19253	9.1	9.1	63.8
	2009	19918	9.4	9.4	73.1
	2010	19210	9.0	9.0	82.2
	2011	18339	8.6	8.6	90.8
	2012	19547	9.2	9.2	100.0
	Total	212477	100.0	100.0	

The stage distances can now be divided by the individual sample (the frequency column). It should be noted that the stage distances are for a single week so therefore need to be multiplied by 52.14 (number of weeks in a year).

e.g. Trips per male in 2002 $= \frac{66278.50}{17,494} \times 52.14$

$$= 197.54$$

As in Annex A, repeat the above step using W1 to get the unweighted number of individuals. This needs to be known to ensure that the base sample sizes are sufficient.

Select the '**Individual.sav**' file.

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W1** (unweighted diary weight) as the frequency variable.

Select OK.

The tabulation can now be specified.

From the menu select **'Analyze' -> 'Descriptive Statistics' -> 'Frequencies'**

Move **Survey Year** across to the Column box

Select OK.

The following SPSS output should be produced:

Survey year - actual year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2002	16886	8.0	8.0	8.0
	2003	19467	9.3	9.3	17.3
	2004	19199	9.1	9.1	26.5
	2005	19904	9.5	9.5	36.0
	2006	19490	9.3	9.3	45.2
	2007	19735	9.4	9.4	54.6
	2008	18983	9.0	9.0	63.7
	2009	19914	9.5	9.5	73.2
	2010	19072	9.1	9.1	82.3
	2011	18069	8.6	8.6	90.9
	2012	19154	9.1	9.1	100.0
	Total	209873	100.0	100.0	

In addition, the unweighted number of stages for each mode should be obtained to ensure that the sample sizes on which the miles are based are sufficient.

Annex C: Example tabulation 3 – Multi-coded data

Reasons for not learning to drive by age

Load the individual level data file 'Individual.sav'

Before producing the tabulation of individuals and their reasons for not learning to drive by age, the weighting must be specified.

From the menu select **Data -> Weight cases**

Check '**Weight cases by**' and move across **W3** (interview sample weight) as the frequency variable.

Select OK.

In this case we are only interested in data for 2012. To restrict the data file to cases for 2012

From the menu select **Data -> Select cases**

Check '**If condition is satisfied'** and select **If…**

Move **SurveyYear** across to the definition box and enter '**=2012**'.

The box should read '**SurveyYear=2012**'.

Select OK.

only:

Any outputs will now only use individuals from the 2012 survey.

The tabulation can now be specified.

From the menu select **'Analyze' -> 'Descriptive Statistics' -> 'Crosstabs'**

Move **ResNDN_A_B01ID** across to the row specification box

Repeat this step for all **ResNDN** variables (B to P).

Move **Age_B04ID** across to the column specification box

Select **Cells** and in the **Non-integer weights** section select '**No adjustments**'

Select OK.

The following SPSS output should be produced (for each response variable – there are 16 possible responses, A-P):

Reason individual does not drive - Family or friends can drive me when necessary * Age of person - banded age - Band D - All ages - 9 categories * Survey year - actual year Crosstabulation

Count

Survey year - actual year			Age of person - banded age - Band D - All ages - 9 categories									Total
			0 - 4 years	5 - 10 years	11 - 16 years	17 - 20 years	21 - 29 years	30 - 39 years	40 - 49 years	50 - 59 years	60 years +	
2012	Reason individual does not drive - Family or friends can drive me when necessary	Yes	.000	.000	.000	132.966	166.050	133.267	111.638	134.147	417.959	1096.027
		No	1348.569	1454.680	1525.500	910.225	2419.969	2700.392	3063.582	2520.721	4345.085	20288.725
	Total		1348.569	1454.680	1525.500	1043.191	2586.019	2833.659	3175.221	2654.868	4763.045	21384.752
Total	Reason individual does not drive - Family or friends can drive me when necessary	Yes	.000	.000	.000	132.966	166.050	133.267	111.638	134.147	417.959	1096.027
		No	1348.569	1454.680	1525.500	910.225	2419.969	2700.392	3063.582	2520.721	4345.085	20288.725
	Total		1348.569	1454.680	1525.500	1043.191	2586.019	2833.659	3175.221	2654.868	4763.045	21384.752

The weighted number of individuals also needs to be known.

From the menu select **Data -> Weight cases**

Check '**Do not weight cases**'

Select OK.

First, turn the weighting off:

Next, we must ensure that we are only counting individuals who have been asked their

From the menu select **Data -> Select cases**

Select **If…** condition.

Edit the definition to select individuals answering the question.

The box should read '**SurveyYear=2012 & (ResNDN_NA_B01ID=2 & ResNDN_DNA_B01ID=2 & ResNDN_DEAD_B01ID=2)**'.

Select OK.

reasons for not learning to drive.

From the menu select **Analyze -> Compare Means -> Means**

To count the weighted number of individuals we need to count **W3**.

Move **W3** across to 'Dependent List'

Select **Age_B04ID** as the layer 1 variable by moving it across to the independent list selection

Select '**Options**' to determine the type of statistics to perform.

Move '**Sum**' across to the '**Cell statistics**' box. Remove all other statistics.

Select continue.

Select OK.

The tabulation can now be specified:

The following SPSS output should be produced:

Report

Sum

Age of person - banded age - Band D - All ages - 9 categories	Weighted interview sample
17 - 20 years	447.1548920
21 - 29 years	696.8862122
30 - 39 years	531.8084065
40 - 49 years	449.5771186
50 - 59 years	470.9010953
60 years +	1161.584938
Total	3757.912662

To calculate all percentages (as in NTS0203) divide the weighted number of individuals for each response and age group by the relevant total number of individuals.

e.g. Percentage of individuals aged 17-20 with the reason 'Family or friends can drive me when necessary' given =

$$\frac{132.9655693}{447.1548920} \times 100 = 30\%$$

Repeat, the above step with **W0** (no weight) to get the unweighted number of individuals.

The following SPSS output should be produced:

Report

Sum

Age of person - banded age - Band D - All ages - 9 categories	Unweighted interview sample
17 - 20 years	429
21 - 29 years	585
30 - 39 years	465
40 - 49 years	401
50 - 59 years	470
60 years +	1240
Total	3590