Department for Transport

Weighting short walks in the National Travel Survey

Methodology report

Moving Britain Ahead

December 2015

The Department for Transport has actively considered the needs of blind and partially sighted people in accessing this document. The text will be made available in full on the Department's website. The text may be freely downloaded and translated by individuals or organisations for conversion into other accessible formats. If you have other needs in this regard please contact the Department.

Department for Transport
Great Minster House
33 Horseferry Road
London SW1P 4DR
Telephone 0300 330 3000
General enquiries https://forms.dft.gov.uk
Website www.gov.uk/dft

Contents

Introduction

The National Travel Survey (NTS) is a continuous survey of households run by the Department for Transport (DfT). Information on trips travelled is collected by means of a 7-day travel diary. In order to reduce the burden on respondents, information on short walks of greater than 50 yards and less than one mile, are only collected on day 7. Information on longer walks, and other modes of transport, is collected on all days.

Results from the travel diary show that the number of trips recorded per day on the travel diary drops as the week progresses. This is corrected for by applying weights by trip purpose, calculated from the observed tailing off of responses during the travel week. However, it is not possible to do this for short walks that involve no other form of transport (called short walk trips) because they are only collected on day 7. As a consequence, short-walk trips currently have no weights applied to them to correct for under-reporting.

This report sets out proposals for weighting short walk trips, describing the source data and the proposed methods. It has been prepared by Jim O'Donoghue from the Office for National Statistics' Methodology Advisory Service.

Summary of recommendations

In relation to the weighting of short walks within the National Travel Survey:

1 Short walk only trips and short walks that form part of multi-mode trips should be considered separately.

2 No further under-reporting adjustments are necessary for short walk stages that are part of multi-mode trips.

3 The recommended approach to weighting for short walk trips is to apply a weight based on the likelihood of the individual undertaking a short walk trip on day 1 compared to day 7.

4 The logistic models should be run with age/economic status and car access as the independent variables.

5 The preferred option for time series of short walks for England is to reweight earlier periods.

6 Data to reweight earlier periods of England time series should be collected on both day 1 and day 7 bases for an overlap period covering one year with half the households using the day 1 collection and the other half day 7. This would also present an opportunity to ensure that the day 1 and day 7 sub-samples cover comparable areas.

7 The NTS team should consider the need for Wales and Scotland to be weighted given that they are not included in the current NTS.

8 If it is decided to reweight Scotland and Wales, in the absence of any other information, the weights should be the same as those used for England.

9 Users should be made aware of the limitations of these weights for Scotland and Wales.

1. Background

1.1 In order to establish whether under-reporting of short walk trips was an issue, DfT conducted an experiment during April to June 2013 whereby one-quarter of households in the NTS sample were asked to record short walk stages on day 1 as opposed to day 7. This found that for individuals who reported any short walks:

a. there was only a small difference in the number of short walks reported: 2.58 on day 1 compared with 2.68 on day 7;

b. there was no change in the average distance of the short walks reported: 0.37 miles for both groups;

c. when short walks were part of a multi-stage trip, there was a very small change in the average number of short walks per adult from 0.20 on day 7 to 0.21 on day 1.

But it was also found that:

d. there was an increase in reporting of short walk only trips, from 0.34 short walks per adult on day 7 to 0.54 on day 1.

1.2 These findings suggest that the difference in reporting of short walk trips is at the person level. If someone reports short walks at all, they don't forget individual short walks at a different rate to those reporting on day 1. The issue is that a larger proportion of people fail to report any short walks if collected on day 7.

1.3 The very small increase identified in point c is likely to be due to the already identified drop-off in reporting during the travel week. Therefore, the increase in the reporting of short walks predominantly comes from the increase in reporting of short walk only trips.

1.4 The impact of under-reporting of short walks on trip rates is likely to be significant. In 2012, short walk stages made up 74% of all walking stages and of these 68% were single stage trips, i.e. meaning the main mode for the trip would be walking. This equates to half of all short walk stages also being short walk trips and therefore affected by under reporting. The other half are part of multi-mode trips where another mode of transport would be classified as the main mode and therefore the historic trip rate result for these would be unaffected.

1.5 DfT reported on the results of the experiment and sought the views of users on the way forward in a public consultation document[1]. The clear message received from users[2] was that:

a. in future, short walks should be collected on day 1;

b. historic results should be re-weighted to take account of under-reporting of short walks.

1.6 Before implementing these recommendations, DfT is conducting a repeat of the experiment in 2015 Q2. These results will be used to validate and supplement those from 2013.

[1] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/336429/short-walks-consultation.pdf
[2] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/387524/NTS-short-walks-consultation-response.pdf

2. Proposed approach

2.1 **R1: It is recommended that short walk only trips and short walks that form part of multi-mode trips should be considered separately.**

2.2 Under-reporting of short walk stages that are part of multi-mode trips should be accounted for by the weights that are applied to take account of the drop in reporting that takes place during the course of the travel week. These weights are applied at the trip level. **R2: It is recommended that no further under-reporting adjustments are necessary for these trips.** The rest of this report therefore focuses solely on short walk trips.

2.3 **R3: The recommended approach to weighting for short walk trips is to apply a weight based on the likelihood of the individual undertaking a short walk trip on day 1 compared to day 7.** It is proposed to do this by running logistic regressions, where the dependent variable is an indicator variable that takes the value 0 if there are no short walks reported by the individual, and the value 1 if short walks are reported. The aim is to develop a set of weights that can be applied consistently back through time.

2.4 A good background to logistic regressions is given in Wikipedia[3]. The aim is to fit a model where the log of the odds of the dependent variable (whether there are any short walks) is a linear combination of relevant independent variables. Each possible value of the explanatory variable is treated as a separate variable which takes the value 0 or 1 to indicate whether it has the desired characteristic or not: e.g. for sex, there are 2 explanatory variables, male and female. If the respondent is a man, the male variable will take the value 1, and the female variable the value zero. When analysing the output of the regression, the coefficient of one of the values for each explanatory variables will be a reference category that it set to zero (exponential equal to one); and the exponential of the coefficients of other values represents their likelihood relative to the reference category of having the desired characteristic.

2.5 The logistic function can be written as:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n)}}$$

F(x) is interpreted as the probability of the dependent variable equalling a "success" or "case".

[3] http://en.wikipedia.org/wiki/Logistic_regression

2.6 The inverse of the logistic function, g, is the logit (log odds); this is the regression equation:

$$g\big(F(x)\big) = ln\frac{F(x)}{1-F(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots . \beta_n x_n$$

2.7 Separate models were developed for individuals recording short walks on day 1 and for those recording short walks on day 7. F(x) was calculated for each combination of the explanatory variables for each model. The short walk trip under-reporting weights were then calculated as the ratio of the odds obtained for day 1 and day 7, for the same combination of values for the explanatory variables.

2.8 The first step in deriving the models was to identify the variables that were related to an individual's likelihood of reporting short walks. The following were examined:

a. age
b. sex
c. economic status
d. region
e. access to a car
f. settlement type (urban/rural)
g. mobility
h. income

2.9 The first six variables were tested in the logistic regression. Mobility was excluded on the grounds that there were relatively few individuals with limited mobility, so results were likely to be insignificant. Income was excluded because:

- the bandings would be difficult to apply consistently back through time (e.g. the proportion of households with an income of £50k now will be much greater than 10 years ago);
- there are a large number of records with missing income in the day 1 dataset;
- income is strongly correlated with car ownership and hence car access.

2.10 Sex was found to be not significant. Age and economic status were both found to have a significant effect, although the significance of their values differed for day 1 and day 7, reflecting the degree of overlap between these two variables - e.g. children tend to be not in work, and elderly people tend to be retired. A composite variable called age_ecostat was therefore set up for use in the models. It took the following values:

1 Age 0-16
2 Age > 16, full time employment
3 Age > 16, part time
4 Age > 16, retired

5 Age > 16, other non-work

2.11 Access to a vehicle was found to be significant. A small number of individuals had their car access status missing. These were recoded as having no access to a car.

2.12 Region and settlement type were both found to be significant to differing extents in the two models but there was no consistent pattern. For instance, for day 7, a summary settlement type variable with just two categories, urban and rural, was found to have odds for urban-living individuals twice that for rural, but for day 1, it did not differ significantly from one. A composite variable combining region and settlement type was also tested, but this too was found to produce inconsistent results. Closer examination of the two samples showed that the day 1 sample did not include any individuals living in small towns for five regions, and had no residents in conurbations for three regions. This suggests that the two samples are not strictly comparable in the urban and region dimensions, and it was therefore decided to exclude these two variables from the models.

2.13 **R4: It is therefore recommended that the logistic models are run with age/economic status and car access as the independent variables.**

3. Results

3.1 Separate logistic regressions were run for day 1 and day 7 short walks where the survey start week was in April, May or June 2013. The models were run with age/economic status and car access as the independent variables. Note that there were also a few day 1 short walks recorded in July 2013, but these made up a much smaller proportion of the monthly total number of trips compared with the previous three months, and so were excluded from the analysis.

3.2 The results of the two logistic models are shown in the table below:

Variables in the Equation - day7

		B	S.E.	Wald	df	Sig.	Exp(B)	Frequency
age_ecostat				45.525	4	.000		
age_ecostat(1)	Age 0-16	.377	.177	4.558	1	.033	1.458	657
age_ecostat(2)	Full time	-.623	.167	13.837	1	.000	.536	1128
age_ecostat(3)	Part-time	-.032	.189	.029	1	.865	0.968	367
age_ecostat(4)	Retired	-.591	.171	11.921	1	.001	.554	756
age_ecostat(5)	Other non-work	.000					1.000	364
CarAccess				17.924	3	.000		
CarAccess(1)	With car - main driver	-.484	.137	12.490	1	.000	.616	1468
CarAccess(2)	With car - other driver	-.100	.183	0.299	1	.584	.905	318
CarAccess(3)	With car - non driver	-.434	.150	8.412	1	.004	.648	841
	Without car/van	.000					1.000	645
Constant		-1.059	.145	53.594	1	.000	.347	

Variables in the Equation - day1

		B	S.E.	Wald	df	Sig.	Exp(B)	Frequency
age_ecostat				10.402	4	.034		
age_ecostat(1)	Age 0-16	.297	.280	1.126	1	.289	1.346	219
age_ecostat(2)	Full time	-.500	.268	3.474	1	.062	.607	292
age_ecostat(3)	Part-time	.044	.306	.021	1	.885	1.045	108
age_ecostat(4)	Retired	-.315	.265	1.409	1	.235	.730	215
age_ecostat(5)	Other non-work	.000					1.000	118
CarAccess				29.001	3	.000		
CarAccess(1)	With car - main driver	-.909	.215	17.837	1	.000	.403	418
CarAccess(2)	With car - other driver	-.393	.293	1.802	1	.180	.675	88
CarAccess(3)	With car - non driver	-.998	.232	18.558	1	.000	.369	251
CarAccess(4)	Without car/van	.000					1.000	195
Constant		-.307	.234	1.721	1	.190	.736	

This translates into the following probabilities of an individual recording a short walk:

Likelihood of reporting a short walk trip on day 7

	With car			Without
	Main driver	Other driver	Non-driver	car/van
Age 0-16	..	..	0.247	0.336
Full time	0.103	0.144	0.108	0.157
Part-time	0.171	0.233	0.179	0.251
Retired	0.106	0.148	0.111	0.161
Other non-work	0.176	0.239	0.183	0.257

Likelihood of reporting a short walk trip on day 1

	With car			Without
	Main driver	Other driver	Non-driver	car/van
Age 0-16	..	..	0.267	0.498
Full time	0.152	0.232	0.141	0.309
Part-time	0.237	0.342	0.221	0.435
Retired	0.178	0.266	0.165	0.349
Other non-work	0.229	0.332	0.213	0.424

Which in turn translates into the following weights:

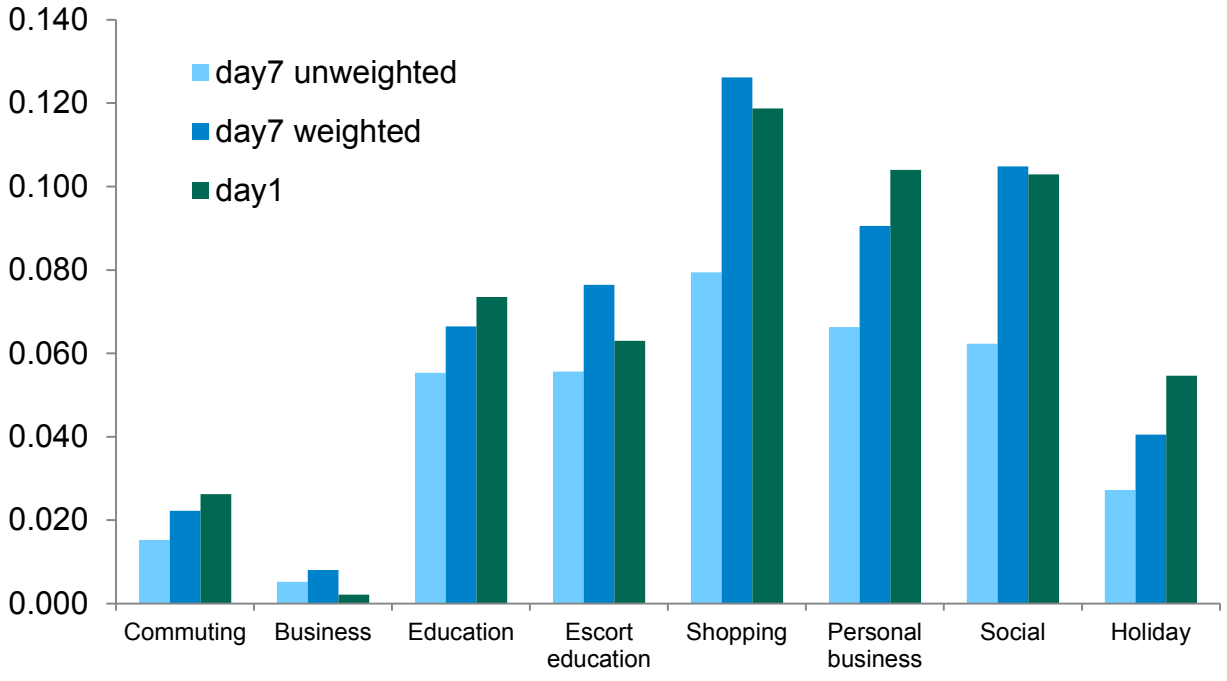
Relative likelihood - i.e. weight

	With car			Without
	Main driver	Other driver	Non-driver	car/van
Age 0-16	..	..	1.084	1.482
Full time	1.483	1.608	1.314	1.968
Part-time	1.380	1.467	1.237	1.730
Retired	1.682	1.799	1.494	2.170
Other non-work	1.299	1.390	1.163	1.647

3.3 It can be seen that retired people are much more likely to under-report short walks on day 7 relative to day 1, with children under-reporting by the least. Residents of households without access to a car or van are also more likely to under-report short walks, while non-drivers with access to a car are less likely to under-report. This results in weights ranging from 1.084 for child non-drivers to 2.170 for retired individuals without access to a car.

3.4 One way of checking that these results are reasonable is to compare the weighted day 7 results by trip purpose against those for day 1. These, and the day 7 unweighted short walk trip rates, are shown in the following table and chart:

Comparison of weighted and unweigthed short walk trips rates: 2013 Q2



Weighted number of short walk trips per person reported on day 7 compared with day 1 by trip purpose

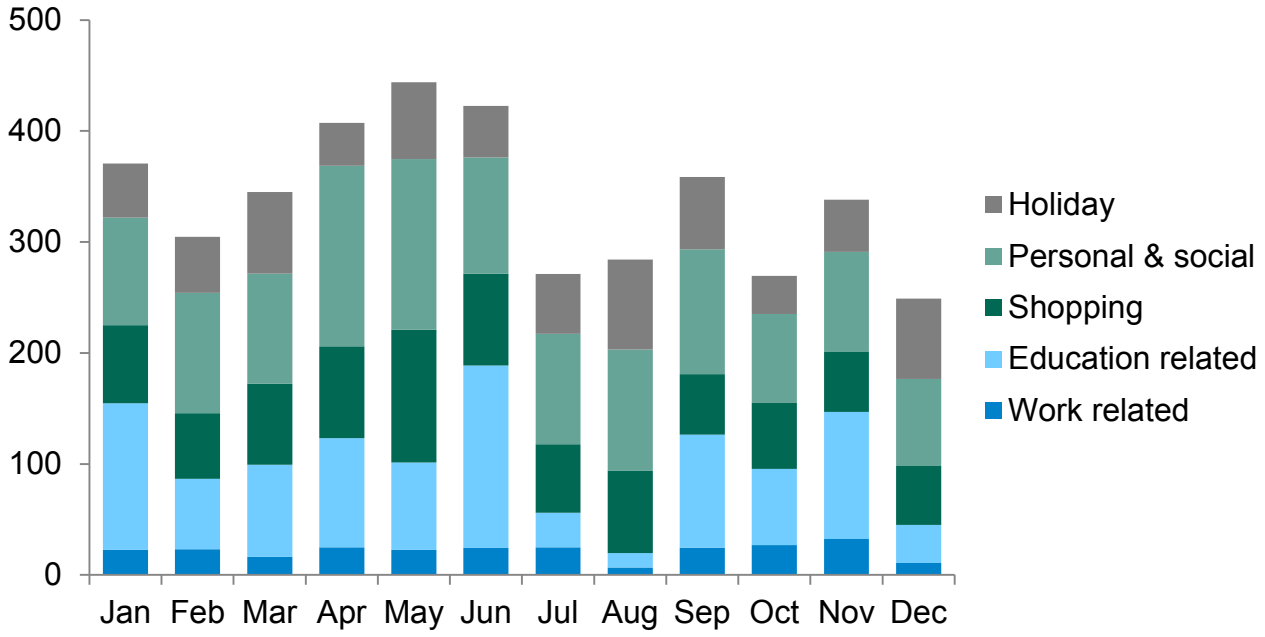
	Commuting	Business	Education	Escort education	Shopping	Personal business	Social	Holiday	All
day7 unweighted	0.015	0.005	0.055	0.056	0.079	0.062	0.066	0.027	0.367
day7 weighted	0.022	0.008	0.066	0.076	0.126	0.091	0.105	0.041	0.535
day1	0.026	0.002	0.074	0.063	0.119	0.104	0.103	0.055	0.545
difference	-0.004	0.006	-0.007	0.013	0.007	-0.013	0.002	-0.014	-0.010

3.5 It can be seen that the results seem plausible.

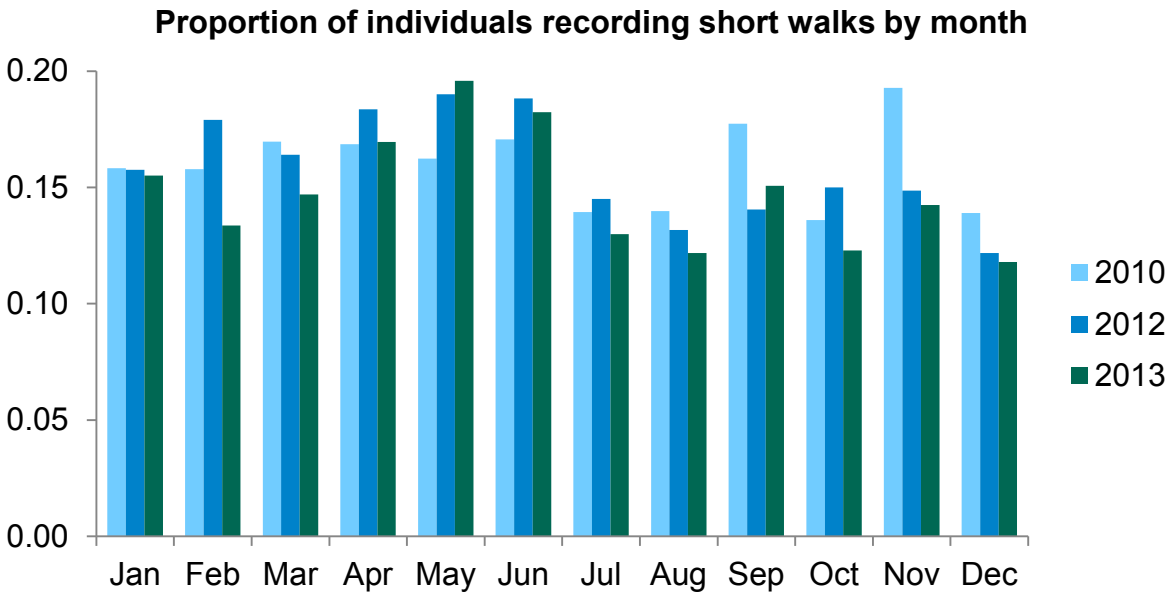
4. Applying the weights to other time periods

4.1 The weights calculated above are based on survey data relating to the April to June quarter in 2013. Their applicability to other times of the year and to other years needs to be assessed. The chart below shows that short walk trip rates in 2013 are highest in Q2 and lowest in July, August, October and December. The chart below shows that this is primarily due differences in the number of education-related and personal business/social trips, partly offset by an increase in holiday trips, although there is a fair amount of random variation.

Day 7 short walk trip rate per 1000 people by journey purpose and month: all areas: 2013



4.2 The proportion of individuals recording any short walk also varies by month as can be seen in the next chart, which also shows that the trip rates by month for 2013 are not particularly out of line with those for 2010 and 2012 and that Q2 tends to be when short walk trip rates are highest. From this we conclude that the 2013 Q2 results for day 7 short walks look reasonable. The repeat experiment taking place in 2015 Q2 should help validate this result but will provide no information on the relative likelihood of recording short walks on day 1 in other months. Ideally, the experiment should be repeated at other times of the year.



4.3 In order to apply these weights to earlier periods, it is necessary to assume that the relative propensity to record short walks on day 1 compared with day 7, after the individual's characteristics have been controlled for, is constant through the years. It is not possible to test this assumption directly, but we can examine the probability of recording short walks on day 7 in earlier years by running logistic regressions to see if the likelihoods classified by the associated variables are relatively stable. The resulting probabilities for 2002 to 2013 are charted in the annex. Some categories are relatively stable, particularly for main drivers, children and full-time workers, but some others show a slight downward trend. Overall, these results do not suggest that there has been a major shift in the likelihood of reporting short walks on day 7.

5. Options for time series for England

5.1 There are essentially three main options for time series of short walks for England:

a. switch to a day 1 collection, with no overlap period;

b. publish results on both bases (day 1 and day 7) for an overlap period (e.g. 2016/17) but do not revise previously published data;

c. reweight earlier periods.

These are considered in more detail below.

5.2 An immediate switch (from 2016/17) has the merit of introducing the new methodology as quickly as possible, and requiring the minimum of work by the NTS team. However, it introduces a break in the series and is not the preferred option for the majority of respondents to the consultation. The consultation found a clear preference for switching to day 1 collection of short walks. This was favoured by 12 out of the 13 who expressed a preference. Of these, 8 wanted historic results to be uplifted by the application of weights.

5.3 Calculating results for an overlap period appears to involve less work by the NTS team in the short run, as they would not have to reproduce weighted analyses for other periods. However, it is likely to be of limited help to users interested in time series analysis, as they would have to apply uplift factors based on the overlap period; the uplifted results may not be consistent, as the factors will depend on which analysis they are based; uplift factors are unlikely to be readily available for every possible analysis; and if a 50:50 split with the same overall sample size is used in the overlap period the results in that period will be less precise than in the periods before and after. Overall, there seems little to recommend this option.

5.4 The third option is the one that most closely aligns with user interests, as represented by the responses to the consultation exercise. The analysis in preceding sections of this report suggests that this option would be workable, but that further parallel data collections of day 1 and day 7 travel diaries covering all times of the year would be advisable to ensure robust results The NTS team advise that the earliest this could be done is 2016/17. The disadvantages of this approach are the delay in its implementation (which would not be before late 2017), and the possible need to re-run earlier analyses on a weighted basis. However, on balance **R5: this approach is the preferred option**. In order to implement it, **R6: data should be collected on both day 1 and day 7 bases for an overlap period covering one year with half the households using the day 1 collection and the other half day 7. This would also present an opportunity to ensure that the day 1 and day 7 sub-samples cover comparable areas**, as defined by the settlement variable, perhaps by

requiring alternate households in each primary sampling unit to be allocated to day 1 and day 7.

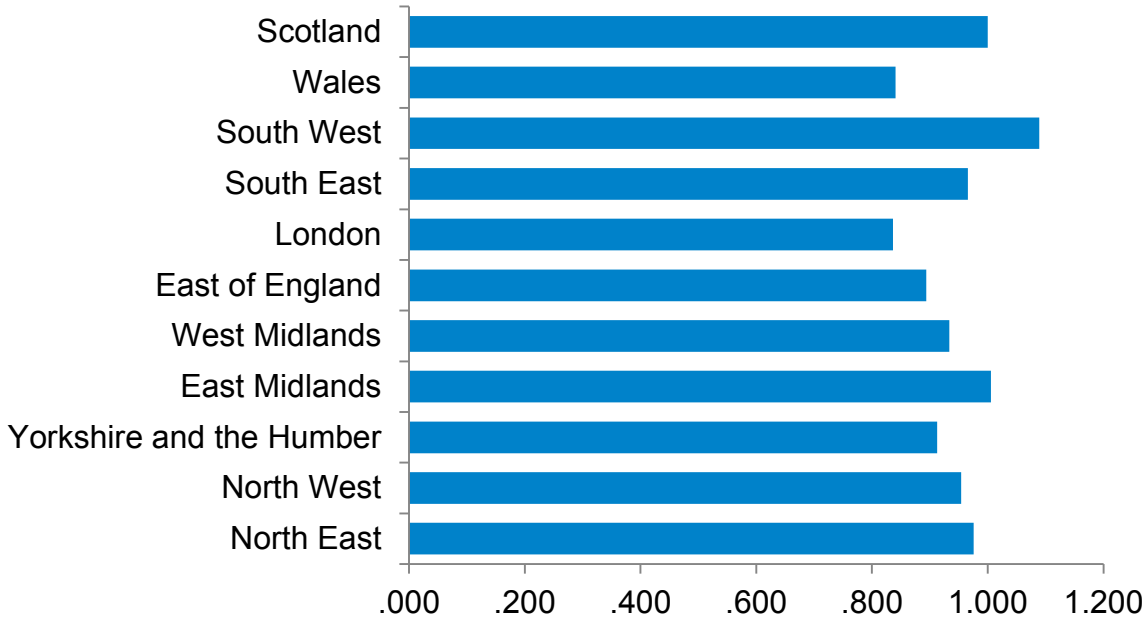
6. Weighting Scotland and Wales

6.1 Since 2013, DfT have only run the NTS in England, with the devolved administrations taking responsibility for running the survey in their own nations. However, DfT's pre-2013 NTS data does include Wales and Scotland, and their short walk trips should in principle be weighted to take account of under-reporting.

6.2 There is no information available on the relative likelihood of day 1 short walks being reported in Wales and Scotland, relative to day 7, as the two countries were not in the sample when the 2013 experiment was conducted.

6.3 However, it is possible to check the relative likelihood of the regions recording short walk trips, by running a logistic regression with car access, age/economic status and region on the pre-2013 data. The results for region are shown in the chart below. It can be seen that Scotland is above average with only the South West having a greater relative likelihood, while only London is below Wales. For Wales, the picture changes little when settlement type (i.e. how urban the area is) is taken into account. Unfortunately, it is not possible to test this for Scotland because a breakdown by settlement type is not available.

Relative likelihood of recording short walks: Scotland = 1.00



6.4 The evidence from the 2013Q2 trial in England is clear that short walk trips are likely to be substantially under-reported when collected on day 7. It is likely that the same applies in Wales and Scotland, although the degree of under-reporting may be different, and that the short walk trip data for these two countries should in principle be reweighted. However, before doing this, **R7: it is recommended that the NTS team should consider the need for Wales and Scotland to be weighted given that they are not included in the current NTS**. Essentially, there are two issues that need to be considered:

a. are time series to the present day now based just on England? If so, this suggests that weighting Wales and Scotland is not necessary;

b. how much interest is there in historic GB trends, and historic comparisons of regional data, including Scotland and Wales? Significant interest would suggest that weighting is required.

6.5 If it is decided to reweight Scotland and Wales, **R8: It is recommended that, in the absence of any other information, the weights should be the same as those used for England.** However, care should be taken when using these estimates, and **R9: it is recommended that users should be made aware of the limitations of these weights**.

6.6 An alternative approach for Scotland and Wales might be to use unweighted data for historic GB results but weighted data for English time series. The danger with this approach is that it may lead to unweighted GB results being compared against weighted England results.

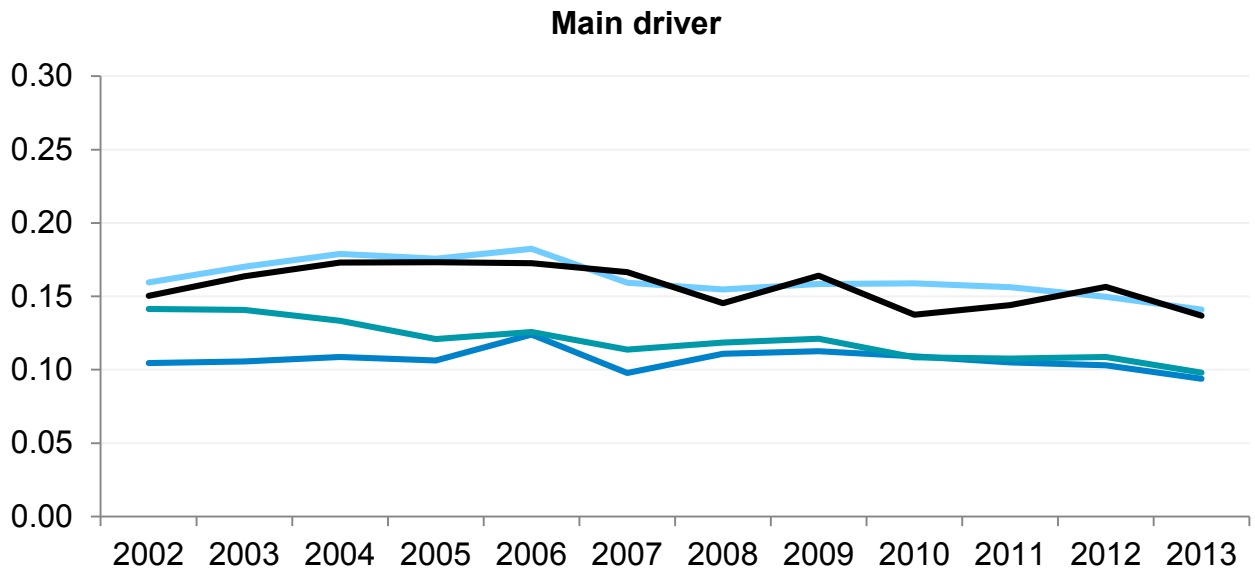
7. Conclusions and next steps

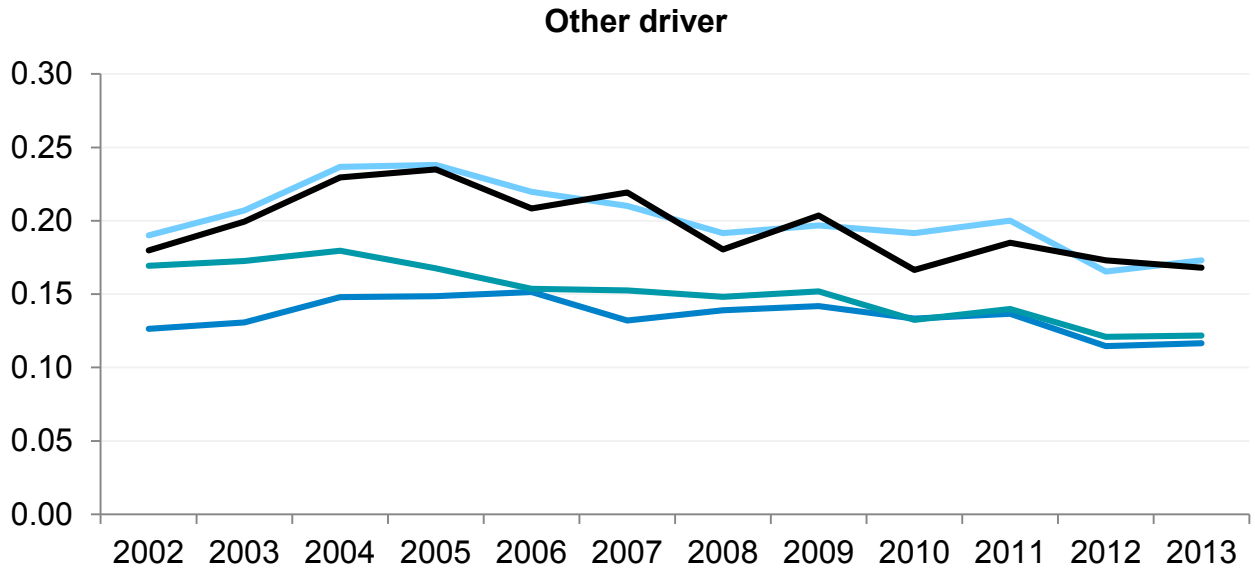
7.1 This report has found that it is possible to develop a methodology, using logistic regression models, to adjust for under-reporting of short walk trips on day 7 of the travel diary, based on the age/economic activity and access to a car of individual respondents. The repeat experiment in 2015Q2 will be used to confirm (or otherwise) the results of the 2013Q2 experiment, and their stability through time.

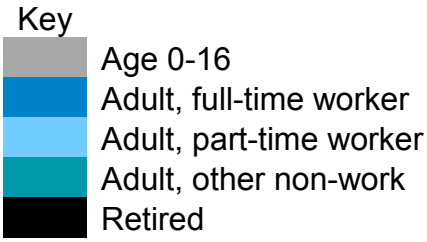
7.2 Analysis of the Q2 results between 2002 and 2013 shows that the likelihood of recording short walks on day 7, after controlling for age, economic status and vehicle accessibility, is fairly stable, suggesting that the weights derived from the 2013 Q2 (and 2015 Q2) experiment can be safely used. The main question mark is over the applicability of using results from Q2 at to other times of the year. To address this concern, it would be advisable to derive weights from a parallel data collection for a full year, split 50:50 between day 1 and day 7 diaries, during 2016/17. With careful selection of households, this could let the significance of the settlement variable in the logistic regression model be assessed, which could in turn lead to more sophisticated weights.

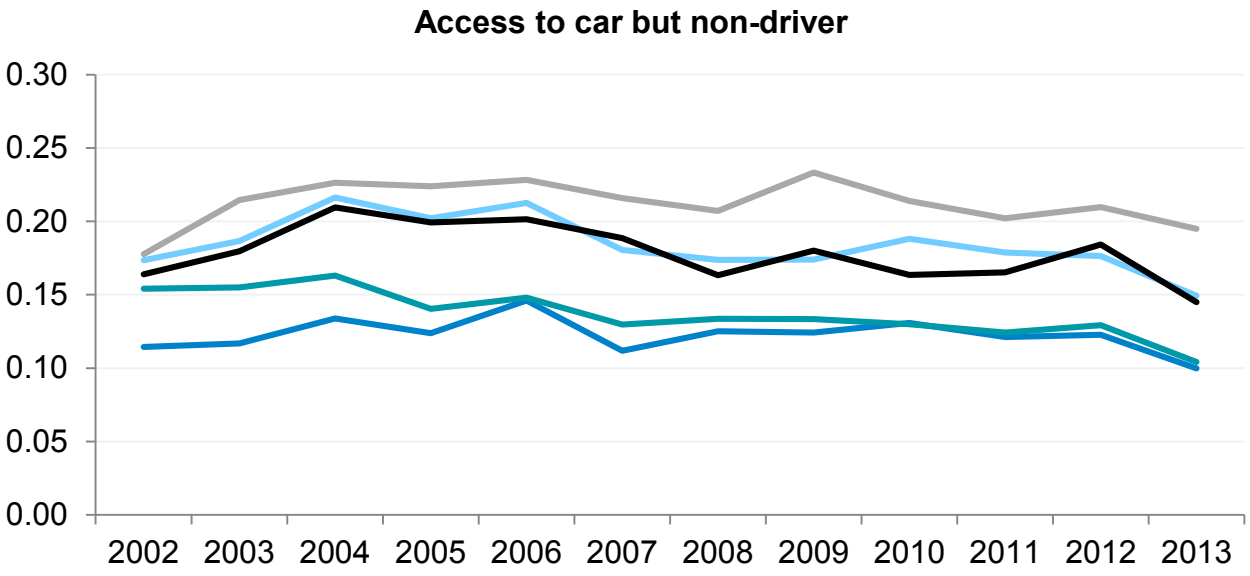
7.3 The situation with respect to Wales and Scotland is unsatisfactory, particularly as their likelihood of reporting short walk trips are at opposite ends of the spectrum. There is little that can be done about this, in the absence of any data for short walks on day 1, and there seems to be few options other than to apply the England-based weights, if it is decided that it is necessary to weight the results for Scotland and Wales.

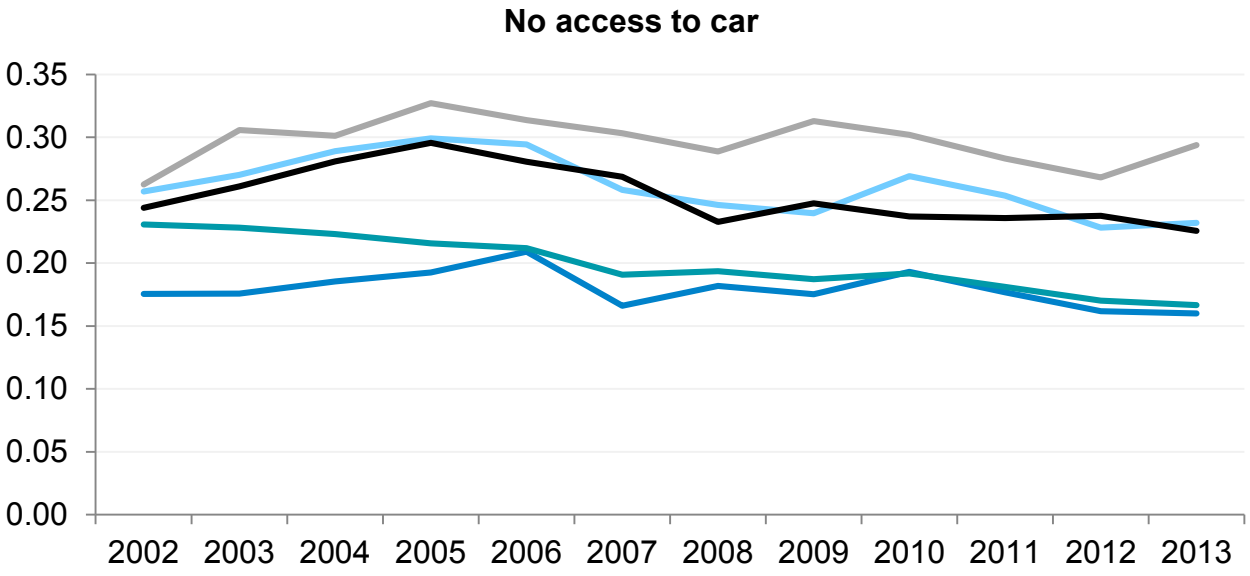
8. Annex: Probability of recording short walks by car access and age/economic status






Key
Age 0-16
Adult, full-time worker
Adult, part-time worker
Adult, other non-work
Retired


Access to car but non-driver
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013


No access to car
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013