Technical Report: Developing the datasets used in the modelling and analysis for the study:

Distribution of carbon emissions in the UK: Implications for domestic energy policy

Part I: Survey harmonisation (Patsios, D.)

Part II: Deriving emissions estimates from survey data (White, V., Thumim, J. and Brand, C.)

Part III: Working with imputed data (White, V.)

March 2013

This paper provides detail on the methodology applied in developing the datasets used in the study:

'Distribution of carbon emissions in the UK: Implications for domestic energy policy' *(Preston, I., White, V., Thumim, J., Bridgeman, T. and Brand, C. (2013) Research report funded by the Joseph Rowntree Foundation)*

The full project report is available at: http://www.jrf.org.uk/focus-issue/climate-change

Part I: Survey harmonisation

Technical Report Part I: Survey Harmonisation

Demi Patsios (University of Bristol)

1 Background

The United Kingdom has a wide range of Government surveys that provide sources of social and economic information. However, to date, no large scale UK dataset has been developed with the specific objective of understanding the profile of personal carbon emissions. The research project – '*Distribution of carbon emissions in the UK: Implications for domestic energy policy*'[1] - therefore combines data from a number of survey sources in order to advance knowledge in this area. The comparability of definitions and data in the different surveys is an important issue. Existing surveys have been designed at different times, to meet different needs, and have been commissioned by a range of departments. Consequently, the surveys were developed mostly in isolation from each other. This has undermined comparability with regard to concepts, definitions, design, fieldwork and processing practices ('inputs'), and also in the way results are released ('outputs').

The harmonisation of datasets was therefore essential in ensuring that key concepts are defined and measured in the same way, and to ensure that the data themselves were comparable with regard to their distribution and weighting. For example, the classification of age bands for household reference person (HRP) must be the same for each survey if the variable is to be used in the linear regression, CHAID and multiple imputation modelling later in the project. This report sets out in detail the survey harmonisation process.

2 Donor/recipient surveys

The original project proposal included four surveys which would be used for the harmonisation procedure.

- Expenditure and Food Survey (EFS)[2]
- National Travel Survey (NTS)
- English House Condition Survey (EHCS)
- International Passenger Survey (IPS)

Following a review of the IPS dataset, a decision was made by the research team to replace it with the Civil Aviation Authority's (CAA) Air Passenger Survey (APS), as the former offered very little in the way of socio-demographic variables, which are necessary for data matching and imputation. A brief overview of the four final survey datasets used in the harmonisation is provided below.

[1] Preston, I., White, V., Thumim, J., Bridgeman, T. and Brand, C. (2013) Research report funded by the Joseph Rowntree Foundation

[2] From January 2008, the EFS became known as the *Living Costs and Food* (LCF) module of the *Integrated Household Survey* (IHS).

2.1.1 Expenditure and Food Survey (EFS)

The EFS brings together two previously independent surveys: the Family Expenditure and National Food Surveys (FES and NFS)[3]. The EFS is a representative survey of around 7,000 UK households undertaken on a rolling annual basis by the Office for National Statistics and supported by the Department for Environment, Food and Rural Affairs (Defra). The main purposes of the EFS are to gather information for the Retail Price Index (RPI); National Accounts estimates of household expenditure; the analysis of the effect of taxes and benefits; and national patterns of food consumption and nutrition.

The EFS collects data through both householder questionnaires and a respondent diary-based approach. The former includes information about the household and large, infrequent expenditures (including on utility bills). The diary is completed by every householder aged 16 or over to record all daily expenditure over a two week period.

The datasets resulting from the EFS are published for financial years up until 2006, since when it has been published by calendar year. Datasets are published in 'raw' and 'derived' format, at the household level, which takes account of all expenditure of household occupants, and at the individual person level where individual expenditure and income (including from benefits) of each householder interviewed is documented.

The survey results comprise a wealth of socio-economic descriptive information on households, alongside weekly income (disposable and equivalised) and detailed weekly expenditure on a huge range of items. These items include the full range of household fuels, as well as petrol and diesel used in private road transport. Using region and time-specific fuel price data, it is possible to estimate household consumption of electricity, heating fuels and petrol and diesel for every case in the dataset. Using published carbon factors for each individual fuel (link to DECC/Defra figures) consumption values can be converted to carbon emissions.

However, the nature of the survey and diary approach is such that case level expenditure data (and thus the resulting consumption and carbon estimations) cannot necessarily be assumed to give accurate representation of annual consumption for any one household. The dataset is designed to be accurate at the aggregate level and 'clusters' can be created for the purpose of analysing policy impacts.

2.1.2 English House Condition Survey (EHCS)

The EHCS is a national survey of housing in England, commissioned by the department for Communities and Local Government (CLG). It covers all tenures and involves a physical inspection of property by professional surveyors. The information obtained through the survey provides an accurate picture of the type and condition of housing in England, the people living there, and their views on housing and their neighbourhoods.

The EHCS was originally carried out every five years until 2001. The EHCS now operates on a continuous format in which a population of approximately 8,000 dwellings is sampled in each year.

[3] From January 2008, the EFS became known as the *Living Costs and Food* (LCF) module of the *Integrated Household Survey* (IHS).

For the majority of analyses, including the monitoring of fuel poverty in England, two years' datasets are combined to give a dataset of approximately 16,000 dwellings / households.

In April 2008 the EHCS was merged with the Survey of English Housing to form the English Housing Survey (EHS). The physical survey, previously undertaken as part of the EHCS, is now part of the EHS.

The EHCS consists of a number of component surveys, including the interview survey, physical survey, market value survey and the private landlord survey. The interview and physical surveys contain the key categorical variables required for determining domestic emission abatement opportunities.

Interview Survey

The EHCS includes an interview with the householder. This covers: household characteristics, satisfaction with the home and the area, disability and adaptations to the home, work done to the property and income details.

The survey collects detailed information about the income of the Household Reference Person (HRP) and any partner from different sources (wages, pensions, benefits, savings and investments and other sources e.g. rent from property). Respondents are asked separately about each source and which, if any, state benefits they receive. This information is collated and modelled to produce total net income for the primary benefit unit (HRP and any partner/spouse). Less detailed information is collected from the respondent (who is always in the primary benefit unit) about the income of any additional benefit units in the household. The EHCS also collects information on housing and council tax benefit, council tax bands, ISMI and MPPI.

Physical Survey

The interview is followed by a visual inspection of the property, both internally and externally, by a qualified surveyor. Data collected includes the number and type of rooms and facilities contained in the property, the condition of a wide range of aspects of the physical structure, details of the heating systems, and parking provision, and assessment of neighbourhood quality. Where there are shared facilities and common areas, these are also assessed.

2.1.3 National Travel Survey

The National Travel Survey (NTS) is a survey of personal travel in Great Britain. The survey is continuous and has a sample size of around 8,000 private households. It provides information on the use of transport facilities by different sectors of the population, including: origin and destination of journeys, distance travelled, purpose and model of travel. It also includes socio-demographic information, at the household and individual level.

The NTS involves a face-to-face interview with householders and a diary, in which householders record all personal travel over a seven day period, including trip purpose, method of travel, time of day and trip length. The resulting datasets are published at various levels, including the household, person, journey stage and vehicle level.

2.1.4 CAA Air Passenger Survey

Since 1968 a series of surveys has been undertaken to obtain information about air travellers and the determinants of the travel market[4]. These surveys have been designed to obtain information of a kind that could not be collected on a routine basis from the air transport industry. The surveys have included questions on journey purpose, final and intermediate surface origins/destinations, means of transport to and from airports, route flown, country of residence and income. This information is used in assessing the type of market served by airports and consequently for forecasting air transport demand and for planning airport facilities.

CAA surveys, in general, operate for twelve month periods with anywhere from 4,000 to 500,000 interviews being achieved. All survey interviews are fed through a number of validation and editing software packages before being weighted to the true traffic levels. Surveys tend to run on an annual basis. The year was divided into twelve month-long sub periods. Shifts were selected so that during each sub period, each shift (e.g. Monday am) was sampled approximately the same number of times.

At all airports only departing passengers were interviewed, previous surveys having shown that differences in characteristics of arriving and departing passengers were not significant. Information is collected from air passengers with the aid of a skilled interviewing field force. Departing passengers are selected for interview using a strict sampling system. All passengers are eligible, apart from children under two years of age. An interview usually lasts between 5 and 7 minutes, and involves around 30 individual questions being asked. All of the information collected is based on face-to-face interviewing. A constant monitor of flights sampled is taken by supervisors, to ensure that as far as possible all routes, and in most cases individual flights, are covered regularly during each month. It is assumed that those passengers who are candidates for an interview, but who for one of a variety of reasons are not able to be interviewed, have the same characteristics as those who are successfully interviewed.

2.1.5 Obtaining a sample of CAA APS data

The CAA Air Passenger Survey data is not freely available. Due to project budget constraints it was not possible to obtain the full dataset of relevance to the research (i.e. UK passengers departing on international flights for non-business purposes). The project team therefore devised a 'random sampling' approach, which met with the constraints of CAA procedures. This necessitated applying a 'filter' to the CAA dataset to select two months at random (February and July). The following fields were included in the purchased dataset:

Table 1. Variables included in the CAA APS data sample for this study

• Survey airport	• Nationality
• Destination airport (immediate airport flying to)	• Population density
• Final destination airport (if not first destination)	• Gender
• Single or return	• Household Size
• Type of ticket	• Household annual income before tax
• Total cost of flight	• Socio-Economic Group
• Currency used to purchase ticket	• Age group of respondent
• Government office region	• No. children in household under 16

[4] http://www.caa.co.uk/default.aspx?catid=81&pagetype=90

3 The harmonisation procedure

The team needed to review each 'donor' survey to identify common variables and recode these (where possible) to match the recipient dataset (the Expenditure and Food Survey (EFS) for phase 1 of the study and the EHCS for phase 2). The following list contains a few examples of the variables in question:

- Income
- Dwelling type (i.e. detached, terrace etc.)
- Method of payment for fuel
- Main heating fuel
- Tenure
- Number of rooms in dwelling
- Household composition

A protocol for selecting and harmonising the various donor and recipient variables/measures was established early in the research period. The harmonisation procedure took part in three steps:

1) producing a 'crosswalk' of candidate variables available in each of the survey datasets;
2) coding household level socio-demographic variables necessary for the regression, CHAID and multiple imputation modelling;
3) coding household level energy consumption variables necessary for calculating carbon emissions.

These are explained in detail below.

3.1.1 Crosswalk of candidate variables available in each of the datasets

The crosswalk[5] (or comparison table) essentially involves building up a comprehensive picture of all potential variables for harmonisation. This was done by first creating variable groupings or 'sets', based upon past research carried out by CSE and the University of Bristol, as well as a review of more recent literature on carbon emissions in the UK. The following variable sets were used:

- Area
- Benefits
- Dwelling
- Employment
- Expenditure
- Household
- Income
- Journey
- Journey long distance
- Region
- Sample
- Socio-demographic
- Taxes
- Vehicle
- Weighting

The aim of the crosswalk was to create a set of mutually exclusive household and dwelling measures which could potentially be used in the imputation and modelling phases of the research.

[5] The crosswalk was created in MS Excel.

Each of the survey datasets was approached in turn, starting with matching or harmonising the NTS and EFS datasets. The EHCS was then matched to the combined NTS and EFS datasets and latterly the CAA APS dataset to all three. However, the process of producing the crosswalk was very much an iterative one, as new variables had to be added or removed following closer examination of the variable dictionaries of the individual datasets.

A section of the crosswalk, as it appears in the MS Excel spreadsheet, is shown below. This shows an example of the extent of different variables and measures that may be relevant to any one variable set (in this case the 'dwelling' set). Wherever possible, the actual variable name from the specific dataset was used in the crosswalk to avoid confusion as to which variable was chosen for the harmonisation.

Some of the surveys used in the study consist of multiple datasets. For example, the EFS includes both a person and household level dataset. In trawling through the various surveys and constituent datasets, we noted in which specific dataset the variable could be found (this is denoted by the letters in parentheses in table 1; see the legend for reference). Some variables, which may be defined in the user guides or survey materials, were not included in the publicly available datasets. These therefore needed to be accessed through a special request to the survey depositor namely ONS's Microdata Release Panel (MRP) in the case of the EFS and the Department for Transport for the NTS. In the case of the latter, these requests were quite straight forward and access was granted in a timely fashion. The EFS special access request took considerably longer. Any variables sourced in this way will not be included as part of the output dataset from this study due to restricted access issues.

Once it was established that there might be potential for harmonising variables at the household level across the datasets, the specific variable name (e.g. A116 and dwtypenx, for dwelling category in the EFS and EHCS respectively) was registered in the crosswalk. The definition of the variables (e.g. to establish whether the definitions of a 'household' or 'household reference person' were the same across the surveys) and the value categories (e.g. 1 = detached, 2 = semi-detached etc) were then explored more closely (as well as running raw frequencies) to ensure that we were comparing like with like. If there did not appear to be any notable differences in terms of the definitional values, the variable was considered a candidate measure for harmonisation.

Table 2. A sample of the Crosswalk showing: Variable set, variable and variable name

Variable set	Variable	Variable name	NTS 02-06	NTS 02-06	EFS 04-07	EFS 04-07	EHCS 07	EHCS 07
Dwelling	Dwelling- category	Dwelling category			A116	(h)	dwtypenx	(ph)
Dwelling	Dwelling- category type	Address Type (2000 onwards)	H12n	(h)				
Dwelling	Dwelling- characteristics bedrooms	Number of bedrooms					nbedsx	(i)
Dwelling	Dwelling- characteristics rooms	Rooms in accommodation - anonymised			a114p			
Dwelling	Dwelling- characteristics rooms	Rooms in accommodation - unanonymised			a114	(h) (ar)		
Dwelling	Dwelling- characteristics rooms occupied (de basis)	Number of rooms occupied - anonymised			p200p	(h)		
Dwelling	Dwelling- characteristics rooms occupied (de basis)	Number of rooms occupied - unanonymised			p200	(h) (ar)		
Dwelling	Dwelling- characteristics rooms shared	Rooms shared by household			A112	(h)		

Table 3. Crosswalk dataset legend

NTS/EFS	EHCS
ar=available by request	fs=fieldwork sample dataset
bu=benefit unit dataset	i=interview dataset
fp=fuel poverty dataset	g=general dataset
h=household dataset	ph=physical dataset
j=NTS journey dataset	e=employment dataset
ldj=NTS long distance journey dataset	inc=income dataset
p= person/individual dataset	p=people dataset
psu=NTS primary sampling unit dataset	a=amenity dataset
v = NTS vehicle dataset	

3.1.2 Coding household level socio-demographic variables

Once the crosswalk of candidate variables was ready, the research team selected a range of core measurement sets which would be needed (and available, ideally) across the datasets for the imputation and modelling phases of the research project (a summary of the core harmonised variables can be found in Appendix 2). The core harmonised variables included:

Table 4. Variables harmonised

- Household composition
- Number of adults
- Number of children
- Household size
- Number of pensioners
- Economic status of HRP
- NS-SEC of HRP
- Urban and rural areas
- Government Office Region (GOR)
- Relationship to HRP
- Household income (gross annual)
- Household income (net disposable annual)
- Individual income (gross)
- Age of HRP
- Gender of HRP
- Marital status and living arrangements
- Ethnic group of HRP
- Number of cars/light vans
- Petrol and Diesel consumed (cost)
- Accommodation type
- Tenure
- Number of workers in household
- Survey year
- Sample calendar year
- Sample month
- Sample quarter
- Sample weighting
- Unique household identifier
- Country

Not all core variables were harmonisation-ready or harmonisation-friendly. Some variables needed to be computed anew based on the combination of existing person-level or household-level variables and/or computed based on aggregating person-level data (or data at a level less than the household level, e.g. private vehicle use in the NTS) up to the household level. Any such computations/re-codes/aggregations have been noted in the SPSS syntax, which was used to create the harmonised data set. The key harmonisation challenges were encountered for the following variables:

- Household composition
- Economic status of HRP
- NS-SEC of HRP
- Urban and rural areas
- Household income
- Ethnic group of HRP
- Accommodation type
- Tenure

3.1.3 Additional data needed for modelling

In addition to harmonising the core socio-demographic variables, a number of other energy-consumption and CO_2 emissions-related variables, required for the imputation and modelling, needed to be harmonised. The approach taken to estimating CO_2 emissions for all household fuels

(EFS), domestic travel (NTS) and international aviation (CAA APS) are explained in detail Part II of this document.

As comparable variables are not available in all datasets dummy variables had to be created to enable the imputations to take place.

In addition, the following 22 variables needed for analysing policy impacts and household emission abatement opportunities (necessary for phase 2 of the study), were harmonised across the three datasets, using the EHCS survey as the reference dataset (dummy-coded for EFS, NTS and APS):

Table 5. Variables needed for policy and abatement modelling in the EHCS

- Dwelling age
- Type of wall insulation
- Wall type and insulation
- Loft insulation thickness
- Number of bedrooms
- Floor area
- Double glazing
- Rooms with low energy lighting
- Main heating system
- Main heating fuel
- Boiler group
- Age of heating system
- Energy efficiency (SAP05) rating
- Draughts and need for draught proofing
- Predominant type of roof covering
- Predominant type of roof structure
- Number of floors in flat
- Electrical system - off peak supply
- Number of dwellings in area
- Space heating (gjs)
- Water heating (gjs)
- Lights and appliances (gjs)
- Cooking (gjs)

4 Outcomes of the harmonisation procedure

A total of 288 variables were identified as candidates for harmonisation, of which 240 variables were harmonised (or dummy-coded) across the four datasets. Full details of the variables in the crosswalk and information on valid/missing data for each variable/measure are included in the Appendix at the end of this report.

The final harmonised data file is made up of a flat-file (in SPSS) comprising 107,933 households in the combined/harmonised dataset, made up of:

45,750 households spanning 5 years (2002-2006) of the NTS
24,649 households spanning 4 years (2004-2007) of the EFS
15,604 households spanning 3 years (2006-2008) of the EHCS
21,930 interviewees spanning 5 years (1999, 2005-2008) of the CAA APS

Table 6. Sample size and years of each survey dataset used

	NTS	EFS	EHCS	APS	TOTAL
1999	0	0	0	291	291
2002	8,567	0	0	0	8,567
2003	9,208	0	0	0	9,208
2004	8,968	5,072	0	0	14,040
2005	9,517	6,696	0	4017	20,230
2006	9,223	6,745	5,800	2975	24,743
2007	267	6,136	7,744	1866	16,013
2008	0	0	2,060	12781	14,841
Total	**45,750**	**24,649**	**15,604**	**21,930**	**107,933**

Part II: Deriving CO_2 emissions estimates from survey data

Technical Report Part II: Deriving CO_2 emissions estimates from survey data

Vicki White (Centre for Sustainable Energy)
Christian Brand (University of Oxford)
Joshua Thumim (Centre for Sustainable Energy)

This technical report provides detail to accompany the main project report for the JRF-funded study: *Distribution of carbon emissions in the UK: Implications for domestic energy policy.*

As described in Part I of this report, the study combines data from four different surveys to create two final comprehensive datasets[6] of household carbon dioxide emissions from energy consumption in the home and all personal travel, including by public transport and domestic and international aviation. Three separate datasets provide the source data for these carbon emissions estimates, as shown below. This report describes the methodology behind deriving estimated household CO_2 emissions from data contained in each survey. As the surveys used in this study differ in design and purpose, different methods and approaches had to be developed for each.

Table 7. Source surveys for deriving estimates of household carbon emissions

Survey	Input (raw survey data)	Output
1. Expenditure and Food Survey (EFS)	Expenditure on all household fuels	Annual consumption of all household fuels (kWh) and associated CO_2 emissions ($kgCO_2$)
2. National Travel Survey (NTS)	Private vehicle mileage Distance travelled public transport Distance travelled domestic flights	Annual CO_2 emissions from all personal (non-business) travel by private vehicle, public transport and domestic aviation
3. CAA Air Passenger Survey (APS)	Start airport, destination airport (international only) and flight class for all UK leisure passengers	Distance travelled and associated CO_2 emissions from (non-business) international aviation

EFS: Deriving energy consumption and emissions from expenditure on household fuels

1.1 Deriving energy consumption from energy expenditure

The Expenditure and Food Survey (EFS)[7] uses interview and diary-based methods to collect information on a range of personal and household expenditure, including on household fuels (Table 8). For the purpose of the research, data from four EFS datasets were combined, (financial years

[6] One representative of households in Great Britain and the second representative of households in England.
[7] In April 2001 the *Family Expenditure Survey* (FES) and *National Food Survey* (NFS) were combined to form the *Expenditure and Food Survey* (EFS), which completely replaced both series. From January 2008, the EFS became known as the *Living Costs and Food* (LCF) module of the *Integrated Household Survey* (IHS).

2004/5, 2005/6, and calendar years 2006 and 2007) generating a sample size of over 24,000 cases. Values for annual energy expenditure for each case (household) in the EFS were converted to energy use using survey variables method of payment (for electricity and mains gas), month and year, combined with a series of look-up tables containing time- and location-specific fuel price data. A look-up table was devised for each fuel and, where applicable, method of payment (see Table 9 for example). Queries were scripted in Microsoft sql to join the fuel unit price from the look-up tables to the EFS dataset on year, month and government office region. Annual energy consumption was then calculated for each fuel ((expenditure on fuel * 52.14)/unit price [p/kWh]).

Where a household has electric central heating, the unit price of electricity applied to derive annual consumption was reduced by a factor of 0.7[8], to allow for the a likely lower unit rate (households using electricity for heating are more likely to be on special tariffs, e.g. Economy 7). Without applying this factor, we would risk underestimating electricity consumption in electrically heated households.

Table 8. EFS expenditure variables for deriving household fuel consumption

Variable name	Label
B175	Electricity amount paid in last account
C45114t	Electricity slot meter payment
C45112t	Second dwelling: electricity account payment
B222	Elec - Amt of payment (budget scheme)
B170	Gas amount paid in last account
C45214t	Gas slot meter payment
C45212t	Second dwelling: gas account payment
B221	Gas - Amt of payment (budget scheme)
B018	Bottled gas for central heating
C45222t	Bottled gas - other
C45411t	Coal and coke
B017	Oil for central heating - last quarter
C45312t	Paraffin
C45412t	Wood and peat
C45511t	Hot water, steam and ice

Table 9. Section of look-up table for converting survey expenditure data to consumption (for mains gas by direct debit)

Gorx	Government Office Region	EFS Year	Actual Year	Year_code	EFS_Month	Date	Fuel	Fuel	MOPn	MOP	p/kWh
1	North East	2004/05	2004	2004	4	Apr-04	2	Gas	Direct Debit	1	0.0172
3	Yorkshire & Humberside	2004/05	2004	2004	4	Apr-04	2	Gas	Direct Debit	1	0.0171
2	North West	2004/05	2004	2004	4	Apr-04	2	Gas	Direct Debit	1	0.0171

[8] This figure was derived from the average ratio of off-peak tariffs to standard electricity rates, as provided by Sutherlands tables data.

1.2 Converting energy consumption to carbon emissions estimates

Annual consumption estimates for each fuel were then converted to carbon using Defra/DECC CO_2 emissions factors[9] (on a net calorific value basis – Table 10). For electricity, the 5-year rolling average was applied. All calculations were scripted in SPSS syntax.

Table 10. Carbon emissions factors applied for household fuels in the EFS

EFS_fuel	kg CO_2 per kWh
Electricity	0.541600
Gas (mains)	0.205150
Bottled gas	0.223570
LPG	0.229990
Coal	0.311390
Oil (for CH)	0.258570
Paraffin	0.259000
Wood	0.024563

NTS: Estimating household level emissions from private and public transport and domestic air travel

The National Travel Survey (NTS) provides information on all journeys, defined by purpose and mode, for all personal and business travel by private and public transport and domestic aviation for GB households.

This project is principally concerned with personal carbon emissions, and the potential for policy influence. It was therefore deemed appropriate to focus on all non-business travel, and maintain a distinction between commuting and leisure. However, for modelling purposes, it is also preferable to limit the number of variables to be imputed. It was therefore agreed by the project team that the following five variables would be derived from NTS travel data, for later imputing to the EFS for distributional analysis:

- Private car - Commute CO_2 emissions
- Private car - Leisure CO_2 emissions
- Public transport - Leisure CO_2 emissions
- Public transport - Commute CO_2 emissions
- All non-business domestic air travel CO_2 emissions

The NTS is made up of a number of different datasets, from the household level (providing socio-demographic information) to the individual, journey, stage and vehicle level. For the purpose of estimating CO_2 emissions from personal travel, as detailed above, three different datasets were used:

- the vehicle dataset: for private vehicle travel
- the journey stage dataset: for public transport

[9] 2010 Guidelines to Defra / DECC's GHG Conversion Factors for Company Reporting

- the long-distance journey dataset: for domestic aviation.

1.3 Calculating emissions for private vehicle travel

Whilst the NTS journey stage dataset provides detailed information on all private vehicle travel, it lacks detailed information about the vehicle used[10]. The vehicle dataset provides detailed information about the vehicle type, its engine size, fuel type and annual mileage, split by purpose (commuting, private and business). For the purpose of calculating CO_2 from private vehicle use for this study, two alternative methods were considered. Both are based on the common travel activity method, which converts travel activity (e.g. distance travelled) into travel emissions using vehicle emissions factors (e.g. $kgCO_2$ per distance travelled):

1. **Method A** uses 'official' vehicle emissions factors specific to the make, model, fuel type, engine size, transmissions type and age of the vehicle. These are linked to the CO_2-graded taxing of vehicle use (VED) and give the best resolution of emissions factors. However, they are based on test-cycle conditions, so do not include an uplift for real world driving conditions. The other caveat is that no detailed emissions factors are available for cars and motorcycles registered for the first time before 2001. As a result, the 2002-2006 NTS dataset contains only 25% of 'official' CO_2 emissions factors.
2. **Method B** uses banded vehicle emissions factors that are based on estimated average values for the UK car fleet in 2009 travelling on average trips in the UK (Defra & DECC, 2009). They are calculated based on data from SMMT (Society of Motor Manufacturers and Traders) on new car CO_2 emissions from 1997 to 2009 combined with factors from TRL (Transport Research Laboratory) as functions of average speed of vehicle derived from test data under real world testing cycles and an uplift of 15% to take into account further real-world driving effects on emissions relative to test-cycle based data. The emissions factor set is complete but lacks the resolution of Method A.

Given the obvious drawbacks of Method A the team decided to use banded vehicle emissions factors (Method B), with the Defra/DECC greenhouse gas (GHG) conversion factors as the primary source[11]. To avoid double counting all private vehicle emissions were allocated to the driver. Table 11 lists the NTS variables that were used in combination with Defra emissions factors. All calculations were scripted in SPSS syntax. The descriptive statistics of vehicle CO_2 emissions and comparison of the two CO_2 emissions factor calculation methods for the NTS 2002-06 dataset are shown in Table 12. The CO_2 calculations were done at vehicle level and then aggregated to NTS household level using a new household ID (newhhid) and survey variables H88 (household reference number), psuid (PSU reference number), H96 (quota allocation year).

Table 11. 'Vehicle' variables used to estimate CO_2 emissions from private vehicle use

Variable name	Variable label
V139	Annual vehicle commuting mileage (unbanded)

[10] Although the NTS travel diary forms ask for a cross reference to the vehicle information provided by the respondent, this cross referencing information was not included in the NTS dataset we received.

[11] Defra & DECC, 2009. Guidelines to Defra/DECC's GHG Conversion Factors for Company Reporting. Annex 6, Passenger Transport. http://www.defra.gov.uk/environment/business/reporting/pdf/20090928-guidelines-ghg-conversion-factors.xls

V143	Annual vehicle private mileage (unbanded)
V86	Engine capacity (cc) (unbanded)
V3b	Type of Vehicle
V88	Type of fuel

Table 12. Descriptive statistics of vehicle CO_2 emissions and comparison of the two CO_2 emissions factor calculation methods based on the NTS 2002-2006 dataset

V3B - Type of Vehicle		Method A: official CO_2 factor method		Method B: Defra e-factor method	
		V139 & annual CO_2 from commuting [kg]	V143 & annual CO_2 from private/leisure travel [kg]	V139 & annual CO_2 from commuting [kg]	V143 & annual CO_2 from private/leisure travel [kg]
Car	N	12319	12278	45623	45335
	% of Total N	99.5%	99.5%	98.6%	98.6%
	Mean	910.62	1566.58	906.53	1698.98
	Std. Error of Mean	12.919	11.046	6.860	6.710
	Std. Deviation	1433.870	1223.982	1465.195	1428.655
	Sum	11217885	19234439	41358659	77023094
	% of Total Sum	99.1%	99.2%	97.3%	98.9%
	Maximum	19984	19115	24700	27397
Motorcycle / Scooter / Moped	N	1	1	132	131
	% of Total N	.0%	.0%	.3%	.3%
	Mean	463.39	463.39	314.71	426.39
	Std. Error of Mean	.	.	48.979	44.343
	Std. Deviation	.	.	562.721	507.534
	Sum	463	463	41541	55857
	% of Total Sum	.0%	.0%	.1%	.1%
	Maximum	463	463	3313	2982
Landrover, Jeep	N	61	60	226	217
	% of Total N	.5%	.5%	.5%	.5%
	Mean	1557.49	2622.07	1494.66	2500.46
	Std. Error of Mean	259.142	312.354	136.172	144.538
	Std. Deviation	2023.963	2419.483	2047.117	2129.173
	Sum	95007	157324	337794	542599
	% of Total Sum	.8%	.8%	.8%	.7%
	Maximum	10961	14755	11487	15463
Light Van	N	3	3	243	241
	% of Total N	.0%	.0%	.5%	.5%
	Mean	723.51	249.40	2795.82	798.83
	Std. Error of Mean	525.629	249.395	263.414	108.237
	Std. Deviation	910.416	431.965	4106.224	1680.291
	Sum	2171	748	679385	192518
	% of Total Sum	.0%	.0%	1.6%	.2%
	Maximum	1746	748	26228	15737
Total private	**N**	**12386**	**12343**	**46279**	**45977**
	% of Total N	**100.0%**	**100.0%**	**100.0%**	**100.0%**
	Mean	**913.76**	**1571.21**	**918.31**	**1693.78**
	Std. Error of Mean	**12.918**	**11.113**	**6.981**	**6.710**
	Std. Deviation	**1437.681**	**1234.615**	**1501.841**	**1438.707**
	Sum	**11317893**	**19393419**	**42498252**	**77874861**
	% of Total Sum	**100.0%**	**100.0%**	**100.0%**	**100.0%**
	Maximum	**19984**	**19115**	**26228**	**27397**

1.4 Calculating emissions for travel by public transport

The NTS journey 'stage' dataset provides the most detailed information on personal travel patterns. It includes each individual stage of a trip, whereas the 'journey' dataset captures the main stage only (for example, a journey may consist of a 1 mile walk to a bus stop; a 5 mile bus ride; and a 40 mile train journey; each of these represents a 'stage', but in the NTS journey dataset the trip would be classed as a single train journey). The NTS 'stage' dataset was therefore considered most appropriate for estimating CO_2 emissions from travel by public transport.

However, the stage dataset does not include 'purpose' (needed to distinguish between leisure and commute) as this is a journey-level variable. The overall journey purpose (J28A - Journey Purpose) for each stage was therefore extracted from the NTS journey dataset. This was done in SPSS syntax, by first creating a new journey ID in both datasets, and then merging on this, taking variable J28a from the 'journey' to the 'stage' dataset.

To be consistent in our emissions factor methodology, we again used Defra/DECC CO_2 emissions factors[12], in combination with the following variables, to estimate emissions for each stage.

Table 6. 'Stage' and 'journey' variables used to estimate CO_2 emissions from public transport

Variable name	Variable label
SD	Stage Distance (tenths miles - unbanded - grossed up for short walks)[13]
S24	Means of travel
J28A	Journey Purpose

All calculations were scripted in SPSS syntax and carried out at the journey stage level. The results were then weighted (by w5 – journey weight) and aggregated to NTS household level, using a new household ID (newhhid) and survey variables H88 (household reference number), psuid (PSU reference number), H96 (quota allocation year).

1.5 Calculating emissions for domestic aviation

The NTS journey stage dataset includes data on domestic air travel. However, this data is based on a two week travel diary, within which long-distance trips are limited in occurrence. To overcome this issue, and improve the coverage of these less common trips, the NTS collects additional data on 'long-distance journeys', including those by air, via an interview. Respondents are asked about all trips of 50 miles or more (within GB), made in the three weeks preceding the travel diary week. This information is recorded in the long-distance journey (LDJ) dataset and was considered most appropriate for the purpose of estimating CO_2 emissions from domestic aviation.

The following variables (table 7) were used in combination with Defra/DECC emissions factors to estimate CO_2 emissions for all non-business domestic air travel. Domestic aviation emissions are not split by purpose, as with private car and public transport travel, (i.e. by commute and leisure), due to the overall small sample size, particularly in relation to the former. All calculations were scripted in

[12] Defra & DECC, 2009. Guidelines to Defra/DECC's GHG Conversion Factors for Company Reporting. Annex 6, Passenger Transport. http://www.defra.gov.uk/environment/business/reporting/pdf/20090928-guidelines-ghg-conversion-factors.xls

[13] Stage distance was converted to kilometres to correspond with the Defra emissions factors.

SPSS syntax and carried out at the long-distance journey dataset level. The results were then weighted (by w4 – long-distance journey weight) and aggregated to NTS household level, using a new household ID (newhhid) and survey variables H88 (household reference number), psuid (PSU reference number), H96 (quota allocation year).

Table 13. 'Long-distance journey' variables used to estimate CO_2 emissions from domestic aviation

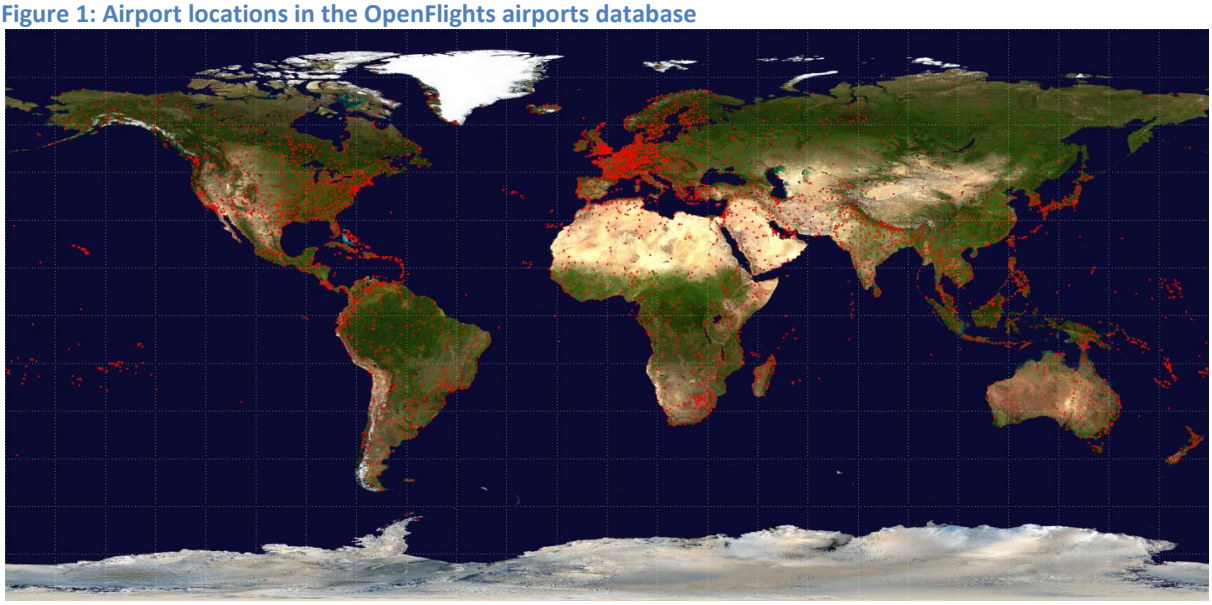
Variable name	Variable label
L9	LDJ Journey Distance[14]
L8	LDJ Mode of Transport
L12A	LDJ Overall Journey Purpose

CAA APS: Computing distance travelled and CO_2 for international air travel

The Civil Aviation Authority Air Passenger Survey (CAA APS) is an annual survey (with data collected on a monthly basis) of departing passengers, covering a wide range of UK airports. The survey collects data on passenger travel (e.g. flight origin and destination) and socio-demographics. For the purpose of modelling carbon emissions from international air travel for this project, a subset of the data was purchased from CAA to include only UK passengers, travelling outside of the UK, for leisure purposes. Data was purchased for all airports surveyed from 2005 – 2008, but due to budget constraints, the sample size had to be limited to two randomly selected months.

To calculate CO_2 emissions, air distance travelled had to first be derived from the flight origin/destination data. First, the team obtained the most up-to-date airports database from OpenFlights.org, a free flight mapping, search, statistics and sharing tool as well as the name of the open-source project to build the tool. As of October 2009, the OpenFlights airports database contained 5,497 airports spanning the globe, as shown in the map below.

Figure 1: Airport locations in the OpenFlights airports database



Source: OpenFlights.org (http://openflights.org/data.html)

[14] Long distance journey distance was converted to kilometres to correspond with the Defra emissions factors.

Amongst other variables, the database contains airport names, country, the 3-letter IATA/FAA code necessary for matching airport names contained in the CAA APS dataset data, and latitude and longitude of the airport.

All cases (journeys) in the CAA APS were first converted to stages, to take account of all individual legs of the journey. An SQL query was then created to compute great circle distance (GCD, in kilometres) for every stage, by matching origin and destination airport codes contained in the CAA APS dataset[15] with the airports database and applying the following code[16]:

$$theta = lon2 - lon1$$
$$dist = acos(sin(lat1) \times sin(lat2) + cos(lat1) \times cos(lat2) \times cos(theta))$$
$$if\ (dist < 0)\ dist = dist + pi$$
$$dist = dist \times 6371$$

This was done for every stage and it was assumed that all journeys were return, hence the calculated distance was doubled. A 50km 'distance penalty' was added to each flight leg to account for real world routing (i.e. not following the straight line of great circle distances between origin and destination) and detours[17].

To be consistent in our emissions factor methodology for other modes of travel, Defra/DECC CO_2 emissions factors (table 6l[18]) were again used in deriving CO_2 emissions from the calculated air distance. For international air travel, different factors are provided according to cabin class. The CAA APS data includes information on 'Ticket type'. Each category of ticket type in the APS dataset (n=43, including 'missing') were matched to the Defra/DECC cabin class definitions. In addition, distance thresholds were used to identify 'short distance', 'short haul' and 'long haul' journeys in the APS, creating a look-up table, summarised below and provided in full in the Appendix.

Table 14. CO_2 emission factors applied for international air travel[18]

Flight type (flight distance threshold)	Cabin class	kg CO_2 per pkm
Domestic[19] (short distance under 750km per leg)	All – Average	0.17147
Short-haul international (750km – 3,700km per leg)	Average	0.09700
	Economy class	0.09245
	Business class	0.13867
Long-haul international (over 3,700km per leg)	Average	0.11319
	Economy class	0.08263
	Premium economy class	0.13221
	Business class	0.23963
	First class	0.33052

[15] There were 32 airport codes in the CAA APS, which were not contained in the database. The airport name, latitude and longitude for these codes were sought manually. There were a further 48 cases in the CAA APS where only the country/state code was provided. In these cases, the location of the largest (in terms of flight frequency) airport for the location was used.
[16] Assuming the Earth is a perfect sphere of radius 6371 km
[17] Based on Lufthansa, 2002. Audit, data and facts 2001/2002, Frankfurt, Lufthansa. www.lufthansa.com
[18] Defra & DECC, 2010. Guidelines to Defra/DECC's GHG Conversion Factors for Company Reporting. Annex 6, Passenger Transport. http://www.defra.gov.uk/environment/business/reporting/pdf/100805-guidelines-ghg-conversion-factors.xls.
[19] For international flights over very short distance emissions are significantly greater than for 'short haul', hence the domestic emissions factor is applied here. Examples include flights from Manchester to the Republic of Ireland, or London to Paris or Brussels.

CO_2 emissions were derived for each leg of the air journey and then aggregated to give an estimation of total CO_2 emissions for every record in the CAA APS dataset.

The CAA APS data is provided with a weighting ('population') to inflate survey data to annual estimates. However, as we only have a two-month sample of the full survey data, the resulting values have to be further inflated. The team considered a range of approaches to do this.

Option 1. At the most simplistic level the data may be inflated using a ratio of the number of days in the sample year to days in the sample months (approximately a factor of 6 – see Table 15). However this assumes the months of February and July are an equal and representative $^1/_{12}$ of all annual departures. This results in an overestimation of the total number of (UK international leisure) passengers from each airport, compared to figures reported by CAA in its annual passenger survey reports[20]. Whilst this method was applied in generating some interim descriptive results, it was considered overly simplistic for application in the final analysis.

Option 2. Using figures published in CAA's annual passenger survey reports, it is possible to calculate an annual multiplying factor for each individual airport based on the number of records in our (weighted) sample and CAA reported total departures by airport for UK international leisure passengers[21] (Table 16). This takes into account multiple month samples for smaller airport as advised by CAA (and therefore corrects for the slight overestimate of inflating by a factor of 6), but does not enable us to differentiate between the two months.

Option 3. A final option which addresses the limitations of the two approaches above is to calculate a multiplying factor for every airport month based on the monthly count of departures from all airports in the sample and the annual total as reported in the CAA passenger survey reports. This is the preferred option and has been applied to generate multiplying factors as shown in Table 17 below.

Table 15. Inflating CAA APS data by number of days in month

N_days	Yrs 2005 - 2007	Yr 2008
February	28	29
July	31	31
Total days	59	60
Multiplying factor	***6.186***	***6.100***

Table 16. Inflating CAA APS data by airport

Airport	Weighted count (2 month sample)	CAA Reported Counts[20]20	Multiplying factor
ABERDEEN	48,980	237,000	4.84
BELFAST INTERNATIONAL	302,552	1,285,000	4.25
BELFAST CITY (GEORGE BEST)	3,451	17,000	4.93
BIRMINGHAM	1,031,330	5,226,000	5.07
BOURNEMOUTH	72,778	443,000	6.09
BRISTOL	751,605	3,886,000	5.17
CARDIFF WALES	218,086	1,252,000	5.74

[20] http://www.caa.co.uk/default.aspx?catid=81&pagetype=90&pageid=7640

[21] Table 4a. Characteristics of terminal passengers at the survey airports

DONCASTER SHEFFIELD	123,044	814,000	6.62
EDINBURGH	257,281	1,190,000	4.63
EAST MIDLANDS INTERNATIONAL	534,054	3,400,000	6.37
EXETER	82,558	491,000	5.95
GLASGOW	669,627	3,150,000	4.70
HUMBERSIDE	55,671	311,000	5.59
INVERNESS	2,665	3,000	1.13
LEEDS BRADFORD	312,927	1,475,000	4.71
LONDON CITY	93,153	616,000	6.61
CITY OF DERRY (EGLINTON)	2,804	16,000	5.71
GATWICK	3,564,278	19,519,000	5.48
HEATHROW	2,837,384	17,213,000	6.07
LIVERPOOL	541,568	2,906,000	5.37
LUTON	895,563	4,998,000	5.58
MANCHESTER	2,561,595	13,431,000	5.24
DURHAM TEES VALLEY	101,333	522,000	5.15
NEWCASTLE	513,847	2,665,000	5.19
NORWICH	40,766	165,673	4.06
PRESTWICK	182,044	938,000	5.15
STANSTED	1,656,180	9,495,000	5.73

Table 17. Inflating CAA APS data by airport and month

Airport	Feb weighted count	July weighted count	Feb Factor	July Factor	CAA reported counts
ABERDEEN	12,664	36,315	9.36	3.26	237,000
BELFAST INTERNATIONAL	82,311	220,241	7.81	2.92	1,285,000
BELFAST CITY (GEORGE BEST)	0	3,451	0.00	4.93	17,000
BIRMINGHAM	333,401	697,929	7.84	3.74	5,226,000
BOURNEMOUTH	0	72,778	0.00	6.09	443,000
BRISTOL	319,529	432,076	6.08	4.50	3,886,000
CARDIFF WALES	58,469	159,617	10.71	3.92	1,252,000
DONCASTER SHEFFIELD	40,926	82,117	9.94	4.96	814,000
EDINBURGH	72,049	185,232	8.26	3.21	1,190,000
EAST MIDLANDS INTERNATIONAL	200,281	333,773	8.49	5.09	3,400,000
EXETER	23,790	58,769	10.32	4.18	491,000
GLASGOW	166,498	503,129	9.46	3.13	3,150,000
HUMBERSIDE	9,929	45,742	15.66	3.40	311,000
INVERNESS	0	2,665	0.00	1.13	3,000
LEEDS BRADFORD	139,158	173,769	5.30	4.24	1,475,000
LONDON CITY	45,230	47,923	6.81	6.43	616,000
CITY OF DERRY (EGLINTON)	267	2,537	29.97	3.15	16,000
GATWICK	1,285,326	2,278,952	7.59	4.28	19,519,000
HEATHROW	1,417,645	1,419,739	6.07	6.06	17,213,000
LIVERPOOL	207,742	333,826	6.99	4.35	2,906,000
LUTON	360,916	534,648	6.92	4.67	4,998,000

MANCHESTER	894,720	1,666,874	7.51	4.03	13,431,000
DURHAM TEES VALLEY	38,566	62,767	6.77	4.16	522,000
NEWCASTLE	160,488	353,359	8.30	3.77	2,665,000
NORWICH	6,409	34,357	12.92	2.41	165,673
PRESTWICK	69,524	112,520	6.75	4.17	938,000
STANSTED	629,854	1,026,326	7.54	4.63	9,495,000

Part III: Deriving final emissions estimates from imputed data

Technical Report Part III: Deriving final emissions estimates from imputed data

Vicki White, Centre for Sustainable Energy

Introduction

This report describes in detail the steps taken to derive final estimates of household emissions following the multiple imputation process. This is the final stage in the method applied in developing the datasets used in the analysis for the JRF-funded study: *Distribution of carbon emissions in the UK: Implications for domestic energy policy.*

As is outlined in the main project report, with further detail provided in this document, the key stages in the methodology include: harmonising surveys; deriving estimates of household emissions from survey data; and applying multiple imputation models to impute emissions estimates from one survey to another.

The imputation process seeks to replicate the distribution of household emissions based on harmonised variables in the recipient and donor surveys. It does not, however, seek to replicate the sum total of emissions from the donor survey. The team therefore developed a methodology to adjust or "re-gross" the imputed values in the recipient surveys (i.e. the EFS and EHCS) such that the sum total corresponds with that of the donor survey (for the equivalent population – Great Britain for the EFS, England only for the EHCS).

Whilst the re-grossing process is immaterial in terms of the distribution of emissions per se, it is of fundamental importance for the modelling of policy impacts (for example, for modelling carbon taxation). The process applied in grossing each of the imputed variables is described below. The approach was slightly different for the EFS and EHCS datasets, due to different imputation models being applied to each.

Methodology

Below is a detailed outline of all the steps applied for adjusting imputed data to match the survey from which it was derived.

1. Transform logged values:

The first step in deriving meaningful values of household carbon emissions following the imputation is to take the inverse log of imputed values (values are logged to give a normal distribution required for multiple imputation, see Table 18 and Table 19).

Table 18. Variables imputed to the EFS

Donor Survey	Imputed Variable name	Imputed Variable label	Unlogged variable name
NTS	NTSTotal	Log of total CO2 from NTS Travel	NTStotalCO2
NTS	CarComute	Log of Private car - Commute CO2 emissions	CarCommuteCO2annual
NTS	CarLeisure	Log of Private car - Leisure CO2 emissions	CarLeisureCO2annual
NTS	PTleisure	Log of Public transport - Leisure CO2 emissions	PTLeisureCO2
NTS	PTcomute	Log of Public transport - Commute CO2 emissions	PTCommuteCO2
NTS	UKair	Log of Air travel CO2 (domestic)	DomesticAirCO2

APS	logkgCO2_aps	Log of International Air Leisure flight CO2 (APS)	IntlAirCO2annual
APS	Log_EU_FlightCO2	Log of Total CO2 from International Air Leisure Travel (APS) to EU Destinations	EU_FlightCO2
APS	Log_Non_EUFlightCO2	Log of Total CO2 from International Air Leisure Travel (APS) to Non-EU Destinations	Non_EUFlightCO2
EFS	hhfuel	LN(C02 hhld fuel) minus zeros/negatives	hhfuel

Table 19. Variables imputed to the EHCS

Donor Survey	Imputed Variable name	Imputed Variable label	Unlogged / Renamed* variable name
NTS	nV139C2	Private car - Commute CO2 emissions	CarCommuteCO2annual
NTS	nV143C2	Private car - Leisure CO2 emissions	CarLeisureCO2annual
NTS	nPT_CO2_leisure	Public transport - Leisure CO2 emissions	PTLeisureCO2
NTS	nPT_CO2_commute	Public transport - Commute CO2 emissions	PTCommuteCO2
NTS	nldjAirCO2_private	Air travel CO2 (domestic)	DomesticAirCO2
NTS	ncars	Number of cars/vans (harmonised)	
APS	logkgCO2_aps	Log of International Air Leisure flight CO2 (APS)	IntlAirCO2annual
APS	intflight	Non-flyer (Dummy variable to identify non-flyers – if Intflight = 0 then compute logkgCO2_aps=0	
APS	nshort	Number of short haul flights to Europe in past year	
APS	nlong	Number of long haul flights further than Europe in past year	
EFS	LN_Heat_load	Log of CO2 from Heat Load	CO2_Heat_load
EFS	LN_Power_Load	Log of CO2 from Power Load	CO2_Power_Load
EFS	appliances	Total number of appliances in hhld (harmonised)	

Where imputed values represented weekly or fortnightly estimates (i.e. those derived from survey diary data) these were multiplied up to give an annual estimate[22]. In addition, imputed estimates of emissions from private car use were set to zero where a case was recorded as not having a car. Similarly, imputed values for international aviation emissions were set to zero where a case was flagged as having taken no international flights.

2. Grossing imputed values:

The transformed and adjusted imputed values were then grossed to give an annual sum total estimate of household carbon emissions to match the totals from the donor surveys. Sum totals were derived from the NTS vehicle (for private car CO_2), journey (for public transport CO_2) and long-distance journey (for domestic aviation) datasets and the APS for international aviation, for GB and England only (for EFS and EHCS imputations respectively).

For the EFS to EHCS imputation (of household fuel CO_2) the target sum total for grossing was derived from the EFS England subset.

The final resulting annual estimates of carbon emissions in the GB EFS and England-only EHCS datasets are shown below.

[22] NTS Public Transport emissions based on one week diary (therefore *52.14); NTS Long-distance journey (domestic aviation) based on two week period (therefore *26.07).

Table 20. Resulting estimates of carbon emissions in the (GB) EFS imputed dataset ($kgCO_2$)

Emissions source	Sum ('000s)	Mean
Household fuels	137,383,481	5,675
Private car - Leisure	41,017,737	1,694
Private car - Commute	22,982,540	949
Total Private Car CO2	64,000,276	2,644
Public transport - Leisure	5,087,185	210
Public transport - Commute	2,215,574	92
Public transport - Total	7,302,758	302
Domestic aviation	787,347	33
International aviation	28,618,966	1,182
Total transport emissions	100,709,348	4,160
Total carbon emissions^	238,092,830	9,836

^ Sum total of emissions from all direct sources as modelled in this study

Table 21. Resulting estimates of carbon emissions in the (England only) EHCS imputed dataset ($kgCO_2$)

Emissions source	Sum ('000s)	Mean
Household fuels	110,964,817	5,190
Private car - Leisure	36,587,548	1,711
Private car - Commute	20,225,539	946
Total Private Car CO2	56,813,087	2,657
Public transport - Leisure	4,511,724	211
Public transport - Commute	2,052,853	96
Public transport - Total	6,564,577	307
Domestic aviation	360,131	17
International aviation	24,036,848	1,124
Total transport emissions	87,774,643	4,105
Total carbon emissions^	198,739,460	9,296

Appendix

Annex Table 1. Summary of the harmonised socio-demographic variables

Variable	Harmonised name	Harmonisation Status
Household composition	**htype**	Different definition of 'child'. Suggest using A063 as the definition of child is similar in NTS and EFS, but need to address definition of benefit unit. Computed new household type in NTS, EFS, EHCS based on combination of age, sex, number of couples aggregated from personal level datasets.
Number of adults	**nadult**	Different definitions, used combination of EFS age (I5) / gender of person (I3) to compute new harmonised measure. For EHCS used combination of AGE and SEX variables to compute new harmonised measures. For APS calculated by subtracting Child from Household to give number of adults.
Number of children	**nchild**	Different definitions, used combination of EFS age (I5) / gender of person (I3) to compute new harmonised measure. For EHCS used combination of AGE and SEX variables to compute new harmonised measures. APS included number of children in household (Child).
Household size	**hhsize**	New variable HHSIZE calculated, which is based on summing up of new ADULT and CHILD variable counts. Cross checked against H58 and sums equal. In EHCS cross-checked with HHSIZEX. APS included variable on household size (Household).
Number of pensioners	**npens**	Calculated using individual age (I5) and gender (I3), then aggregated to the household level.
Economic status of HRP	**emphrp**	There were some harmonisation issues around definitions of economically active/inactive, so had to compute new harmonised variable in NTS using person level variable I177 and in EFS using combination of A206 and A015 (EFS) then assigned to HRP. EHCS had derived variables already computed on economic position and self-employment, so computed new variable to match NTS, EFS.
NS-SEC of HRP	**nssec** **nssec8** **nssecnew**	Different operationalisation in both NTS and EFS, although same categories used in both. NTS: used I398 to create new NS-SEC measure for all members of household and then select and compute for HRP (I2=0). Then recoded NTS into new 8 SEC variable to match EFS but found distributions very different. EFS: Individuals that have retired within the last 12 months are classified according to their employment. Other retired individuals are assigned to the 'Not classifiable for other reasons' category. Created new harmonised variable (NSSECNEW) which puts retired into 'Not classifiable: HRP Retired'. ****EHCS missing categories 'Never worked, 'LT unemployed' and 'Students'. Can create 'student' based on employment status but not never worked/long-term unemployed** APS- does not include NS-SEC, but does have SEG (see below)
SEG of HRP	**seghrp**	Different classification schemes used in the surveys. However, able to compute new harmonised **seghrp** based on recoding **nssec** categories in NTS, EFS, EHCS to match SEG variable in APS .

Variable	Harmonised name	Harmonisation Status
Urban and rural areas	**urbrur**	Different classification schemes used in the surveys. Able to compute basic urban/rural category URBRUR across surveys. In addition, EFS and EHCS variables are similar for England and Wales, so created additional harmonised variable URBRUREW for comparison. EFS classifications different for Scotland and missing for Northern Ireland. **For APS can possibly use combination of county and district to compute urban/rural divide.**
	urbrurew	
Government Office Region (GOR)	**gorx**	NTS missing NI in NTS, and some minor recoding issues. Merged PSU and HHLD NTS files and computed new variable to match EFS GORX. NTS H170 have 1/3 missing responses so have to correct. EHCS- minor recodes and missing Wales, Scotland, Ni, but otherwise fine.
Household income (gross annual)	**hhginc_en** **hhginc_ea**	Banded annual for NTS. Copied H70 into new HHGINC_EN variable and then recoded EFS p344 to compute first income dyad (EFS-NTS). Banded annual for APS. Copied INCOME into new HHGINC_EA and then recoded EFS p344 to compute second income dyad (EFS-APS). Computed HHGINC_EN and HHGINC_EA and set to missing (-99) in EHCS. ***EHCS refers to total annual NET househld income. Total (gross) income in income.sav file refers to HRP and partner only and is in £520 bands. New variable HHGINC computed and set to missing (-99) in EHCS.**
Household income (net disposable annual)	**hhninc**	*EHCS refers to total annual NET househld income. Total income in income.sav file refers to HRP and partner only and is in £520 bands. hhincx top-coded.
Individual income (gross)		Banded annual for NTS
		Weekly top-coded anonymised for EFS.
		Can created harmonised bands as done for household income.
Age of HRP	**agehrp**	Computed in NTS by choosing age HRP and then recoding at household level to match EFS age bands. Computed in APS by using age of respondent.
Gender of person		Same definitions for NTS, EFS, EHCS. APS needed to be recoded.
Gender of HRP	**sexhrp**	Computed in NTS by choosing gender of HRP. Computed in APS by using sex of respondent.
Ethnic group of HRP	**ethhrp**	Computed in NTS by using ethnic group of HRP. Definitional problems as NTS refers to ethnic group and EFS to ethnic origin. EFS has roughly 20% not classifiable and many differences between the ethnic origin of HRP and members of the household. Modal value in person file= 0. *EHCS categories of ethnic origin of HRP match up with NTS only.
Number of cars/light vans	**ncars**	Recoded H55 into new top-coded (5+) NTS variable (6=NA, recoded to 0). A124 in EFS recoded (5 tru hi=5+) to harmonise into top-coded with NTS. ***EHCS- not available, dummy variable created**

Variable	Harmonised name	Harmonisation Status
Petrol and Diesel consumed (cost)	**npetrolcost ndieselcost nvehfuelcost**	NTS diary is for one week and is banded, EFS is for two weeks and unbanded, but can possibly recode/compute. Computed new harmonised petrol, diesel and combined fuel costs. Mutiplied EFS by 0.5 to match NTS. *__EHCS- not available, dummy variables created__
Accommodation type	**acctyp**	Minor recoding issues around particular/unknown types and others, so new harmonised variables created by retaining main types. EFS dwellcat had to be recategorised, but otherwise fine. ***EHCS Bungalow, detached combined to form bungalow-detached - otherwise categories look fine.**
Tenure	**tenure**	Harmonised into smaller categories using NTS EFS. NTS- Not possible to split renters (private/social housing), but can do for EFS, EHCS. EHCS- missing Part rent/part mortgage and live rent free/squatting & NA, so only 3 categories (Own outright, own with mortgage, rent).
Number of workers in household	**nworker**	For NTS, calculated using aggregating workers as identified in I177 (=0 or =1). Computed for EFS using same procedure. EHCS- computed counting up the number of people in the household who were employed or on a government training scheme. Then aggregated to the household level and merged.
Survey year	**NTSyear EFSyear EHCSyear APSyear**	Computed new survey years. Created new survey year, month, quarter for NTS and EFS. Then created EFS filter to remove common data in 2005-06/2006 (n=1715 cases removed)
Sample calendar year	**nyear**	Created new harmonised variable based on combination of sample year and month
Sample month	**nmonth**	Created new harmonised variable based on existing sample month
Sample quarter	**nqrtr**	Created new harmonised variable based on recoding newly created sample month to harmonise across EFS surveys
Sample weighting	**nweightA nweightQ**	Computed new annual and quarterly weights by copying in original weights from 2005, 2005, and 2007 and new weights from 2006. ***NB: EHCS- computed nweighta equals household weight (aagch67). No dummy variables created for NTS & EFS.** APS- dummy variables created to match other datasets
Unique household identifier	**newhhid**	Computed new household identifier (NEWHHID) which assigns a unique survey and household number combination based on calendar year and survey number. A 10 digit household ID number which sequentially adds cases by survey year.
Country	**ngorx3**	Recoded harmonised GORX variable into three countries (NI set as missing in EFS as not available in NTS). EHCS- missing Wales, Scotland, NI. APS- Recoded into same GOR. Also included 13=South Ireland
Method of payment – Gas	**gasMOP**	Created new harmonised variable based on recoding EHCS and EFS variables into fewer categories. Dummy variables created in NTS and APS.

Variable	Harmonised name	Harmonisation Status
Method of payment – Electricity	**elecMOP**	Created new harmonised variable based on recoding EHCS and EFS variables into fewer categories. Dummy variables created in NTS and APS.
Heating fuel	**Heatfuel**	Created new variable in EFS using central heating type and expenditure on fuels where no central heating exists. Recoded EHCS heating fuel to match EFS categories. Dummy variables created in NTS and APS.
Number of bedrooms	**DVbedp**	EHCS variable recoded to match categories in EFS. Dummy variable created in NTS and APS.
Gas_supply	**Gas_supply**	Harmonised EFS – EHCS to flag households that have a mains gas supply. Dummy variables created in NTS and APS.
Tumble drier	**ndrier**	Harmonised EFS – EHCS to flag households that have a tumble drier. Dummy variable created for NTS and APS.
Washing machine	**nwashmach**	Harmonised EFS – EHCS to flag households that have a a washing machine. Dummy variable created for NTS and APS.

Annex Table 2. Crosswalk of household level socio-demographic and energy-consumption variables in the NTS, EFS, EHCS and APS

Variable	Variable name	NTS	EFS	EHCS	APS
newhhid	Household ID (harmonised)	X	X	X	X
NTSyear	NTS survey year	X			
EFSyear	EFS survey year		X		
EHCSyear	EHCS survey year			X	
APSyear	APS survey year				X
NTS	NTS survey (dummy coded)	X	X	X	X
EFS	EFS survey (dummy coded)	X	X	X	X
EHCS	EHCS survey (dummy coded)	X	X	X	X
APS	APS survey (dummy coded)	X	X	X	X
nyear	Survey calendar year (harmonised)	X	X	X	
nmonth	Survey month (harmonised)	X	X	X	
nqrtr	Survey quarter (harmonised)	X	X	X	
Gorx	GOR (harmonised)	X	X	X	
ngorx3	GOR country (harmonised)	X	X	X	
urbrurew	Urban/rural England and Wales (harmonised)		X	X	
urbrur	Urban/rural classification (harmonised)	X	X	X	
nadult	Number of adults (harmonised)	X	X	X	
nchild	Number of children (harmonised)	X	X	X	
npens	Number of pensioners (harmonised)	X	X	X	
ncpl	Number of couples (harmonised)	X	X	X	
nworker	Number of workers (harmonised)	X	X	X	
hhsize	Number of persons (harmonised)	X	X	X	
htype	Household type (harmonised)	X	X	X	
acctyp	Accommodation type (harmonised)	X	X	X	
tenure	Tenure (harmonised)	X	X	X	
ncars	Number of cars/vans (harmonised)	X	X		
hhgincEN	EFS-NTS Gross annual income (harmonised)	X	X		
hhgincEA	EFS-APS Gross annual income (harmonised)		X		X
hhninc	Disposable annual income (harmonised)		X	X	
agehrp	Age of HRP (harmonised)	X	X		X
SexHRP	Gender of HRP (harmonised)	X	X	X	X
ethhrp	Ethnic group of HRP (harmonised)	X	X	X	
emphrp	Employment status of HRP (harmonised)	X	X	X	
nssec	NS-SEC of HRP long (harmonised)	X	X		
nssec8	NS-SEC of HRP short (harmonised)	X	X	X	
nssecnew	NS-SEC of HRP retired re-classfied (harmonised)	X	X		
seghrp	SEG HRP computed (harmonised)	X	X	X	X
W1	W1 - Unweighted diary sample	X			
W2	W2 - Diary sample weight	X			
W3	W3 - Interview sample weight	X			
weightq	Quarterly Weight		X		
weighta	Annual weight		X		
nweightQ	New quarterly weight (dummy coded)		X		
nweightA	New annual weight (dummy coded)		X		
aagcd67	Dwelling weight (core cases 2006-7 & 2007-8)			X	
aagch67	Household weight (core cases 2006-7 & 2007-8)			X	

Variable	Variable name	NTS	EFS	EHCS	APS
population	Journey to household weight (APS)				X
annualWeight	Journey annual weight (APS)				X
npetrolltrs	petrol litres - hhld		X		
ndieselltrs	diesel litres - hhld		X		
nvehfuelltrs	vehicle fuel litres – hhld		X		
npetrolcost	petrol cost – hhld		X		
ndieselcost	diesel cost – hhld		X		
nvehfuelcost	vehicle fuel cost – hhld		X		
nV163	CO2 emissions – hhld	X			
nV139C2	Private car - Commute CO2 emissions	X			
nV143C2	Private car - Leisure CO2 emissions	X			
nPT_CO2_leisure	Public transport - Leisure CO2 emissions	X			
nPT_CO2_commute	Public transport - Commute CO2 emissions	X			
nsdjwalkdist	SDJ- walk distance - hhld (miles)	X			
nsdjbikedist	SDJ- bike distance - hhld (miles)	X			
nsdjcarddist	SDJ- car driver distance - hhld (miles)	X			
nsdjcarpdist	SDJ- car passenger distance - hhld (miles)	X			
nsdjmotcdist	SDJ- motor cycle distance - hhld (miles)	X			
nsdjoprvdist	SDJ- other private transport distance - hhld (miles)	X			
nsdjlbusdist	SDJ- London stage bus distance - hhld (miles)	X			
nsdjobusdist	SDJ- other stage bus distance - hhld (miles)	X			
nsdjnlcbdist	SDJ- non local bus distance - hhld (miles)	X			
nsdjltundist	SDJ- LT underground distance - hhld (miles)	X			
nsdjraildist	SDJ- surface rail distance - hhld (miles)	X			
nsdjtaxidist	SDJ- taxi minicab distance - hhld (miles)	X			
nsdjopubdist	SDJ- other public transport distance - hhld (miles)	X			
nsdjwalkcost	SDJ- walk cost - hhld (pence)	X			
nsdjbikecost	SDJ- bike cost - hhld (pence)	X			
nsdjcardcost	SDJ- car driver cost - hhld (pence)	X			
nsdjcarpcost	SDJ- car passenger cost - hhld (pence)	X			
nsdjmotccost	SDJ- motor cycle cost - hhld (pence)	X			
nsdjoprvcost	SDJ- other private transport cost - hhld (pence)	X			
nsdjlbuscost	SDJ- London stage bus cost - hhld (pence)	X			
nsdjobuscost	SDJ- other stage bus cost - hhld (pence)	X			
nsdjnlcbcost	SDJ- non local bus cost - hhld (pence)	X			
nsdjltuncost	SDJ- LT underground cost - hhld (pence)	X			
nsdjrailcost	SDJ- surface rail cost - hhld (pence)	X			
nsdjtaxicost	SDJ- taxi minicab cost - hhld (pence)	X			
nsdjopubcost	SDJ- other public transport cost - hhld (pence)	X			
nJDUNGROSS	SDJ- journey distance – hhld (miles)	X			
nJD	SDJ- journey distance (grossed up for short walks) – hhld (miles)	X			
nJJXSC	SDJ- number of journeys (grossed up for short walks) – hhld	X			
nJOTXSC	SDJ- overall journey time (minutes - grossed up for short walks) – hhld	X			
nJTTXSC	SDJ- overall travelling time (minutes -	X			

Variable	Variable name	NTS	EFS	EHCS	APS
	grossed up for short walks) – hhld				
nJTOTCOST	SDJ- overall journey cost – hhld (pence)	X			
nldjwalkdist	LDJ- walk distance - hhld (miles)	X			
nldjbikedist	LDJ- bike distance - hhld (miles)	X			
nldjcarddist	LDJ- car driver distance - hhld (miles)	X			
nldjcarpdist	LDJ- car passenger distance - hhld (miles)	X			
nldjmotcdist	LDJ- motor cycle distance - hhld (miles)	X			
nldjoprvdist	LDJ- other private transport distance - hhld (miles)	X			
nldjlbusdist	LDJ- London stage bus distance - hhld (miles)	X			
nldjobusdist	LDJ- other stage bus distance - hhld (miles)	X			
nldjnlcbdist	LDJ- non local bus distance - hhld (miles)	X			
nldjltundist	LDJ- LT underground distance - hhld (miles)	X			
nldjraildist	LDJ- surface rail distance - hhld (miles)	X			
nldjtaxidist	LDJ- taxi minicab distance - hhld (miles)	X			
nldjopubdist	LDJ- other public transport distance - hhld (miles)	X			
nldjairtdist	LDJ – air travel distance - hhld (miles)	X			
nL9	LDJ – journey distance - hhld (miles)	X			
nldjAirCO2_private	Air travel CO2	X			
A128	Gas - method of payment		X		
A130	Electricity - method of payment		X		
B017	Oil for central heating - last quarter		X		
B018	Bottled gas for central heating		X		
B170	Gas amount paid in last account		X		
B175	Electricity amount paid in last account		X		
B221	Gas - Amt of payment (budget scheme)		X		
B222	Elec - Amt of payment (budget scheme)		X		
P249t	Gas - slot meter payments less rebates		X		
P250t	Elec - slot meter payments less rebates		X		
C45112t	Second dwelling: electricity account payment		X		
C45212t	Second dwelling: gas account payment		X		
C45222t	Bottled gas - other		X		
C45312t	Paraffin		X		
C45411t	Coal and coke		X		
C45412t	Wood and peat		X		
C45511t	Hot water, steam and ice		X		
C45114t	Electricity slot meter payment		X		
C45214t	Gas slot meter payment		X		
C72211t	Petrol		X		
C72212t	Diesel oil		X		
C72213t	Other motor oils		X		
C73112t	Railway and tube fares other than season		X		
C73212t	Bus and coach fares other than season tickets		X		
C73213t	Taxis and hired cars with drivers		X		
C73214t	Other personal travel		X		
C73311t	Air fares (within UK)		X		
C73312t	Air fares (international)		X		
C73411t	Water travel		X		

Variable	Variable name	NTS	EFS	EHCS	APS
C73512t	Combined fares other than season tickets		X		
C73513t	School travel		X		
C71311t	Purchase of bicycle		X		
ECredit_pkWh			X		
EDD_pkWh			X		
Eppm_pkWh			X		
GCredit_pkWh			X		
GDD_pkWh			X		
Gppm_pkWh			X		
Oil_pkWh			X		
Coal_pkWh			X		
LPG_pkWh			X		
Bottled_pkWh			X		
Wood_pkWh			X		
Communal_pkWh			X		
Petrol_pndl			X		
Diesel_pndl			X		
kWh_Elect_Account_estimate			X		
kWh_Elect_Slot_estimate			X		
kWh_Elect_2ndDwell_estimate			X		
kWh_Elect_BrdBudget_estimate			X		
kWh_Gas_Account_estimate			X		
kWh_Gas_Slot_estimate			X		
kWh_Gas_2ndDwell_estimate			X		
kWh_Gas_BrdBudget_estimate			X		
kWh_GasBottled_estimate			X		
kWh_LPG_estimate			X		
kWh_Coal_estimate			X		
kWh_OilCentralHtg_estimate			X		
kWh_Paraffin_estimate			X		
kWh_Wood_estimate			X		
kWh_Steam_estimate			X		
kWh_Elec_Total_Estimate			X		
kWh_Gas_Total_Estimate			X		
kWh_Petroleum_Total_Estimate			X		
kWh_AllOther_estimate			X		
l_Petrol_Estimate			X		
l_Diesel_Estimate			X		
CO2_Elec			X		
CO2_Gas			X		
CO2_GasBottled			X		
CO2_LPG			X		
CO2_Coal			X		
CO2_Oil			X		
CO2_Paraffin			X		
CO2_Wood			X		
CO2_Petrol_efs			X		
CO2_Diesel_efs			X		
CO2_household_fuels			X		
CO2_road_fuels			X		
CO2_EFS_total			X		

Variable	Variable name	NTS	EFS	EHCS	APS
CO2_Heat_Load			X		
CO2_Power_Load			X		
dwage9x	Dwelling age (DIMPSA)			X	
wallinsx	DIMPSA- Type of wall insulation			X	
wall	Wall type and insulation - Calculated by us (DIMPSA)			X	
Loftins4	Loft insulation thickness (DIMPSA)			X	
nbedsx	Number of bedrooms (DIMPSA)			X	
floorx	Floor area (DIMPSA)			X	
dblglaz4	Double glazing (DIMPSA)			X	
lowenlt	Rooms with low energy lighting - Calculated by us (DIMPSA)			X	
nolowlt	Rooms without low energy lighting - Calculated by us (DIMPSA)			X	
heat7x	Main heating system (DIMPSA)			X	
Finmhfue	Main heating fuel (DIMPSA)			X	
Finmhboi	Boiler group (DIMPSA)			X	
sysage	Age of heating system (DIMPSA)			X	
sap05	Energy efficiency (SAP05) rating (DIMPSA)			X	
draughty	Draughts and need for DP - Calculated by us (DIMPSA)			X	
typercov	Predominant type of roof covering (DIMPSA)			X	
typerstr	Predominant type of roof stucture (DIMPSA)			X	
Fdffloor	Number of floors in flat (DIMPSA)			X	
Finopele	Electrical system - off peak supply (DIMPSA)			X	
fardwell	Number of dwellings in area (DIMPSA)			X	
sph_gjs	Space heating gjs (DIMPSA)			X	
wht_gjs	Water heating gjs (DIMPSA)			X	
lae_gjs	Lights and appliances gjs (DIMPSA)			X	
cke_gjs	Cooking gjs (DIMPSA)			X	
Gas_supply	Gas supply		X	X	
Gas_pay	Gas payment method		X		
Elec_pay	Electric payment method		X		
gasMOP	Method of payment - gas		X	X	
elecMOP	Method of payment - electricity		X	X	
Central_heating	Central heating type		X		
CSE_heatfuel	Central heating type (CSE)		X		
Heatfuel	Heating fuel (harmonised)		X	X	
a114	Rooms in accomodation - total		X		
p200	Number of rooms occupied (de basis)		X		
DVbedp	Number of bedrooms occupied (harmonised)		X	X	
ncounty	Home county (APS only)				X
ndistrict	Home district (APS only)				X
rtn_distance_km_total	Air (non-business, non-UK) distance return (APS)				X
kgCO2_total	Air (non-business, non-UK) CO2 emissions (APS)				X
nvideo	Video recorder (EFS)		X		
nfreezer	Freezer (EFS)		X		

Variable	Variable name	NTS	EFS	EHCS	APS
ncomputer	Personal computer (EFS)		X		
ndrier	Tumble drier (EFS/EHCS)		X	X	
nmicrowve	Microwave oven (EFS)		X		
ndishwash	Dishwasher		X		
ncdplay	CD player (EFS)		X		
ndvd	DVD player (EFS)		X		
nwashmach	Washing machine (EFS/EHCS)		X	X	
ntelephon	Telephone/mobile (EFS)		X		
nTVNum	Number of TVs in hhld (EFS)		X		
appliances	Total number of appliances in hhld (Harmonised)		X		

Annex Table 3. Valid and missing data available in harmonised dataset

Variable and variable name	Valid	Missing
newhhid Household ID (harmonised)	107,933	0
NTSyear NTS survey year	45,750	62,183
EFSyear EFS survey year	24,649	83,284
EHCSyear EHCS survey year	15,604	92,329
APSyear APS survey year	21,930	86,003
NTS NTS survey (dummy coded)	107,933	0
EFS EFS survey (dummy coded)	107,933	0
EHCS EHCS survey (dummy coded)	107,933	0
APS APS survey (dummy coded)	107,933	0
nyear Survey calendar year (harmonised)	107,933	0
nmonth Survey month (harmonised)	107,933	0
nqrtr Survey quarter (harmonised)	107,933	0
gorx GOR (harmonised)	107,262	671
ngorx3 GOR country (harmonised)	107,235	698
urbrurew Urban/rural England and Wales (harmonised)	36,094	71,839
urbrur Urban/rural classification (harmonised)	83,939	23,994
nadult Number of adults (harmonised)	99,499	8,434
nchild Number of children (harmonised)	99,611	8,322
npens Number of pensioners (harmonised)	86,003	21,930
ncpl Number of couples (harmonised)	86,003	21,930
nworker Number of workers (harmonised)	86,003	21,930
hhsize Number of persons (harmonised)	100,652	7,281
htype Household type (harmonised)	86,003	21,930
acctyp Accommodation type (harmonised)	86,003	21,930
tenure Tenure (harmonised)	86,003	21,930
ncars Number of cars/vans (harmonised)	70,399	37,534
hhgincEN EFS-NTS Gross annual income (harmonised)	70,399	37,534
hhgincEA EFS-APS Gross annual income (harmonised)	35,593	72,340
hhninc Disposable annual income (harmonised)	40,253	67,680
agehrp Age of HRP (harmonised)	101,687	6,246
sexhrp Gender of HRP (harmonised)	107,763	170
ethhrp Ethnic group of HRP (harmonised)	85,524	22,409
emphrp Employment status of HRP (harmonised)	86,003	21,930
nssec NS-SEC of HRP long (harmonised)	70,399	37,534
nssec8 NS-SEC of HRP short (harmonised)	86,003	21,930
nssecnew NS-SEC of HRP retired re-classfied (harmonised)	70,399	37,534
seghrp SEG HRP computed (harmonised)	87,364	20,569
W1 W1 - Unweighted diary sample	45,750	62,183
W2 W2 - Diary sample weight	45,750	62,183
W3 W3 - Interview sample weight	45,750	62,183
weightQ Quarterly Weight	24,649	83,284
weightA Annual weight	24,649	83,284
nweightQ New quarterly weight (dummy coded)	17,851	90,082
nweightA New annual weight (dummy coded)	33,455	74,478
aagcd67 Dwelling weight (core cases 2006-7 & 2007-8)	15,604	92,329

Variable and variable name	Valid	Missing
aagch67 Household weight (core cases 2006-7 & 2007-8)	15,604	92,329
population CAA APS weight	21,930	86,003
annualWeight CAA APS annual weight	21,930	86,003
npetrolltrs petrol litres - hhld	27,843	80,090
ndieselltrs diesel litres - hhld	8,570	99,363
nvehfuelltrs vehicle fuel litres – hhld	31,602	76,331
npetrolcost petrol cost – hhld	43,139	64,794
ndieselcost diesel cost – hhld	12,585	95,348
nvehfuelcost vehicle fuel cost – hhld	49,115	58,818
nV163 CO2 emissions – hhld	11,103	96,830
nV139C2 Private car - Commute CO2 emissions	32,440	75,493
nV143C2 Private car - Leisure CO2 emissions	32,289	75,644
nPT_CO2_leisure Public transport - Leisure CO2 emissions	17,894	90,039
nPT_CO2_commute Public transport - Commute CO2 emissions	5,628	102,305
nsdjwalkdist SDJ- walk distance - hhld (miles)	22,619	85,314
nsdjbikedist SDJ- bike distance - hhld (miles)	3,614	104,319
nsdjcarddist SDJ- car driver distance - hhld (miles)	29,343	78,590
nsdjcarpdist SDJ- car passenger distance - hhld (miles)	29,360	78,573
nsdjmotcdist SDJ- motor cycle distance - hhld (miles)	703	107,230
nsdjoprvdist SDJ- other private transport distance - hhld (miles)	2,855	105,078
nsdjlbusdist SDJ- London stage bus distance - hhld (miles)	2,504	105,429
nsdjobusdist SDJ- other stage bus distance - hhld (miles)	11,654	96,279
nsdjnlcbdist SDJ- non local bus distance - hhld (miles)	502	107,431
nsdjltundist SDJ- LT underground distance - hhld (miles)	1,579	106,354
nsdjraildist SDJ- surface rail distance - hhld (miles)	4,811	103,122
nsdjtaxidist SDJ- taxi minicab distance - hhld (miles)	6,146	101,787
nsdjopubdist SDJ- other public transport distance - hhld (miles)	933	107,000
nsdjwalkcost SDJ- walk cost - hhld (pence)	22,619	85,314
nsdjbikecost SDJ- bike cost - hhld (pence)	3,614	104,319
nsdjcardcost SDJ- car driver cost - hhld (pence)	29,343	78,590
nsdjcarpcost SDJ- car passenger cost - hhld (pence)	29,360	78,573
nsdjmotccost SDJ- motor cycle cost - hhld (pence)	703	107,230
nsdjoprvcost SDJ- other private transport cost - hhld (pence)	2,855	105,078
nsdjlbuscost SDJ- London stage bus cost - hhld (pence)	2,504	105,429
nsdjobuscost SDJ- other stage bus cost - hhld (pence)	11,654	96,279
nsdjnlcbcost SDJ- non local bus cost - hhld (pence)	502	107,431
nsdjltuncost SDJ- LT underground cost - hhld (pence)	1,579	106,354
nsdjrailcost SDJ- surface rail cost - hhld (pence)	4,811	103,122
nsdjtaxicost SDJ- taxi minicab cost - hhld (pence)	6,146	101,787
nsdjopubcost SDJ- other public transport cost - hhld (pence)	933	107,000
nJDUNGROSS SDJ- journey distance – hhld (miles)	39,560	68,373
nJD SDJ- journey distance (grossed up for short walks) – hhld (miles)	39,560	68,373
nJJXSC SDJ- number of journeys (grossed up for short walks) – hhld	39,560	68,373
nJOTXSC SDJ- overall journey time (minutes - grossed up for short walks) – hhld	39,560	68,373
nJTTXSC SDJ- overall travelling time (minutes - grossed up for short walks) – hhld	39,560	68,373
nJTOTCOST SDJ- overall journey cost – hhld (pence)	39,560	68,373

Variable and variable name	Valid	Missing
nldjwalkdist LDJ- walk distance - hhld (miles)	1	107,932
nldjbikedist LDJ- bike distance - hhld (miles)	11	107,922
nldjcarddist LDJ- car driver distance - hhld (miles)	12,242	95,691
nldjcarpdist LDJ- car passenger distance - hhld (miles)	9,822	98,111
nldjmotcdist LDJ- motor cycle distance - hhld (miles)	148	107,785
nldjoprvdist LDJ- other private transport distance - hhld (miles)	796	107,137
nldjlbusdist LDJ- London stage bus distance - hhld (miles)	1	107,932
nldjobusdist LDJ- other stage bus distance - hhld (miles)	22	107,911
nldjnlcbdist LDJ- non local bus distance - hhld (miles)	1,206	106,727
nldjltundist LDJ- LT underground distance - hhld (miles)	2	107,931
nldjraildist LDJ- surface rail distance - hhld (miles)	2,761	105,172
nldjtaxidist LDJ- taxi minicab distance - hhld (miles)	140	107,793
nldjopubdist LDJ- other public transport distance - hhld (miles)	101	107,832
nldjairtdist LDJ – air travel distance - hhld (miles)	101	107,832
nL9 LDJ – journey distance - hhld (miles)	17,136	90,797
nldjAirCO2_private Air travel CO2	254	107,679
A128 Gas - method of payment	24,649	83,284
A130 Electricity - method of payment	24,649	83,284
B017 Oil for central heating - last quarter	24,649	83,284
B018 Bottled gas for central heating	24,649	83,284
B175 Electricity amount paid in last account	24,649	83,284
B170 Gas amount paid in last account	24,649	83,284
B221 Gas - Amt of payment (budget scheme)	24,649	83,284
B222 Elec - Amt of payment (budget scheme)	24,649	83,284
P249t Gas - slot meter payments less rebates	24,649	83,284
P250t Elec - slot meter payments less rebates	24,649	83,284
C45112t Second dwelling: electricity account payment	24,649	83,284
C45212t Second dwelling: gas account payment	24,649	83,284
C45222t Bottled gas - other	24,649	83,284
C45312t Paraffin	24,649	83,284
C45411t Coal and coke	24,649	83,284
C45412t Wood and peat	24,649	83,284
C45511t Hot water, steam and ice	24,649	83,284
C45114t Electricity slot meter payment	24,649	83,284
C45214t Gas slot meter payment	24,649	83,284
C72211t Petrol	24,649	83,284
C72212t Diesel oil	24,649	83,284
C72213t Other motor oils	24,649	83,284
C73112t Railway and tube fares other than season	24,649	83,284
C73212t Bus and coach fares other than season tickets	24,649	83,284
C73213t Taxis and hired cars with drivers	24,649	83,284
C73214t Other personal travel	24,649	83,284
C73311t Air fares (within UK)	24,649	83,284
C73312t Air fares (international)	24,649	83,284
C73411t Water travel	24,649	83,284
C73512t Combined fares other than season tickets	24,649	83,284

Variable and variable name	Valid	Missing
C73513t School travel	24,649	83,284
C71311t Purchase of bicycle	24,649	83,284
ECredit_pkWh	24,649	83,284
EDD_pkWh	24,649	83,284
Eppm_pkWh	24,649	83,284
GCredit_pkWh	24,649	83,284
GDD_pkWh	24,649	83,284
Gppm_pkWh	24,649	83,284
Oil_pkWh	24,649	83,284
Coal_pkWh	24,649	83,284
LPG_pkWh	24,649	83,284
Bottled_pkWh	24,649	83,284
Wood_pkWh	24,649	83,284
Communal_pkWh	24,649	83,284
Petrol_pndl	24,649	83,284
Diesel_pndl	24,649	83,284
kWh_Elect_Account_estimate	24,649	83,284
kWh_Elect_Slot_estimate	24,649	83,284
kWh_Elect_2ndDwell_estimate	24,649	83,284
kWh_Elect_BrdBudget_estimate	24,649	83,284
kWh_Gas_Account_estimate	24,649	83,284
kWh_Gas_Slot_estimate	24,649	83,284
kWh_Gas_2ndDwell_estimate	24,649	83,284
kWh_Gas_BrdBudget_estimate	24,649	83,284
kWh_GasBottled_estimate	24,649	83,284
kWh_LPG_estimate	24,649	83,284
kWh_Coal_estimate	24,649	83,284
kWh_OilCentralHtg_estimate	24,649	83,284
kWh_Paraffin_estimate	24,649	83,284
kWh_Wood_estimate	24,649	83,284
kWh_Steam_estimate	24,649	83,284
kWh_Elec_Total_Estimate	24,649	83,284
kWh_Gas_Total_Estimate	24,649	83,284
kWh_AllOther_estimate	24,649	83,284
kWh_Petroleum_Total_Estimate	24,649	83,284
l_Petrol_Estimate	24,649	83,284
l_Diesel_Estimate	24,649	83,284
CO2_Elec kg CO2 Elec (EFS)	24,649	83,284
CO2_Gas kg CO2 Gas (EFS)	24,649	83,284
CO2_GasBottled kg CO2 GasBottled (EFS)	24,649	83,284
CO2_LPG kg CO2 LPG (EFS)	24,649	83,284
CO2_Coal kg CO2 Coal (EFS)	24,649	83,284
CO2_Oil kg CO2 Oil (EFS)	24,649	83,284
CO2_Paraffin kg CO2 Paraffin (EFS)	24,649	83,284
CO2_Wood kg CO2 Wood (EFS)	24,649	83,284
CO2_Petrol_efs kg CO2 Petrol (EFS)	24,649	83,284

Variable and variable name	Valid	Missing
CO2_Diesel_efs kg CO2 Diesel (EFS)	24,649	83,284
CO2_household_fuels kg CO2 from household fuel consumption (EFS)	24,649	83,284
CO2_road_fuels kg CO2 from road transport fuel consumption (EFS)	24,649	83,284
CO2_EFS_total kg CO2 from household and road transport fuel consumption (EFS)	24,649	83,284
CO2_Heat_Load kg CO2 heat load (EFS)	24,649	83,284
CO2_Power_Load kg CO2 power load (EFS)	24,649	83,284
dwage9x Dwelling age (DIMPSA)	15,604	92,329
wallinsx DIMPSA- Type of wall insulation	15,604	92,329
wall Wall type and insulation - Calculated by us (DIMPSA)	15,604	92,329
Loftins4 Loft insulation thickness (DIMPSA)	13,518	94,415
nbedsx Number of bedrooms (DIMPSA)	22,402	85,531
floorx Floor area (DIMPSA)	15,604	92,329
dblglaz4 Double glazing (DIMPSA)	15,604	92,329
lowenlt Rooms with low energy lighting - Calculated by us (DIMPSA)	15,604	92,329
nolowlt Rooms without low energy lighting - Calculated by us (DIMPSA)	15,604	92,329
heat7x Main heating system (DIMPSA)	15,604	92,329
Finmhfue Main heating fuel (DIMPSA)	15,417	92,516
Finmhboi Boiler group (DIMPSA)	15,426	92,507
sysage Age of heating system (DIMPSA)	15,604	92,329
sap05 Energy efficiency (SAP05) rating (DIMPSA)	15,604	92,329
draughty Draughts and need for DP - Calculated by us (DIMPSA)	15,604	92,329
typercov Predominant type of roof covering (DIMPSA)	15,604	92,329
typerstr Predominant type of roof stucture (DIMPSA)	15,604	92,329
Fdffloor Number of floors in flat (DIMPSA)	3,351	104,582
Finopele Electrical system - off peak supply (DIMPSA)	15,368	92,565
fardwell Number of dwellings in area (DIMPSA)	15,590	92,343
sph_gjs Space heating gjs (DIMPSA)	15,604	92,329
wht_gjs Water heating gjs (DIMPSA)	15,604	92,329
lae_gjs Lights and appliances gjs (DIMPSA)	15,604	92,329
cke_gjs Cooking gjs (DIMPSA)	15,604	92,329
Gas_supply Gas supply	38,279	69,654
Gas_pay Gas payment method	24,649	83,284
Elec_pay Electric payment method	24,649	83,284
gasMOP Method of payment - gas	40,253	67,680
elecMOP Method of payment - electricity	40,248	67,685
Central_heating Central heating type	24,649	83,284
CSE_heatfuel Central heating type (CSE)	24,649	83,284
Heatfuel Heating fuel (harmonised)	40,253	67,680
a114 Rooms in accomodation - total	24,649	83,284
p200 Number of rooms occupied (de basis)	24,649	83,284
DVbedp Number of bedrooms occupied (harmonised)	40,253	67,680
ncounty Home county (APS only)	21,930	86,003
ndistrict Home district (APS only)	21,930	86,003
rtn_distance_km_total Air (non-business, non-UK) distance return (APS)	21,930	86,003
kgCO2_total Air (non-business, non-UK) CO2 emissions (APS)	21,930	86,003
nvideo Video recorder (EFS)	24,649	83,284

Variable and variable name	Valid	Missing
nfreezer Freezer (EFS)	24,649	83,284
ncomputer Personal computer (EFS)	24,649	83,284
ndrier Tumble drier (EFS/EHCS)	40,223	67,710
nmicrowve Microwave oven (EFS)	24,649	83,284
ndishwash Dishwasher (EFS)	24,649	83,284
ncdplay CD player (EFS)	24,649	83,284
ndvd DVD player (EFS)	24,649	83,284
nwashmach Washing machine (EFS/EHCS)	40,249	67,684
ntelephon Telephone/mobile (EFS)	24,649	83,284
nTVNum Number of TVs in hhld (EFS)	24,649	83,284
appliances Total number of appliances in hhld (harmonised)	24,649	83,284

Note: Variable order based on location in harmonised dataset

Annex Table 4. Look up table for matching CAA APS ticket type and distance with Defra/DECC CO_2 emissions factors

Defra_Flight_type_name	APS_Flight_distance	APS_Ticket	Defra_Cabin_class	e_factor_kgCO2/pkm
Short distance	< 750 km	*ALL*	Average	0.17147
Short haul international	> 750km & < 3700km	*Missing*	Average	0.097
Short haul international	> 750km & < 3700km	Business Charter	Business class	0.13867
Short haul international	> 750km & < 3700km	Business Flexible	Business class	0.13867
Short haul international	> 750km & < 3700km	Business Frequent Flier	Business class	0.13867
Short haul international	> 750km & < 3700km	Business IT	Business class	0.13867
Short haul international	> 750km & < 3700km	Business Other	Business class	0.13867
Short haul international	> 750km & < 3700km	Business Staff	Business class	0.13867
Short haul international	> 750km & < 3700km	Business/club	Business class	0.13867
Short haul international	> 750km & < 3700km	Charter seat only	Economy class	0.09245
Short haul international	> 750km & < 3700km	Charter seat upgrade	Economy class	0.09245
Short haul international	> 750km & < 3700km	Don't Know	Average	0.097
Short haul international	> 750km & < 3700km	Economy Charter Sear	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy Flexible	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy Frequent Flier	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy full fare	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy IT	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy Other	Economy class	0.09245
Short haul international	> 750km & < 3700km	Economy Staff	Economy class	0.09245
Short haul international	> 750km & < 3700km	First	Business class	0.13867
Short haul international	> 750km & < 3700km	First Charter Seat	Business class	0.13867
Short haul international	> 750km & < 3700km	First Flexible	Business class	0.13867
Short haul international	> 750km & < 3700km	First Frequent Flier	Business class	0.13867
Short haul international	> 750km & < 3700km	First IT	Business class	0.13867
Short haul international	> 750km & < 3700km	First Other	Business class	0.13867
Short haul international	> 750km & < 3700km	First Staff	Business class	0.13867
Short haul international	> 750km & < 3700km	Frequent Flier Scheme	Average	0.097
Short haul international	> 750km & < 3700km	Inclusive/package tour	Economy class	0.09245

Short haul international	> 750km & < 3700km	No Answer	Average	0.097
Short haul international	> 750km & < 3700km	Other	Average	0.097
Short haul international	> 750km & < 3700km	Other Charter Seat	Average	0.097
Short haul international	> 750km & < 3700km	Other Frequent Flier	Average	0.097
Short haul international	> 750km & < 3700km	Other IT	Average	0.097
Short haul international	> 750km & < 3700km	Other	Average	0.097
Short haul international	> 750km & < 3700km	Other Staff	Average	0.097
Short haul international	> 750km & < 3700km	P Economy Charter Seat	Business class	0.13867
Short haul international	> 750km & < 3700km	P Economy Flexible	Business class	0.13867
Short haul international	> 750km & < 3700km	P Economy Frequent Flier	Business class	0.13867
Short haul international	> 750km & < 3700km	P Economy IT	Business class	0.13867
Short haul international	> 750km & < 3700km	P Economy Other	Business class	0.13867
Short haul international	> 750km & < 3700km	P Economy Staff	Business class	0.13867
Short haul international	> 750km & < 3700km	Premium Economy	Business class	0.13867
Short haul international	> 750km & < 3700km	Refuse	Average	0.097
Short haul international	> 750km & < 3700km	Standby	Economy class	0.09245
Long haul international	> 3700km	*Missing*	Average	0.11319
Long haul international	> 3700km	Business Charter	Business class	0.23963
Long haul international	> 3700km	Business Flexible	Business class	0.23963
Long haul international	> 3700km	Business Frequent Flier	Business class	0.23963
Long haul international	> 3700km	Business IT	Business class	0.23963
Long haul international	> 3700km	Business Other	Business class	0.23963
Long haul international	> 3700km	Business Staff	Business class	0.23963
Long haul international	> 3700km	Business/club	Business class	0.23963
Long haul international	> 3700km	Charter seat only	Economy class	0.08263
Long haul international	> 3700km	Charter seat upgrade	Economy class	0.08263
Long haul international	> 3700km	Don't Know	Average	0.11319
Long haul international	> 3700km	Economy Charter Sear	Economy class	0.08263
Long haul international	> 3700km	Economy Flexible	Economy class	0.08263
Long haul international	> 3700km	Economy Frequent Flier	Economy class	0.08263
Long haul international	> 3700km	Economy full fare	Economy class	0.08263
Long haul international	> 3700km	Economy IT	Economy class	0.08263
Long haul international	> 3700km	Economy Other	Economy class	0.08263
Long haul international	> 3700km	Economy Staff	Economy class	0.08263
Long haul international	> 3700km	First	First class	0.33052
Long haul international	> 3700km	First Charter Seat	First class	0.33052
Long haul international	> 3700km	First Flexible	First class	0.33052
Long haul international	> 3700km	First Frequent Flier	First class	0.33052
Long haul international	> 3700km	First IT	First class	0.33052
Long haul international	> 3700km	First Other	First class	0.33052
Long haul international	> 3700km	First Staff	First class	0.33052
Long haul international	> 3700km	Frequent Flier Scheme	Average	0.11319
Long haul international	> 3700km	Inclusive/package tour	Economy class	0.08263
Long haul international	> 3700km	No Answer	Average	0.11319

Long haul international	> 3700km	Other	Average	0.11319
Long haul international	> 3700km	Other Charter Seat	Average	0.11319
Long haul international	> 3700km	Other Frequent Flier	Average	0.11319
Long haul international	> 3700km	Other IT	Average	0.11319
Long haul international	> 3700km	Other	Average	0.11319
Long haul international	> 3700km	Other Staff	Average	0.11319
Long haul international	> 3700km	P Economy Charter Seat	Premium economy class	0.13221
Long haul international	> 3700km	P Economy Flexible	Premium economy class	0.13221
Long haul international	> 3700km	P Economy Frequent Flier	Premium economy class	0.13221
Long haul international	> 3700km	P Economy IT	Premium economy class	0.13221
Long haul international	> 3700km	P Economy Other	Premium economy class	0.13221
Long haul international	> 3700km	P Economy Staff	Premium economy class	0.13221
Long haul international	> 3700km	Premium Economy	Premium economy class	0.13221
Long haul international	> 3700km	Refuse	Average	0.11319
Long haul international	> 3700km	Standby	Economy class	0.08263