

**DEPARTMENT OF SOCIOLOGY  
CENTRE FOR RESEARCH IN ECONOMIC SOCIOLOGY AND INNOVATION  
WORKING PAPER SERIES: ISSN 1759-2860**



---

## **HOME ON LINE SURVEY CALL RECORD DOCUMENTATION**

**CRESI WORKING PAPER NUMBER: 2011-01**

Ben Anderson    [benander@essex.ac.uk](mailto:benander@essex.ac.uk)

**Abstract:**

This paper describes the call records collected as part of the Home OnLine Household Panel study conducted by the University of Essex in collaboration with and funded by BT plc in 1997-2000. The paper is a 'living' document and will be updated as and when further details become available.

**Version:**

<b>Date</b>	<b>Version</b>	<b>Author/Editor</b>	<b>Contact</b>	<b>Commentary</b>
June 29, 2011	2.0	Ben Anderson	benander@essex.ac.uk	Revision of initial draft to include basic descriptive analysis and integration of updated code labels for charging and tariff labels as well as flagging of (some) ISP calls Addition of Annex containing codebook
July 7 2011	2.1	Ben Anderson	benander@essex.ac.uk	Incoming calls coding error fixed in datafile
July 11, 2011	2.2	Ben Anderson	benander@essex.ac.uk	Added full tables for charge_band and tariff_group codes to Annex

**Keywords/tags:**

Survey, Home OnLine, mixed-methods, call records, documentation

**Citation:**

Anderson, B. (2011) *Home OnLine Survey Call Record Documentation*, Centre for Research in Economic Sociology and Innovation (CRESI) Working Paper 2011-01, University of Essex: Colchester.

### About CRESI:

Based in the UK's leading Sociology Department, the [Centre for Research in Economic Sociology and Innovation](#) (CRESI) is the first UK centre for research in economic sociology. With a clear focus on innovation, our research programmes highlight contemporary and historical processes of socio-economic transformation. You can [read about our research](#) and [join our conversation](#).

This work is published under the [Creative Commons Attribution-Non-Commercial-No Derivative Works 2.0 UK: England & Wales License](#)

You are free:



to copy, distribute, display, and perform the work

Under the following conditions:



**Attribution.** You must give the original author credit.



**Non-Commercial.** You may not use this work for commercial purposes.



**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

**Your fair dealing and other rights are in no way affected by the above.  
This is a human-readable summary of the [Legal Code \(the full licence\)](#).**

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>5</b>
<b>2</b>	<b>Raw data.....</b>	<b>5</b>
2.1	Variables .....	5
2.2	Descriptive Statistics.....	5
2.3	Processing and error checking .....	6
<b>3</b>	<b>Anonymisation process .....</b>	<b>6</b>
<b>4</b>	<b>Anonymised data .....</b>	<b>6</b>
4.1	Content and format .....	6
4.2	Household Sample .....	7
4.3	Linked Tables.....	8
4.3.1	Call charge bands .....	9
4.3.2	Call Tariff groups .....	9
4.3.3	ISP calls .....	10
4.4	Calling patterns over time .....	11
<b>5</b>	<b>Issues .....</b>	<b>13</b>
5.1	Data 'holes' .....	13
5.2	Calls made to invalid or 'unlikely' phone numbers.....	13
5.3	Zero duration calls.....	13
5.4	Zero cost calls.....	13
<b>6</b>	<b>Summary .....</b>	<b>14</b>
<b>7</b>	<b>References.....</b>	<b>14</b>

## 1 INTRODUCTION

'Home OnLine' was a household panel study carried out by a collaboration between the University of Essex and BT between 1997 and 2000 (B Anderson, 2003, 2005, 2006a, 2006b; B. Anderson, et al., 2002; Ben Anderson & Tracey, 2001; Lacohee & Anderson, 2001).

The study was modelled on the British Household Panel Survey (Berthoud & Gershuny, 2000) and consisted of a panel of 2,434 individuals recruited from a clustered but national sample of 999 households. 842 of the individuals were interviewed in each of three waves to provide a longitudinal individual sample. The sample intentionally over-represented households with computers (50%) and only recruited those with a landline telephone. This design bias can be corrected in any analysis using an appropriate weighting scheme developed by the survey team.

Household response persons (who completed a household survey as well as a personal interview) were asked to give permission for their call records to be collected by BT for the duration of the study irrespective of whether or not they were BT customers.

At wave 2 a further 280 households were recruited to counter attrition from wave 1. Of these only those with BT as a service provider were asked to give permission resulting in 14 new households who were added to the call record collection process.

## 2 RAW DATA

### 2.1 Variables

The raw data contains the variables described in Table 1.

**Table 1: Raw data variables**

Variable name	Type	Format	Variable label
o_date_start	str8	%9s	Date call started
o_time_start	str8	%9s	Time call started
o_tariff_group_code	byte	%8.0g	Tariff group
o_chargeband_code	byte	%8.0g	Chargeband code
o_units	int	%8.0g	Number of units
o_caller_s	str11	%11s	Phone number making call
o_callee_s	str11	%11s	Phone number receiving call
o_cost	float	%9.0g	Cost of call (p)
o_duration	float	%9.0g	Duration (s)

### 2.2 Descriptive Statistics

The raw data consists of 2,622,418 records of all incoming and outgoing calls that were thought to have been made by or to the panel households that were captured by BT's billing database over four years as shown in Table 2.

**Table 2: Raw data frequencies**

Year	Freq.	Percent	Cum.
1998	724,705	27.63	27.63
1999	648,466	24.73	52.36
2000	1,002,207	38.22	90.58
2001	247,040	9.42	100.00
Total	2,622,418	100.00	

## 2.3 Processing and error checking

Inevitably the data requires considerable checking and cleaning. Of the original 2,622,418 calls, 877,980 calls did not match to a household from the panel at all. Of those that did match 92,654 + 61,589 + 4 = 154,346 matched to households who had not given permission to link the survey data to their call records (Table 3). A total of 1,032,326 (39.4%) calls were therefore rejected from the sample to be linked to the survey data.

**Table 3: Call error breakdowns**

	Permission to link call records to survey data?				Total
	No		Yes		
No match to any household	877,980	33.48%	0	0.00%	877,980
Outgoing from panel household	92,654	3.53%	984,587	37.55%	1,077,241
Incoming to panel household	61,589	2.35%	604,336	23.05%	665,925
Call between different panel households	4	0.00%	621	0.024%	625
Call between 2 numbers of the same panel household	0	0.00%	128	0.005%	128
Calls to and from the same number belonging to the same household	99	0.004%	420	0.016%	519
<b>Total</b>	<b>1,032,326</b>	<b>39.36%</b>	<b>1,590,092</b>	<b>60.64%</b>	<b>2,622,418</b>

Having removed these calls the remaining sample of 1,590,092 calls is described in Table 4.

**Table 4: Incoming and outgoing calls after 'cleaning' (variable inout)**

Type of call	1998	1999	2000	2001	Total
Outgoing from panel household	57,174	356,181	458,720	112,512	984,587
Incoming to panel household	39,800	224,339	275,135	65,062	604,336
Call between different panel households	36	271	282	32	621
Call between 2 numbers of the same panel household	9	42	68	9	128
Calls to and from the same number	17	198	161	44	420
<b>Total</b>	<b>97,036</b>	<b>581,031</b>	<b>734,366</b>	<b>177,659</b>	<b>1,590,092</b>

## 3 ANONYMISATION PROCESS

In order to ensure confidentiality and prevent disclosure all telephone numbers in the cleaned dataset were then encoded using the following algorithms:

- The last 4 digits of telephone numbers belonging to panel households have been replaced by their household serial number
- The last 4 digits of telephone numbers belonging to third parties have been transformed into alphabetic characters using a confidential coding scheme

It is therefore possible to link the households to the survey data using the household serial number.

## 4 ANONYMISED DATA

### 4.1 Content and format

**Table 5: Final cleaned data variables**

variable name	type	variable label
---------------	------	----------------

aserno	int	Household serial number
enc_callee	str19	Encrypted callee (received call)
enc_callee_miss	byte	Unknown callee? (yes=1)
enc_caller	str13	Encrypted caller (made call)
enc_caller_miss	byte	Unknown caller? (yes=1)
inout	byte	Incoming/outgoing call?
is_ispcall	byte	ISP call according to BT Dec 1998 file
isp	str26	ISP name (if is an ISP call)
o_chargeband_calldesignation	str70	General chargeband classification
o_chargeband_code	byte	Chargeband code (BT CCBA database)
o_chargeband_description	str54	Chargeband description
o_cost	float	Cost (p)
o_cost_0800	byte	Is a call to an 0800 number? (yes=1)
o_cost_miss	byte	Missing/no cost? (yes=1)
o_duration	float	Duration (s)
o_startdate	int	Date call started
o_starttime	str8	Time call started
o_tariff_group_code	byte	Tariff group code
o_tariff_group_label	str25	Tariff group description
o_units	int	Number of units
o_year	str4	Year

A full codebook for these variables detailing labels and other basic descriptive information can be found in Annex A.

## 4.2 Household Sample

As Table 6 shows there was no obvious difference in permission rates for the wave 1 households for the various telephone service operators of the time.

**Table 6: Wave 1 Household call record collection permissions by service provider (Home OnLine Survey, Wave 1)**

	No	Yes	Total n
BT	29%	71%	749
Mercury/Cable & Wireless	28%	72%	29
Cable company	44%	56%	99
Other	26%	74%	19
Don't know	25%	75%	4
Refused	0%	100%	1
Missing			98
Total	40%	70%	901

However there were clear differences in the agreement rates for different family types (Table 7) with unrelated persons and lone parents much less likely to give permission.

This potential bias means that any inferences made from the data can only be generalised to the sub-populations represented. Further analysis of additional 'non-linkage' bias within the groups who did give permission for linkage to take place would be advisable.

**Table 7: Household call record collection permissions by family type (Home OnLine Survey, Wave 1,2,3)**

	Did not agree	Agreed	Total n
Alone < 56	16%	84%	80
Alone > 55	13%	87%	128
Unrelated persons	83%	17%	35
Lone parent, children < 16	82%	18%	60
Lone parent, children > 15	77%	23%	43
Partner, no children, < 36	21%	79%	72

Partner, no children, < 56	19%	81%	67
Partner, no children, > 55	17%	83%	151
Partner, children < 12	33%	67%	152
Partner, children 11-14	37%	63%	111
Partner, children > 15	33%	67%	108
Total			1007

**Table 8: Household call record collection permissions and whether calls actually collected by family type (Home OnLine Survey, Wave 1,2,3)**

	Did not agree		Agreed		% of agreed who were logged
	Calls not logged	Calls logged	Calls not logged	Calls logged	
Alone < 56	13		38	29	43%
Alone > 55	17		49	62	56%
Unrelated persons	29			6	100%
Lone parent, children < 16	49			11	100%
Lone parent, children > 15	33			10	100%
Partner, no children, < 36	15		39	18	32%
Partner, no children, < 56	13		33	21	39%
Partner, no children, > 55	25		67	59	47%
Partner, children < 12	50		61	41	40%
Partner, children 11-14	41		39	31	44%
Partner, children > 15	36		46	26	36%
Total	321	0	372	314	46%
Total	Did not agree:		321	Agreed: 686	
Total households				1007	

Table 8 shows the breakdown of those who gave (and did not give) permission according to whether their calls were logged in the cleaned dataset. As can be seen of all the households who gave permission (Agreed) just under half actually had their call records collected. It is not clear why this was the case but several reasons are possible:

- Non-BT customers’ call records may not be visible in BT’s call logging database
- Wave 2 households’ telephone numbers failed to enter the logged sample
- Errors were made in the updating of the telephone numbers to be logged

As Table 9 shows the first of these is certainly one explanation but others should be investigated.

**Table 9: Household call record collection permissions and logging by wave 1 provider**

	Did not agree to linkage		Agreed to linkage	
	Calls not logged	Calls logged	Calls not logged	Calls logged
BT	219		246	284
Mercury/Cable & Wireless	8		19	2
Cable company	44		52	3
Other	5		13	1
Don’t know	1		3	
Refused				1

Nonetheless a sample of 314 households is sufficient for a range of analytic purposes. However, as noted above data users wishing to make generalisable claims from the data must do so with extreme caution given the potential (and potentially unknown) selection bias inherent in the data.

### 4.3 Linked Tables

In addition to the call records dataset BT have provided three look-up tables which provide information on the charge band and the tariff group of each call as well as a list of known Internet Service Provider modem numbers. This enables the coding of, for example, regional



and national calls or calls to premium rate services and the identification of at least some modem (Internet) calls. The two tables are given in full in Annex A and Annex B.

### 4.3.1 Call Tariff groups

Table 10 (see Annex A for full details) shows the tariff groups of the calls and it is interesting to note that some proportion of the outgoing calls from the panel households were classified as 'ORD BUS' or ordinary business calls. It is possible therefore that some households were running businesses from home or that there may be mis-classifications in the data. Investigation of the survey data may be able to resolve this and other similarly unexpected results.

**Table 10: Call tariff groups (variable: o\_tariff\_group\_label)**

	Outgoing from panel household	Incoming to panel household	Call between different panel households	Call between 2 numbers of the same panel household	Calls to and from the same number belonging to the same household	Total
<i>Unknown</i>	170,065	109,784	12	1	43	279,905
ISDN 2	6,343	1,769	0	0	0	8,112
ISDN 30	0	20,421	0	0	0	20,421
ORD BUS	16,259	75,423	1	5	1	91,689
ORD RES	791,920	379,690	608	122	376	1,172,716
P.C.O.	0	10,985	0	0	0	10,985
PAYPHONE LINE WITH CT	0	4,318	0	0	0	4,318
PAYPHONE LINE WITHOUT	0	1,628	0	0	0	1,628
SERVICE & TEST	0	318	0	0	0	318
<b>Total</b>	<b>984,587</b>	<b>604,336</b>	<b>621</b>	<b>128</b>	<b>420</b>	<b>1,590,092</b>

### 4.3.2 Call charge bands

As we can see ( and Annex B for full details) the vast majority of both incoming and outgoing calls are defined as local (62%). Mobile calls are subsumed in a number of charge codes including 'Mobile', 'New Services' and 'Personal Numbering Services'.

**Table 11: Call Charge bands (variable: o\_chargeband\_classification)**

	Outgoing from panel household	Incoming to panel household	Call between different panel households	Call between 2 numbers of the same panel household	Calls to and from the same number belonging to the same household	Total
<i>Unknown</i>	193,884	111,001	12	1	43	304,941
Local	585,818	405,054	609	127	377	991,985
Regional	35,012	28,745	0	0	0	63,757
National	86,295	58,822	0	0	0	145,117
International	29,009	0	0	0	0	29,009
Mobile	230	0	0	0	0	230
Other (Directory Enquiries International)	120	0	0	0	0	120
Other (Directory Enquiries)	7,437	0	0	0	0	7,437
Other (Information Services)	30	0	0	0	0	30
Other (Message Services)	6	0	0	0	0	6
Other (MOD calls from Bosnia)	0	714	0	0	0	714
Other (New Services)	17,401	0	0	0	0	17,401
Other (Paging)	278	0	0	0	0	278
Other (Personal Communications Services)	182	0	0	0	0	182
Other (Personal Numbering Services)	17,788	0	0	0	0	17,788
Premium Rate Services	8,928	0	0	0	0	8,928
Premium Rate Services	2,169	0	0	0	0	2,169
<b>Total</b>	<b>984,587</b>	<b>604,336</b>	<b>621</b>	<b>128</b>	<b>420</b>	<b>1,590,092</b>

### 4.3.3 ISP calls

Given the growth of dial-up internet access during the period of the study we should expect a number of the outgoing calls to be modem to modem calls supporting internet access. Using the file of ISP modem numbers known to BT in December 1998, Table 12 suggests that at least 52,160 calls (3.28%) were ISP calls although it is likely that the number of ISP calls is actually higher due to the use of, for example, institutional dial-up modems (e.g. workplace etc) and of ISPs not recorded in the BT file.

**Table 12: ISP (internet) calls (variable: isp)**

ISP	1998	1999	2000	2001	Total	% of Total calls
AOL	113	1,470	661	0	2,244	0.14%
AirTime Internet Reso	50	500	682	142	1,374	0.09%
ArgoNet	3	172	407	25	607	0.04%
Avel PiP	4	0	0	0	4	0.00%
Aviators Network	0	1	0	0	1	0.00%
BT	190	502	644	0	1,336	0.08%
BT Click	0	94	122	16	232	0.01%
BT Click+	1	24	0	0	25	0.00%

ClaraNET	69	178	47	0	294	0.02%
Compuserve	41	191	51	0	283	0.02%
Data International	0	1	0	0	1	0.00%
Demon	155	552	836	38	1,581	0.10%
Direct Connection	0	3	0	0	3	0.00%
EASYNET	1	0	0	0	1	0.00%
Freeserve	822	8,229	15,928	3,539	28,518	1.79%
Global Internet Ltd.	324	742	454	241	1,761	0.11%
Line One	2	702	6,187	1,159	8,050	0.51%
MSN	0	6	2	0	8	0.00%
Sprynet	9	292	339	95	735	0.05%
TescoNet	111	1,107	2,352	375	3,945	0.25%
U-NET Ltd.	1	0	0	0	1	0.00%
Virtual Chester	67	546	529	14	1,156	0.07%
Total ISP calls	1,963	15,312	29,241	5,644	52,160	3.28%
<i>No match</i>	<i>95,073</i>	<i>565,719</i>	<i>705,125</i>	<i>172,015</i>	<i>1,537,932</i>	<i>96.72%</i>
Total Calls	97,036	581,031	734,366	177,659	1,590,092	

Table 13 shows the duration of ISP and non-ISP calls and suggests that such calls are generally much longer than voice calls. In addition the table suggests that some calls are recorded as having zero duration.

**Table 13: Duration of ISP (internet) calls (variables: o\_duration by is\_ispcall)**

ISP call?	mean	min	max	sd	N
No	253.23	0	38807.86	534.12	1537932
Yes	879.32	0	56727.43	1575.77	52160

#### 4.4 Calling patterns over time

Figure 1 shows the total number of calls made each day over the duration of the study period and also the mean cost and duration of the calls. There are several spikes in the data that may represent data errors or duplications and also a number of data 'holes' that should be investigated before any significant time series analysis is carried out.

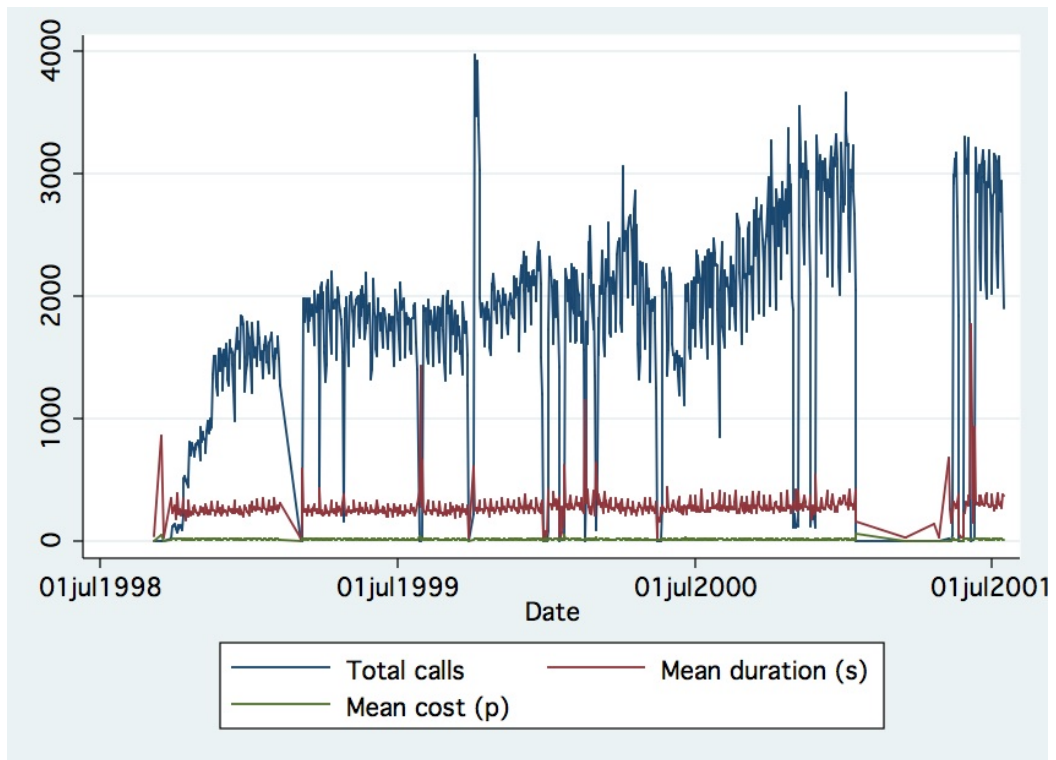


Figure 1: Total calls and mean cost/duration by day

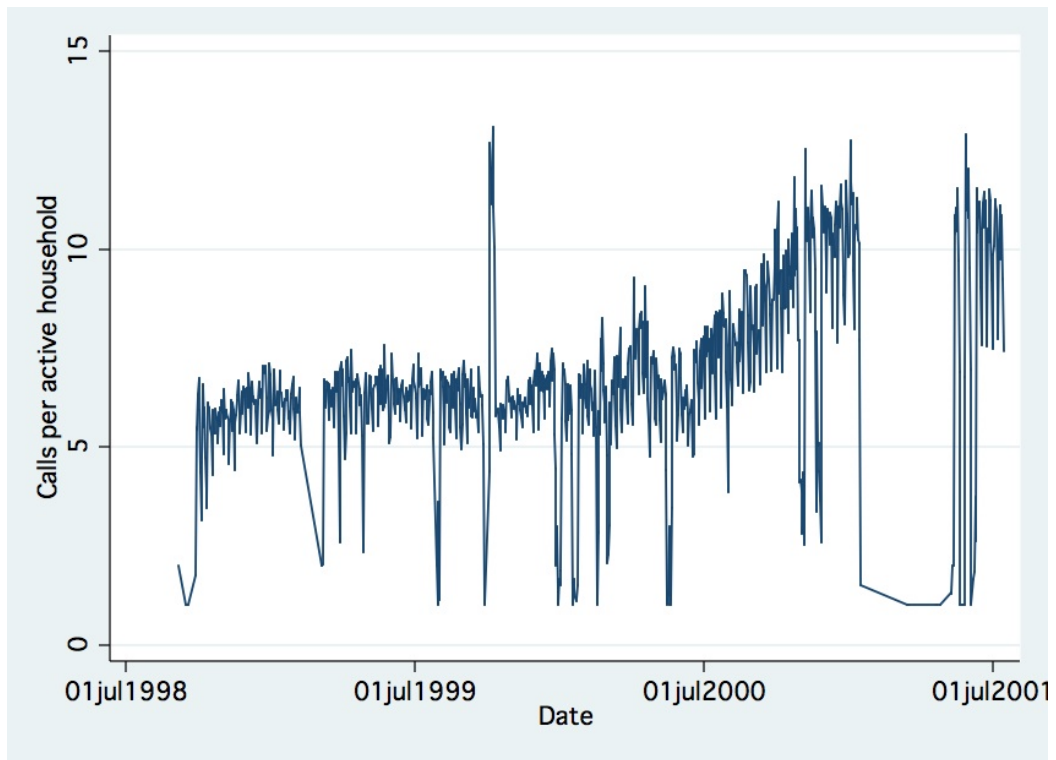


Figure 2: Number of calls per household by day

## 5 ISSUES

### 5.1 Data 'holes'

As Figure 1 makes clear there are potential data holes and/or duplications. These should be carefully investigated before any time-series analysis is performed.

### 5.2 Calls made to invalid or 'unlikely' phone numbers

A number calls were recorded which have invalid 'callee' phone numbers. These were identified as calls made by panel households either by accident or by random dialpad use, possibly by children. To some extent these can be identified using Table 14. Most of the shorter numbers match on to Directory Enquiries or premium rate services.

**Table 14: Called number length by classification (length < 10)**

	Called number (callee) length						Total
	1	4	5	6	8	9	
Unknown	368	294	6,392	375	776	172	8,377
Local	0	897	0	0	0	785	1,682
National	0	0	0	0	0	261	261
Other (Directory Enquiries)	0	7,437	0	0	0	0	7,437
Other (Directory Enquiries International)	0	120	0	0	0	0	120
Premium Rate Services	0	0	0	1	0	0	1
Premium Rate Services	0	1,143	0	0	0	0	1,143
Regional	0	0	0	0	0	196	196

### 5.3 Zero duration calls

It is not clear why there are calls with zero recorded durations (Table 15 shows that there are 19,102). These calls do not have an identified tariff group or charge band and so cannot easily be classified. Further investigation of these calls is required although they represent only  $(19,102/1,590,092) \times 100 = 1.2\%$  of the cleaned data sample. These calls are flagged by `o_duration_zero`.

**Table 15: Calls with no (zero) duration (variables: `o_duration_miss` by inout)**

	Duration > 0	Zero	Total
		duration call	
Outgoing from panel household	974,551	10,036	984,587
Incoming to panel household	595,286	9,050	604,336
Call between different panel households	615	6	621
Call between 2 numbers of the same panel household	127	1	128
Calls to and from the same number belonging to the same household	411	9	420
Total	1,570,990	19,102	1,590,092

### 5.4 Zero cost calls

It is not clear why there are so many zero cost calls although a proportion of them are explainable as 0800 or other 'freephone' calls. Further investigation of these calls is required as they represent  $(267,156/1,590,092)/100 = 16.8\%$  of the cleaned data sample. These calls are flagged by `o_cost_miss`.

Table 16 shows that 18,835 of the zero duration calls also have no cost recorded. There are also 248,321 calls which have a recorded duration but which have no cost and of these some are, as would be expected 0800 (free) numbers (Table 17) and of these 7,655 are to free ISP modem numbers (not shown).

**Table 16: Cost of calls with no (zero) duration (variables: o\_duration\_miss by o\_cost\_miss)**

Call cost?	Duration > 0	Zero duration call	Total
Cost > 0	1,322,669	267	1,322,936
Cost = 0 or missing	248,321	18,835	267,156
<b>Total</b>	<b>1,570,990</b>	<b>19,102</b>	<b>1,590,092</b>

**Table 17: Calls with no (zero) cost which are 0800 numbers (variables: o\_duration\_miss by o\_cost\_miss and o\_cost\_0800)**

	Cost > 0		Cost = 0	
	Not 0800 call	0800 call	Not 0800 call	0800 call
Outgoing from panel household	826,105	3	143,356	15,123
Incoming to panel household	495,688		108,648	
Call between different panel households	609		12	
Call between 2 numbers of the same panel household	127		1	
Calls to and from the same number belonging to the same household	404		16	
<b>Total</b>	<b>1,322,933</b>	<b>3</b>	<b>252,033</b>	<b>15,123</b>
				<b>1,590,092</b>

## 6 SUMMARY

This technical paper describes a set of 2,622,418 call records linked to a longitudinal household panel survey collected in 1998-2001. The raw call sample contains but after a number of necessary cleaning and filtering processes the usable data set comprises 1,590,092 calls made between 5 September 1998 and 15 July 2001 (Figure 1).

## 7 REFERENCES

- Anderson, B. (2003, 17-20 September 2003). *The value of longitudinal panel studies in ICT research: Transitions in and out of 'ICT poverty' as a case in point*. Paper presented at the iCS/OII Information, Communication, Society Research Symposium, Balliol College/OII, Oxford.
- Anderson, B. (2005). The value of mixed-method longitudinal panel studies in ICT research: Transitions in and out of 'ICT poverty' as a case in point. *Information Communication & Society*, 8(3), 343-367.
- Anderson, B. (2006a). *Non-response and attrition in a multi-method longitudinal household panel survey*. Paper presented at the Methodology of Longitudinal Surveys International Conference, 12-14 July 2006.
- Anderson, B. (2006b). Passing by and passing through. In R. Kraut, M. Brynin & S. Keisler (Eds.), *Computers, Phones, and the Internet: Domesticating Information Technology* (pp. 32-42). Oxford: Oxford University Press.
- Anderson, B., Gale, C., Gower, A. P., France, E. F., Jones, M. L., Lacohee, H. V., et al. (2002). Digital Living - People-Centred Innovation and Strategy. *BT Technology Journal*, 20(2), 11-29.
- Anderson, B., & Tracey, K. (2001). Digital Living: The 'Impact' or otherwise of the Internet on Everyday Life. *American Behavioral Scientist*, 45(3), 457-476.
- Berthoud, R., & Gershuny, J. (Eds.). (2000). *Seven Years in the Lives of British Families: Evidence on the Dynamics of Social Change from the British Household Panel Survey*.

Bristol: The Policy Press.

Lacohee, H., & Anderson, B. (2001). Interacting with the telephone. *International Journal of Human Computer Studies*, 54(5), 665-700.

## Annex A TARIFF GROUP CODES

File supplied by BT 29<sup>th</sup> March 1999

o_tariff_group_label	o_tariff_group_code	Freq.	Percent	Cum.
ORD RES	1	1,172,716	73.75	73.75
ORD BUS	2	91,689	5.77	79.52
OSC RES	4			
OSC BUS	5			
ISDN 2	6	8,112	0.51	80.03
ISDN 30	7	20,421	1.28	81.31
P.C.O.	8	10,985	0.69	82
PAYPHONE LINE WITHOUT CTS	9	1,628	0.1	82.11
PAYPHONE LINE WITH CTS	10	4,318	0.27	82.38
SERVICE & TEST	11	318	0.02	82.4
GLOBAL BUS (PARTIAL)	12			
?	86	18,400	1.16	83.55
?	87	39,129	2.46	86.01
?	88	216,516	13.62	99.63
?	89	2,861	0.18	99.81
?	90	2,254	0.14	99.95
?	91	16	0	99.95
?	92	195	0.01	99.97
?	93	56	0	99.97
?	98	478	0.03	100
Total		1,590,092	100	



## Annex B CHARGE BAND CODES

File supplied by BT 17<sup>th</sup> November 1997

<b>o_charge band_code</b>	<b>o_chargeband_description</b>	<b>General Classification</b>	<b>Tariff Reference File</b>	<b>CCBA Call designation</b>	<b>ACCO UNTING CHARGE BAND</b>
1	Local	Local	Local	Local	L
2	Regional (a Rate)	Regional	Regional (a Rate)	Regional	T
3	National (b Rate)	National	National (b Rate)	National	T
4	National (k Rate)	National	National (k Rate)	National	T
5	Irish Republic	International	IR	International	I
6	h Rate (New Services)	Other (New Services)	h Rate	Other	T
7	SO Rate (Premium Rate Services) FF0 Rate (Fixed Fee 0)	Premium Rate Services / Paging	SO RateFF0 Rate	Fixed to Mobile*	T
8	INTERNATIONAL 6	International	IDD 6	International	I
9	FIXED FEE 1 Rate	Premium Rate Services / Paging	FF1 Rate	Fixed to Mobile*	T
10	FIXED FEE 2 Rate	Premium Rate Services	FF2 Rate	Other*	T
11	n Rate (Information Services)	Other (Information Services)	n Rate	Other	T
12	p2 Rate (Valuecall)	Premium Rate Services	p2 Rate	Other*	T
13	p0 Rate (Premium Rate Services)	Premium Rate Services	p0 Rate	Other*	T
14	m Rate (Mobile Phones)	Mobile	m Rate	Fixed to Mobile	T
15	INTERNATIONAL 1	International	IDD 1	International	I
16	INTERNATIONAL 2	International	IDD 2	International	I
17	INTERNATIONAL 3	International	IDD 3	International	I
18	INTERNATIONAL 4	International	IDD 4	International	I
19	INTERNATIONAL 5	International	IDD 5	International	I
20	INTERNATIONAL 8A	International	IDD 8A	International	I
21	INTERNATIONAL 7	International	IDD 7	International	I
22	INTERNATIONAL 8	International	IDD 8	International	I
23	INTERNATIONAL 9	International	IDD 9	International	I
24	INTERNATIONAL 10	International	IDD 10	International	I
25	INTERNATIONAL 11	International	IDD 11	International	I
26	INTERNATIONAL 12	International	IDD 12	International	I
27	INTERNATIONAL 13	International	IDD 13	International	I
28	FIXED FEE 3 Rate	Other (Paging)	FF3 Rate	Fixed to Mobile*	T
29	p1 Rate (Premium Rate Services)	Premium Rate Services	p1 Rate	Other*	T
30	d Rate (Personal Communications Network)	Other (Personal Comms N'wk)	d Rate	Other	T
31	INTERNATIONAL ISDN 2	International	INTL ISDN 2	International	I
32	No Charge	Other (No Charge)	No Charge	Other	L
33	DQ (Int'l)	Other (Directory Enquiries International)	DQ (Int'l)	Other	T
34	Op/Access	Other (Operator Access)	Op/Access	Other*	L
35	DQ (Inland)	Other (Directory Enquiries Inland)	DQ (Inland)	Other*	T
36	FIXED FEE 6 Rate	Other (Paging)	FF6 Rate	Fixed to Mobile*	T
37	FIXED FEE 5 Rate	Other (Paging)	FF5 Rate	Fixed to Mobile*	T
38	g1 Rate	Premium Rate Services	g1 Rate	Other*	T
39	p7 Rate (Premium Rate Service)	Premium Rate Services	p7 Rate	Other*	T
40	SAT DATA 2p5 Rate (Premium Rate Services)	Premium Rate Services	Sat Data 2p5 Rate	Other*	TI
41	SAT DATA 3	International	Sat Data 3	International	I

42	INTERNATIONAL Skyphone 1	International	Skyphone 1	International	I
43	INTERNATIONAL Inmarsat-A	International	Inmarsat-A	International	I
44	INTERNATIONAL BT M-Sat	International	BT M-Sat	International	I
45	INTERNATIONAL ISDN 3	International	INTL ISDN 3	International	I
46	INTERNATIONAL ISDN 5	International	INTL ISDN 5	International	I
47	INTERNATIONAL ISDN 4	International	INTL ISDN 4	International	I
48	INTERNATIONAL BT B-Sat	International	BT B-Sat	International	I
49	INTERNATIONAL ISDN 1	International	INTL ISDN 1	International	I
50	INTERNATIONAL ISDN 6	International	INTL ISDN 6	International	I
51	Unallocated INTERNATIONAL 1a	International	Unallocated IDD 1a	International	I
52	IDD 14	International	IDD 14	International	I
53	e Rate	Mobile	e Rate	Fixed to Mobile*	T
54	c Rate	Other (Message Services)	c Rate	Other	T
55	j Rate	Mobile	j Rate	Fixed to Mobile*	T
56	Mobile Inmarsat TerminalINTC	International	MOBIQ	International	TI
57	p3 Rate (Valuecall)	Premium Rate Services	p3 Rate	Other*	T
58	p4 Rate (Valuecall)	Premium Rate Services	p4 Rate	Other*	T
59	Unallocated INTERNATIONAL 15	International	UnallocatedID D 15	International	I
60	Unallocated INTERNATIONAL 16	International	UnallocatedID D 16	International	I
61	Unallocated f Rate	Other (Personal Numbering Services / Personal Communications Networks)	Unallocatedf Rate	Other*	T
62	p6 Rate (Premium Rate Service)	Premium Rate Services	p6 Rate	Other*	T
63	SAT DATA 1	International	SAT DATA1	International	I
64	INCOMING from MoD (Bosnia)	Other (MOD calls from Bosnia)	AXE10 Bosnia	International*	I
65	National (b Rate)	National	National (b Rate)	National	T
66	Regional (a Rate)	Regional	Regional (a Rate)	Regional	T
67	UNALLOCATED				
68	UNALLOCATED				
69	UNALLOCATED				
70	ISDN NI to IR ONLY	Other (ISDN calls - NI to Eire)	ISDN NI to IR ONLY	National *	T

## Annex C CALL DATA CODEBOOK

This annex contains the codebook for the final cleaned dataset referred to in Section 4 Anonymised data.

aserno	Household serial number
-----	
type: numeric (int)	
range: [7,10198]	units: 1
unique values: 387	missing .: 0/1590092
mean: 4580.89	
std. dev: 3023.23	
percentiles: 10% 25% 50% 75% 90%	
472 1774 4655 7231 8725	
-----	
enc_callee	Encrypted callee (received call)
-----	
type: string (str19)	
unique values: 109091	missing "": 0/1590092
examples: "0128667.JIJK"	
"0153972.LHJJ"	
"0178522.JLQJ"	
"0208808.KLKL"	
-----	
enc_callee_miss	Unknown callee? (=1)
-----	
type: numeric (byte)	
range: [0,1]	units: 1
unique values: 2	missing .: 0/1590092
tabulation: Freq. Value	
1.6e+06 0	
368 1	
-----	
enc_caller	Encrypted caller (made call)
-----	
type: string (str13)	
unique values: 776	missing "": 0/1590092
examples: "0128322.2074"	
"0144078.MNNN"	
"0161834.OLHL"	
"0181682.IIJO"	
-----	
enc_caller_miss	Unknown caller? (=1)
-----	
type: numeric (byte)	
range: [0,0]	units: 1
unique values: 1	missing .: 0/1590092

tabulation: Freq. Value  
1.6e+06 0

---

inout Incoming/outgoing call?

---

type: numeric (byte)  
label: inout

range: [0,4]                      units: 1  
unique values: 5                    missing .: 0/1590092

tabulation: Freq.    Numeric    Label  
9.8e+05      0    Outgoing from panel hh  
6.0e+05      1    Incoming to panel hh  
621          2    Call between diff panel hhs  
128          3    Call between 2 numbers of same  
                 panel household  
420          4    Same number loopback within  
                 panel hh

---

is\_ispcall ISP call according to BT Dec 1998 file

---

type: numeric (byte)  
label: is\_ispcall

range: [0,1]                      units: 1  
unique values: 2                    missing .: 0/1590092

tabulation: Freq.    Numeric    Label  
1.5e+06      0    Not ISP call  
52160        1    ISP call

---

isp ISP

---

type: string (str26)

unique values: 22                    missing "": 1537932/1590092

examples: ""  
""  
""  
""

warning: variable has embedded blanks

---

o\_chargeband\_classification General Classification

---

type: string (str70)

unique values: 16                    missing "": 304941/1590092

examples: "International"  
"Local"  
"Local"  
"Local"

warning: variable has embedded blanks

---

o\_chargeband\_code (unlabeled)



2.7e+05 1 Cost = 0 or missing

---

o\_duration Duration (s)

---

type: numeric (float)

range: [0,56727.43]            units: .01  
 unique values: 160308        missing .: 0/1590092

mean: 273.767  
 std. dev: 608.172

percentiles:    10%    25%    50%    75%    90%  
                  9.49   27.11   78.21   254.28   709.61

---

o\_startdate (unlabeled)

---

type: numeric daily date (int)

range: [14127,15171]            units: 1  
 or equivalently: [05sep1998,15jul2001]    units: days  
 unique values: 876            missing .: 0/1590092

mean: 14670.6 = 01mar2000 (+ 13 hours)  
 std. dev: 267.638

percentiles:    10%    25%    50%    75%    90%  
                  14312   14446   14684   14883   14981  
                  09mar1999 21jul1999 15mar2000 30sep2000 06jan2001

---

o\_starttime (unlabeled)

---

type: string (str8)

unique values: 90740            missing "": 0/1590092

examples: "12:35:54"  
           "16:09:13"  
           "18:29:48"  
           "20:50:56"

---

o\_tariff\_group\_code (unlabeled)

---

type: numeric (byte)

range: [1,98]                    units: 1  
 unique values: 17            missing .: 0/1590092

mean: 16.5185  
 std. dev: 32.956

percentiles:    10%    25%    50%    75%    90%  
                  1      1      2      88

---

o\_tariff\_group\_label (unlabeled)

---

type: string (str25)

unique values: 8                    missing "": 279905/1590092

```

tabulation: Freq. Value
2.8e+05 ""
8112 "ISDN 2"
20421 "ISDN 30"
91689 "ORD BUS"
1.2e+06 "ORD RES"
10985 "P.C.O."
4318 "PAYPHONE LINE WITH CTS"
1628 "PAYPHONE LINE WITHOUT CTS"
318 "SERVICE & TEST"
    
```

warning: variable has embedded blanks

---

o\_units

(unlabeled)

---

type: numeric (int)

range: [0,3682]                      units: 1  
unique values: 431                    missing .: 0/1590092

mean: 3.11177  
std. dev: 13.202

percentiles:	10%	25%	50%	75%	90%
	0	1	3	7	

---

o\_year

(unlabeled)

---

type: string (str4)

unique values: 4                    missing "": 0/1590092

```

tabulation: Freq. Value
97036 "1998"
5.8e+05 "1999"
7.3e+05 "2000"
1.8e+05 "2001"
    
```