Paper 1: Using Sequence Analysis to Better Understand Interviewer Calling Behaviours: An Example from the UK Understanding Society Survey

Gabriele Durrant, Olga Maslovskaya, Peter W.F. Smith and Julia D'Arrigo
Southampton Statistical Sciences Research Institute (S3RI)
University of Southampton

Paper 2: Investigating Call Record Data using Sequence Analysis: Some Practical Considerations

Olga Maslovskaya, Gabriele Durrant, Peter W.F. Smith and Julia D'Arrigo
Southampton Statistical Sciences Research Institute (S3RI)
University of Southampton

Introduction

Survey data are increasingly collected through computer assisted modes which produce a new class of data, so-called *paradata* (Couper 1998). Paradata are by-products of the data collection process capturing information about that process. Many survey researchers hope to use paradata to improve the quality of survey data: for assessment of measurement error, to guide intervention decisions during data collection (i.e. for responsive survey designs) (Groves & Heeringa 2006) and to provide opportunities for significant cost savings.

For interviewer-mediated surveys, many survey agencies are nowadays routinely collecting information at each call to a sampling unit, here households and individuals. Such data may be referred to as call record or call history data. Examples of such data may be recordings of the day, time and outcome of each call and information collected by the interviewer such as physical and social characteristics of the selected housing unit and neighbourhood as well as information about interactions between interviewers and household members. Researchers have increasingly become interested in how best to use such information. However, such data may be complex since it is collected across time, on different sampling units and often by different interviewers. Recent research on paradata has started to analyse such call record and interviewer observation data together with other linked data to explore their potential for informing contact and response processes (Kreuter & Kohler 2009; Durrant *et al.* 2011; Durrant *et al.* 2013).

Recent research on paradata has started to analyse interviewer observation and call record data to explore their potential for informing contact and response processes (Durrant *et al.* 2011; Durrant *et al.* 2013; Kreuter & Kohler 2009). Although survey agencies have become increasingly interested in understanding and improving the process of data collection, to date, analysis of interviewer calling behaviour is still limited. Most previous studies have focussed on the final outcome of a response process, e.g. final refusal, rather than the process leading to it (Weeks *et al.* 1980; O'Muircheartaigh & Campanelli 1999; Durrant *et al.* 2011; Durrant *et al.* 2013). If contact sequences have been considered, often only summary measures from call sequences have been used in response propensity models (Groves & Heeringa 2006) rather than information from the contact sequence as a whole (Kreuter & Kohler 2009). Much of the prior research has focused on the average best times of day and days of the week to establish contact, without controlling for household characteristics and prior call information (Weeks *et al.* 1980). Greenberg & Stokes (1990) and Kulka & Weeks (1988) conditioned on previous call times but did not have householdlevel information available. Few studies controlled for basic information about the household or area, but without deriving household-specific estimates of the probability of contact (Purdon *et al.* 1999; Groves & Couper 1998; Brick *et al.* 1996; O'Muircheartaigh & Campanelli 1999). Most research on optimal calling scheduling has been carried out in the context of telephone surveys (Weeks *et al.* 1987; Greenberg & Stokes 1990; Brick *et al.* 1996) rather than face-to-face surveys, although the latter offer a much wider range of observational information available for each household and call (Greenberg & Stokes 1990). Techniques to analyse such data have often been limited to descriptive statistics and simple logistic regression modelling and usually only one survey was considered (Purdon *et al.* 1999; Groves & Couper 1996; Wood *et al.* 2006). If contact sequences have been considered, often only call record summary measures have been used in response propensity models (e.g. Groves & Heeringa 2006; Durrant *et al.* 2011; Durrant *et al.* 2013) to predict likelihood of contact or response at next call.

According to Kreuter & Kohler (2009) information that is contained in the sequence of calls has been mainly ignored by researchers. A call outcome may have a different meaning if analysed on its own or if seen as part of a sequence. For example, "no contact" call outcome has a different meaning if previous calls were also noncontacts, it then can be regarded as a longer period of no contact or if previous call was an appointment, then the current "no contact" can be regarded as a refusal (Kreuter & Kohler 2009).

To analyse the contact sequence as a whole, Kreuter & Kohler (2009) introduced the idea of sequence analysis to survey methodology. According to Kreuter & Kohler (2009), call sequence is a string of outcomes of all call attempts. A controversial idea of a continuum of resistance is widely employed in nonresponse studies (Kreuter & Kohler 2009; Lin & Schaeffer 1995). The main basis of continuum of resistance is that "with each contact, eventual survey nonrespondents are less likely to respond; hence, nonresponders are seen to be more similar to reluctant or late reponders" (Kreuter & Kohler 2009, p. 205). It is important to be able to identify late responders or nonresponders and to compare them to early respondents or nonresponders.

Sequence analysis is widely used in medicine, genetics and biology to study DNA samples, in labour economics to study employment, in demography to study union formation, childbearing, work-family trajectories and other areas. This approach is still not widely used in survey methodology but it has lots of potential to be employed in this subject area. Kreuter & Kohler (2009) and Hanly *et al.* (2013) used sequence analysis in the context of nonresponse bias adjustment. Kreuter & Kohler (2009) and Hanly *et al.* (2013) findings suggest that sequence analysis is not ideal for nonresponse adjustment. However, according to Kreuter & Kohler (2009) sequence analysis could produce useful results for survey management. This paper employs sequence analysis of call record data to inform survey management, survey design and field process work.

Sequence analysis is used in this paper to better understand the complex patterns of interviewer calls to housing units. The main aim of this paper is to explore the use of sequence analysis for investigating call record data. This paper aims to analyse the contact sequence as a whole rather than focussing on simple summary measures. The results of the sequence analysis are then used to inform the modelling of the call outcomes with the aim of identifying in particular long, unsuccessful sequences. There may be a large number of sequences in a call record data file and it is important to be able to find similarities across the contact histories and to create groups of sequences that are more homogeneous. For example, for a survey practice perspective, it is important to identify sequences with a large number of unsuccessful attempts.

Sequence analysis offers a nice way of visualising, displaying and summarising the sequences of calls. The method allows the identification of groups of sampling units with similar call sequences using a distance matrix obtained through an optimal matching. Multidimensional scaling (MDS) may be employed to summarise the structure of the data into one, two or more dimensions.

The paper explores the use of sequence analysis to inform modelling of complex call patterns such as the identification of characteristics for early and late responders and non-responders. Our modelling strategy makes use of multilevel analysis taking into account the clustering of sampling units within interviewers. The paper also highlights some practical considerations in the analysis of call record data, such as definitions of the calls and outcomes. Call record data from a longitudinal study, the UK Understanding

Society, are explored. Some guidance to survey practitioners is provided on how best to analyse call histories. The results of analysis might inform survey practitioners on strategies on how to reduce the number of unproductive calls through identification of long unsuccessful sequences and in turn how to save costs.

Data

Understanding Society is the UK Household Longitudinal Study of approximately 40,000 households in the United Kingdom (England, Scotland, Wales and Northern Ireland). Two waves of data are already available for the analysis. Wave 1 data collection took place between January 2009 and January 2011. Wave 2 took place between January 2010 and April 2012. All interviews at Wave 1 and the majority of interviews at Wave 2 (with the exception for the households from BHPS which were interviewed by telephone) were carried out face-to-face in respondents' homes by trained interviewers (McFall 2012).

According to McFall (2012, p.3), the main purpose of Understanding Society is "to provide high quality longitudinal data about subjects such as health, work, education, income, family, and social life to help understand the long term effects of social and economic change, as well as policy interventions designed to impact upon the general well-being of the UK population."

A minimum of six calls was made at each sampled address before it was considered a non-contact but interviewers were encouraged to make further calls, were possible (McFall 2012). Paradata about the interview process is available in Understanding Society survey for the analysis. These consist of call records together with outcomes for each call, timings data, interviewer observation data as well as interviewer characteristics data (McFall 2012). Call record data file is available in a long format with a separate record available for each call attempt conducted by an interviewer. It is possible to link call records data to the interviewer observation data as well as to substantive data collected during household interviews. However, the latter is only available for the responding households.

In our analytical sample we excluded Ethnic Minority Boost. Only households with at least one call were included into our sample. Households which were present in 3 or 4 issues (70 households) were also excluded from the analysis.

The call record data were converted from a long format into a wide format to first create sequences of calls and then to enable a sequence analysis using the R software package.

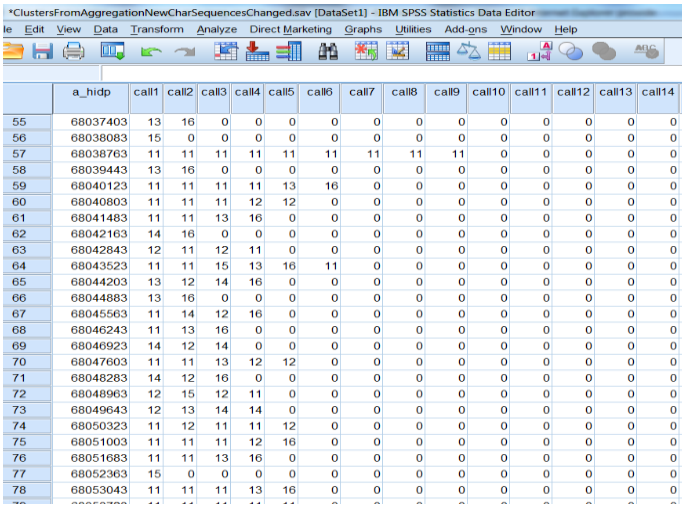
The final main sample contains 47,843 households with call sequences. This main sample was used for descriptive analysis and for modelling of grouped call sequences. These call sequences were linked with a wide range of potential explanatory variables which were used for the descriptive analysis and modelling. Slightly more than 35% of households appear in two issues. For 96% of those who appeared in two issues there was a switch of interviewer (or re-assignment of a household to another interviewer) between two issues.

Due to the limitations of the R software package (it was impossible to conduct optimal matching using all call sequences), the decision was taken to select a subsample of 5% randomly selected interviewers who conducted interviews at issue 1 and all sequences associated with those interviewers. 42 interviewers were selected and the final sub-sample contained 2,438 call sequences. The reason for selection of sub-sample of interviewers rather than households is to enable analysis of interviewer behaviour. This subsample is used to conduct some descriptive techniques available in sequence analysis such as MDS.

In the call record dataset there is a variable "call outcome" which has 5 categories: "no reply" (in our datafile coded as 11), "contact made" (as 12), "appointment made" (as 13), "any interviewing done" (as 14), and "any other status" (as 15). First four categories are clear but unfortunately the fifth category "any other status" combines a wide range of possible call outcomes (anything what would not fit within the first four categories) from different types of refusals to different ineligibility statuses. We need to treat this category with caution when interpreting the results of sequence analysis due to its ambiguity.

During data preparation stage we split the last occurrence of "any interviewing done" in each appropriate sequence into two categories based on the final household outcome variable: the first category was created for those households in which household interview together with all individual interviews were completed (coded as 16) and the second category was created for all other households (still were coded as 14). Figure 1 presents some of the sequences of calls in our datafile. For example, the sequence which has ID number 56 is a very short sequence with only one call and with the call outcome being "any other status". Sequence with ID number 57 is a relatively long sequence with 9 calls and outcomes for each calls are "no contact". Sequence with the ID number 67 consists of 4 calls, at the first attempt "no contact" was recorded, then one interview was done, then contact was made and then another interview was done and in this household household interview and all individual interviews were completed.

Figure 1: Sequences in the dataset



*ClustersFromAggregationNewCharSequencesChanged.sav [DataSet1] - IBM SPSS Statistics Data Editor

le Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	a_hidp	call1	call2	call3	call4	call5	call6	call7	call8	call9	call10	call11	call12	call13	call14
55	68037403	13	16	0	0	0	0	0	0	0	0	0	0	0	0
56	68038083	15	0	0	0	0	0	0	0	0	0	0	0	0	0
57	68038763	11	11	11	11	11	11	11	11	11	0	0	0	0	0
58	68039443	13	16	0	0	0	0	0	0	0	0	0	0	0	0
59	68040123	11	11	11	11	13	16	0	0	0	0	0	0	0	0
60	68040803	11	11	11	12	12	0	0	0	0	0	0	0	0	0
61	68041483	11	11	13	16	0	0	0	0	0	0	0	0	0	0
62	68042163	14	16	0	0	0	0	0	0	0	0	0	0	0	0
63	68042843	12	11	12	11	0	0	0	0	0	0	0	0	0	0
64	68043523	11	11	15	13	16	11	0	0	0	0	0	0	0	0
65	68044203	13	12	14	16	0	0	0	0	0	0	0	0	0	0
66	68044883	13	16	0	0	0	0	0	0	0	0	0	0	0	0
67	68045563	11	14	12	16	0	0	0	0	0	0	0	0	0	0
68	68046243	11	13	16	0	0	0	0	0	0	0	0	0	0	0
69	68046923	14	12	14	0	0	0	0	0	0	0	0	0	0	0
70	68047603	11	11	13	12	12	0	0	0	0	0	0	0	0	0
71	68048283	14	12	16	0	0	0	0	0	0	0	0	0	0	0
72	68048963	12	15	12	11	0	0	0	0	0	0	0	0	0	0
73	68049643	12	13	14	14	0	0	0	0	0	0	0	0	0	0
74	68050323	11	12	11	11	12	0	0	0	0	0	0	0	0	0
75	68051003	11	11	11	12	16	0	0	0	0	0	0	0	0	0
76	68051683	11	11	13	16	0	0	0	0	0	0	0	0	0	0
77	68052363	15	0	0	0	0	0	0	0	0	0	0	0	0	0
78	68053043	11	11	11	13	16	0	0	0	0	0	0	0	0	0

The results of any study are dependent on the quality of data which are used for the analysis. Our preliminary data analysis suggests that there is a problem with the final outcome status and the sequences themselves which require further investigation. For example, in sequence with the ID number 64 (see Figure 1) the interviewer goes back to the household after the entire interviewing process has already been completed. As mentioned earlier, the call outcome category "any other status" presents another challenge for the interpretation of the results due to this category's ambiguity. There is a possibility that the interviewer coding might be imperfect but any other studies will face these limitations. Another potential problem with the data is related to the group of variables which are collected through interviewer observations as interviewers' judgements are subjective and can be also imperfect at times.

Methods

As mentioned earlier, sequence analysis is widely used in medicine, genetics and biology to study DNA samples, in labour economics to study employment, in demography to study union formation, childbearing, work-family trajectories and other areas. This approach is still not widely used in survey methodology but it has lots of potential to be employed in this subject area as large-scale survey data collect call record data which can be easily converted in call sequences.

Once call sequences are obtained, first descriptive analysis of call sequences will be conducted. The main aim of this first part of the analysis is to identify patterns in different sequences and similarities between these sequences in order to group them most effectively. No single characteristic would fully describe the large difference between sequences. A combination of different characteristics of sequences might be necessary to establish an adequate measure of similarity across all sequences in our data. Different descriptive measures of sequences will be obtained and plotted. Then optimal matching (Levenshtein 1966; Sankoff & Kruskal 1983; Abbott & Hrycak 1990; Abbott & Tsay 2000) will be performed to enable cluster analysis (Bartholomew *et al.* 2008) and multidimensional scaling (Kruskal & Wish 1978; Bartholomew *et al.* 2008) which will inform best ways to summarise sequences and to obtain summary measures which will be used in further analysis. Optimal matching produces a matrix of distance measures for all sequences which then can be used to measure similarity/difference between sequences. On the basis of the measures of similarity between sequences, the sequences can then be grouped using cluster analysis. These measures of similarity can also be used for multidimensional scaling which allows choosing a solution with one or more dimensions and then each sequence is assigned a location on one or several dimensions. Summary measures obtained through help of cluster analysis and multidimensional scaling will then be assessed and compared. The combination of their main characteristics will then inform the final most adequate summary measures used for further analysis. These measures will then be modelled in the multilevel environment to account for clustering effect of interviewers. Multilevel multinomial logistic regression will be employed for the analysis of the final summary measure.

SPSS software package version 20 (IBM Corp. 2011) is used at the data preparation stage as well as at initial data modelling stage. R software package and specifically designed *TraMineR* library (Gabadinho *et al.*. 2011) are used to conduct sequences analysis. *WeightedCluster* library (Studer 2013) is used to enable cluster analysis for the whole analytical sample. This library conducts optimal matching only on subset of unique sequences (it conducts analysis on aggregated data); it then performs cluster analysis within this aggregated dataset by partitioning the dataset around medoids. The results of the cluster analysis are then disaggregated to the whole analytical sample. MLwiN 2.24 (Rasbash *et al.* 2012) is used to conduct multilevel modelling. R software packaged is chosen over STATA as according to Brzinsky-Fay *et al.* (2006), STATA is able to perform optimal matching with a moderate number of relatively short sequences. Unfortunately, R software package has the same limitation but is able to obtain distance matrices for larger number of sequences than STATA is, it can also use longer sequences. Due to this limitation of R, optimal matching is not possible on our complete analytical sample. Therefore, optimal matching and multidimensional scaling are performed on a subsample of 5% of randomly selected interviewers discussed in the data section of the paper.

Results

Descriptive Analysis

The results of the descriptive analysis suggest that there are 47,843 call sequences with 12,787 distinct sequences in our analytical sample. The minimum length of sequences is one call and the maximum length is 30 calls. The mean length of the sequences is 5.33 calls whereas the median length of sequences is 4. Therefore, the distribution of call sequences is therefore positively skewed.

Figure 2 presents index plot of all sequences in our analytical sample. This graph shows the distribution of sequences ordered by length and appearance in the dataset. The horizontal axis presents number of calls whereas vertical axis shows the running case count. Different colours represent different call outcomes. Each horizontal line in the plot represents a call record for one household, i.e. one sequence. The results of the descriptive analysis show that 63% of all sequences have length between 1 and 5 calls. Only a small proportion of the sequences (11.6%) has length longer than 10 calls.

Figure 2: Sequence index plot of all sequences

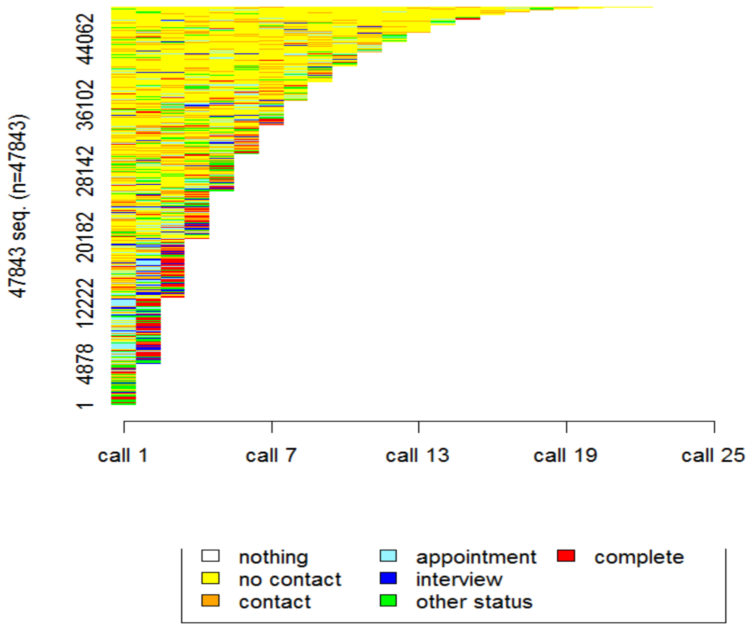


Figure 3 shows 10 most frequent sequences in our analytical sample, they represent 23.8% of the sample. The first most frequent sequence is a sequence which contains 2 calls with the first one being an “appointment made” call and the second call being the successful completion of the interviewing process. These sequences represent 5.8% of the sample. The second and the third most frequent sequences contain one call and the first one (4.5%) has “any other status” outcome which is an ambiguous category containing a wide range of possible outcomes from different types of refusals to different ineligibility statuses. The second one (nearly 3% of all sequences) are the ones which represent successful completion of the interviewing process. Figure 3 also shows that all 10 most frequent sequences are relatively short and contain no more than four calls. It also shows that none of these sequences has a very large occurrence with the largest one being 5.8%. The 9th group which contains “contact made” outcomes is an interesting group as there is no follow up after the contact had been established despite the requirement for all interviewers which suggest that a minimum of six calls were supposed to be made at each household (McFall 2012). The fifth group is also an interesting one as it finishes with a status any interview is done which suggests that there are still individuals in a household who were not interviewed or a household interview was not completed. However, there is no follow up to complete the interviewing process. This graph is helpful in visualising sequences and can also help in assessing compliance with a survey protocol which prescribed what interviewers should be doing.

Figure 3: Ten most frequent sequences

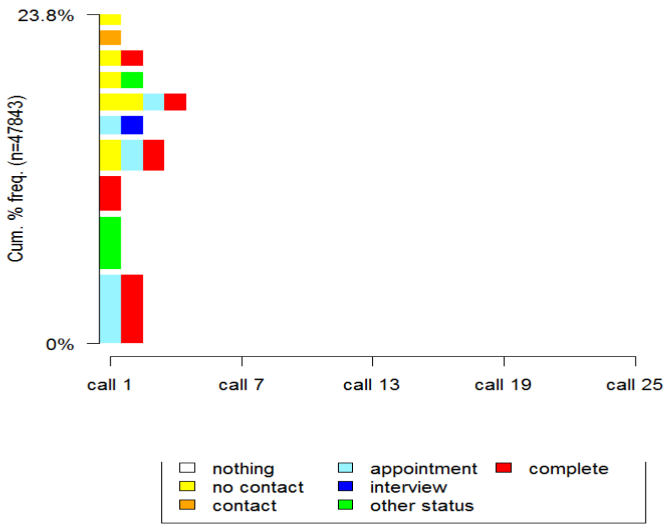


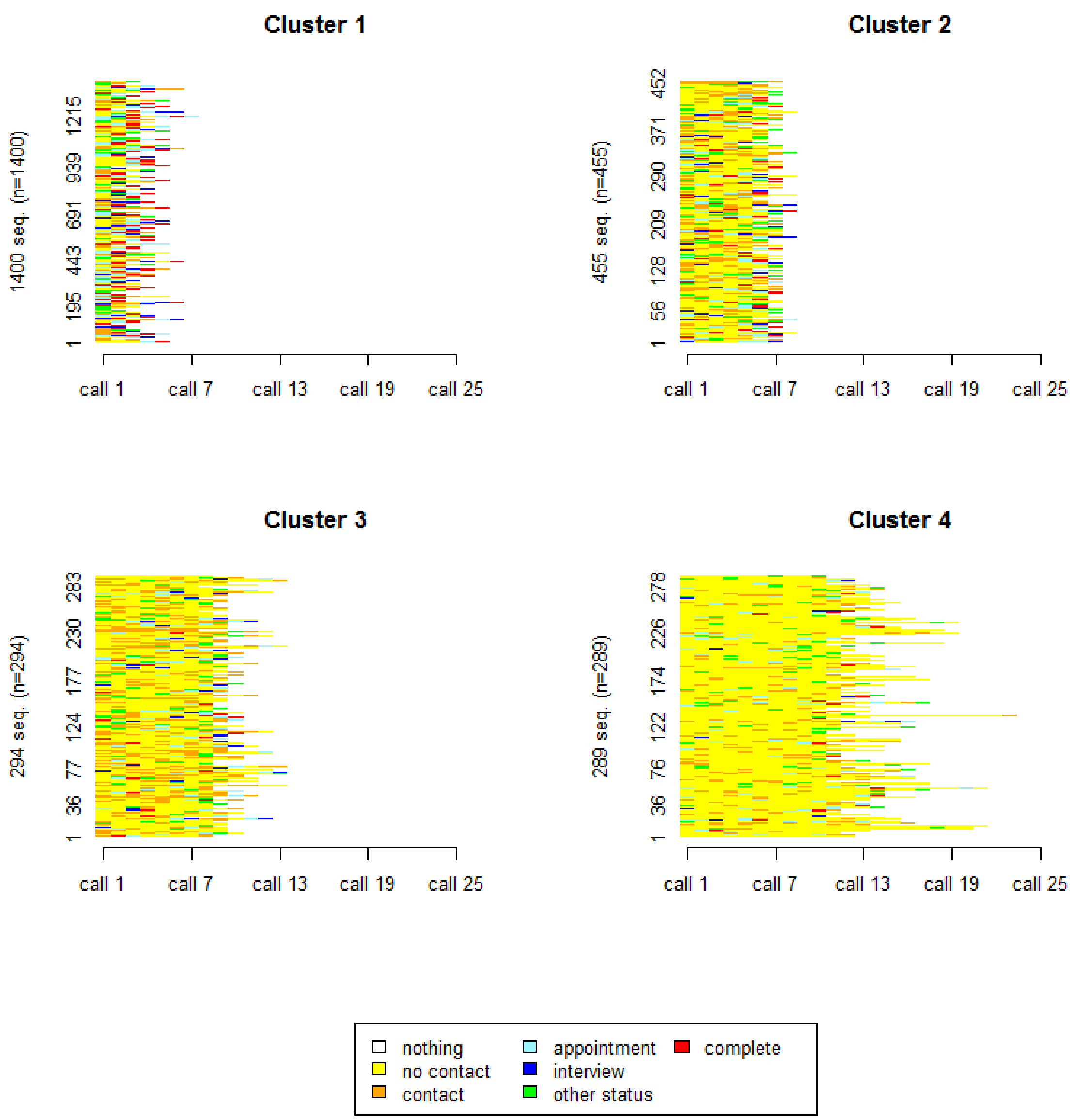
Table 1 presents transition rates between different call outcomes and shows how likely is the next outcome following a particular call outcome. Rates in rows add up to 1. This table shows that the probability of having a "no contact" outcome after a "no contact" is 63%. The probability of having a "no contact" after a "contact made" outcome is 38%. The probability of having a "no reply" outcome after "any interviewing done" is 21% and to get to the end of sequence after an interview is 41%. The probability of having a "no reply" outcome after "any other status" is 29% and to get to the end of sequence after this outcome is 46%. The highest probability of having an interview or a completion of the interviewing process is observed after an "appointment made" status. This table shows unusual cases in which we observe activities (with the probabilities being between 1% and 12%) after the interviewing process has been successfully completed. These cases require further investigations as they unnecessarily increase costs of the survey process.

Table 1: Transition rates

	To end of sequence	To no contact	To contact	To appointment	To any interview	To any other status	To complete
From no contact	0.07	0.63	0.12	0.08	0.01	0.07	0.02
From contact	0.19	0.38	0.20	0.09	0.03	0.08	0.04
From appointment	0.10	0.17	0.06	0.04	0.20	0.04	0.38
From any interview	0.42	0.17	0.11	0.06	0.05	0.02	0.15
From any other status	0.46	0.29	0.10	0.03	0.01	0.10	0.01
From complete	0.78	0.12	0.06	0.02	0.00	0.01	0.00

Cluster analysis is conducted on the complete analytical sample with the help of *WeightedCluster* library in R. Different solutions with different number of clusters (3, 4 and 5) were obtained. Figure 4 presents the solution with four clusters. Figure 4 suggests that the main characteristic which defines cluster membership is length of sequences. Clusters 1 and 2 contains shorter sequences while Clusters 3 and 4 longer ones. The main limitation of cluster analysis is that it is difficult to clearly define each cluster as, for example, sequences which have length 3-10 calls are found in both Cluster 1 and Cluster 2. Therefore, it would be difficult to use the summary measure obtained with the help of cluster analysis for further modelling. However, results of the cluster analysis suggest that length of sequences represent one of the most important characteristics of sequences. The final summary measure should take this characteristic into account.

Figure 4: Cluster analysis: Four-cluster solution



Multidimensional scaling was conducted on the subsample of our analytical sample based on 5% of randomly selected interviewers due to limitations of the R software. Figure 5 presents a one-dimension solution and Figure 6 presents a two-dimension solution. Horizontal axis of Figure 5 shows sequences ordered by the first dimension whereas vertical axis shows number of calls in sequences. One-dimension solution shows the importance of length and is in agreement with the results of cluster analysis discussed earlier and presented in Figure 4. The horizontal axis of Figure 6 shows the first dimension where sequences are ordered by length. The vertical axis shows the second dimension where sequences are ordered by different call outcomes with the top part of the dimension representing long unsuccessful sequences which have many "no contact" outcomes, middle part of the dimension representing short successful sequences and the bottom part of the dimension mainly representing sequences with "contact made" outcomes but without successful completion of the process. Sequences presented in blue colour represent interviewing processes when no interviews were achieved and red sequences represent processes with at least one successful interview. The sequences in the top right corner in blue are the long unsuccessful ones which increase the costs unnecessarily and ideally should be avoided. The red ones in the middle of the second dimension and at the left part of the first dimension are the successful ones. Figure 6 shows that the two-dimension solution orders sequences by length and by outcome and suggest the importance of these two characteristics. These results are in agreement with the findings obtained by Kreuter & Kohler (2009) when they conducted MDS using ESS surveys. The results of the MDS suggest that the two most important characteristics which need to be taken into account when constructing a summary measure for further analysis are length and final household outcome. MDS plots can be effectively used to identify different clusters of sequences with different characteristics. They can also be produced, for example, per interviewer and be used as indicators for interviewer performance.

Figure 5: Multidimensional scaling: One-dimension solution

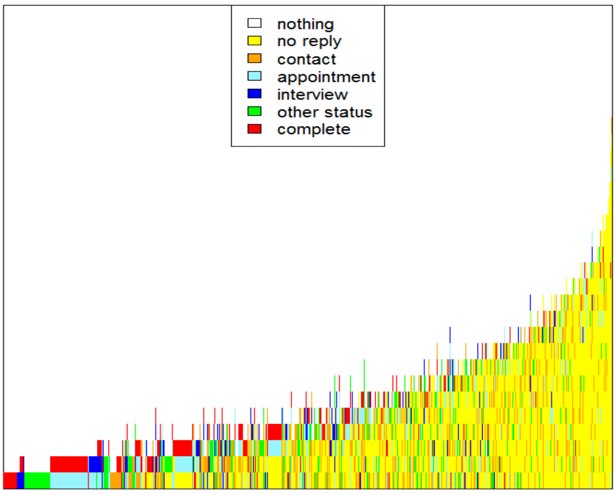
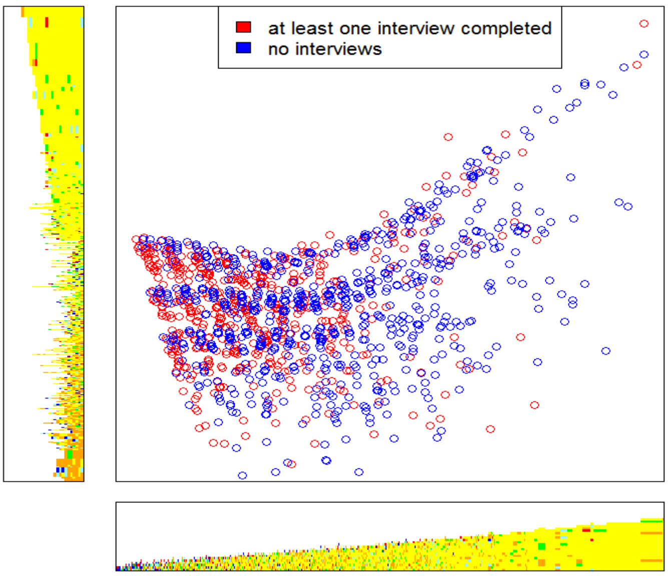


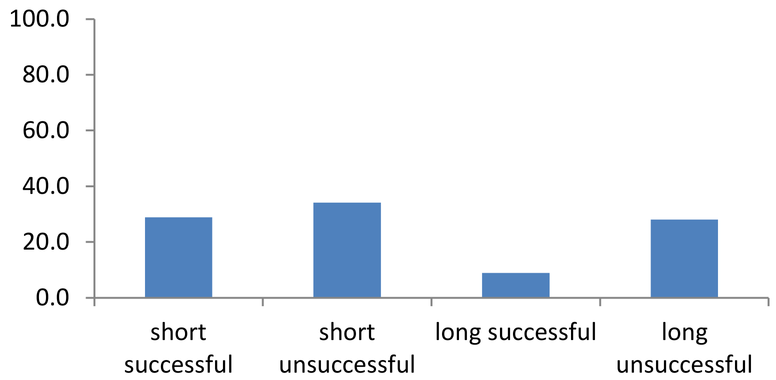
Figure 6: Multidimensional scaling: Two-dimension solution



Multilevel multinomial logistic regression

The results of the cluster analysis and multidimensional scaling informed the main characteristics which are taken into account when the main variable of interest is constructed. The main variable of interest has four categories which differ by length and the final household outcome. Short sequences have 1-5 calls and long sequences have 6-30 calls. Successful sequences are those which have household interview and all individual interviews completed, all other sequences are considered being unsuccessful. On the basis of these two main characteristic the main variable of interest which is modelled later has the following categories: short successful sequences, short unsuccessful sequences, long successful sequences, and long unsuccessful sequences. Figure 7 shows percentages in each group of the main variable of interest. The most prevalent group in our analytical sample is short unsuccessful sequences group whereas the least prevalent is the long successful sequences group.

Figure 7: Types of sequences (main summary measure)



The main variable of interest (types of sequences) is modelled using multilevel multinomial logistic regression to identify factors associated with different types of sequences. Explanatory variables used for the analysis belong to two main groups of variables: geographical and interview process characteristics and interviewers' observations. The following variables are found to be significantly associated with the main variable of interest: month of sample issue, low density area for ethnic minorities, Government Offices for the Regions (GORs), urban or rural area, type of accommodation, number of floors, locked gates, unkempt garden, car or van, children, boarded houses, abandoned buildings, demolished houses or demolished buildings, heavy traffic, condition of residential property, and relative condition of residential property. We also found a significant random interviewer effect. Figures 8-11 show selected results in the form of predicted probabilities for being in a certain group by selected characteristics while all other characteristics are being held at their means.

Figure 8 shows that the probability of having a short successful sequences is higher in bedsitters and sheltered accommodation when compared to other types of accommodation. The probability of having short unsuccessful sequences is the highest in sheltered accommodation and the lowest in flats and maisonettes. The probability of having long unsuccessful sequences is the highest in dwellings with business premises when compared to other types of accommodations.

Figure 8: Multilevel modelling: Types of sequences by types of accommodation

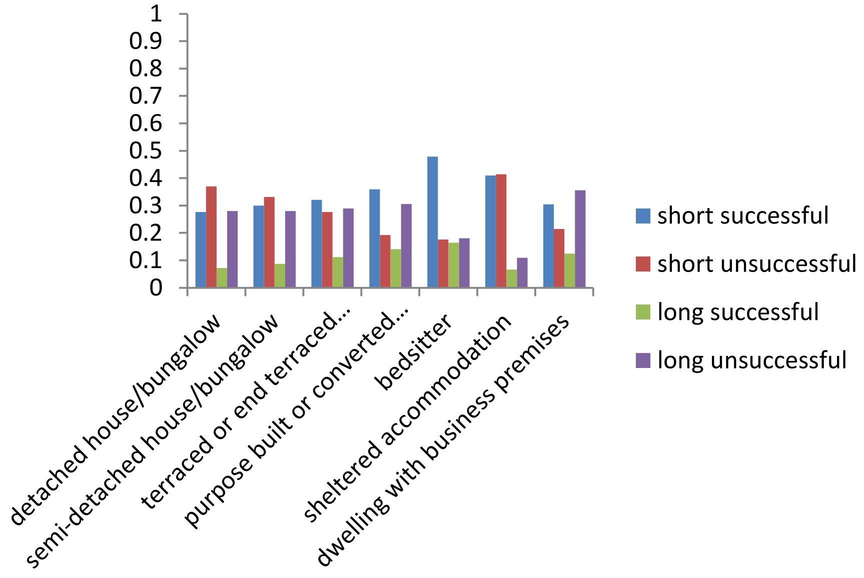


Figure 9 suggests that the probability of having long unsuccessful sequences is higher in buildings where locked gates are reported.

Figure 9: Multilevel modelling: Types of sequences by presence of locked gates

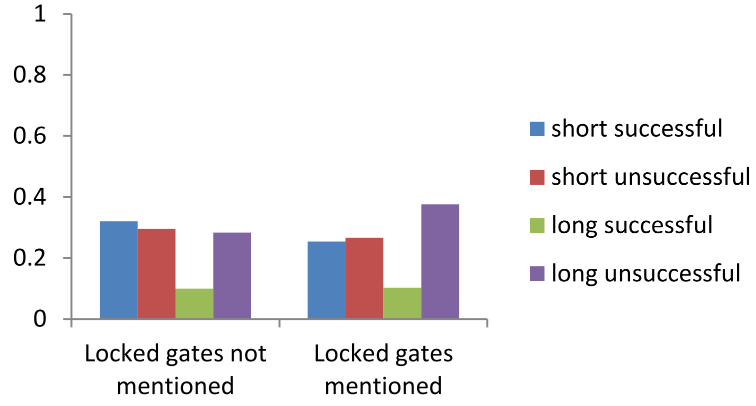


Figure 10 shows that that the probability of having short successful sequences is the highest when the property definitely does not have a car. The probability of having short unsuccessful sequences is the highest when it is unlikely that a car or a van is present.

Figure 10: Multilevel modelling: Types of sequences by presence of a car or a van

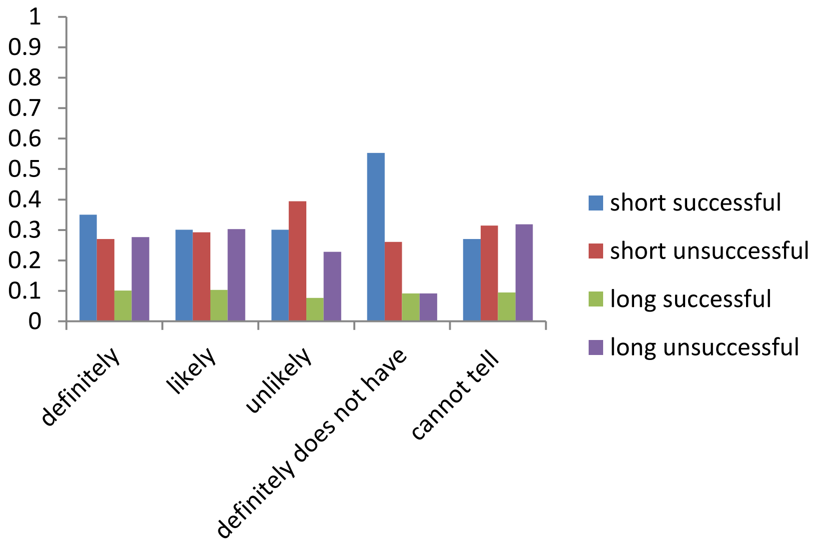
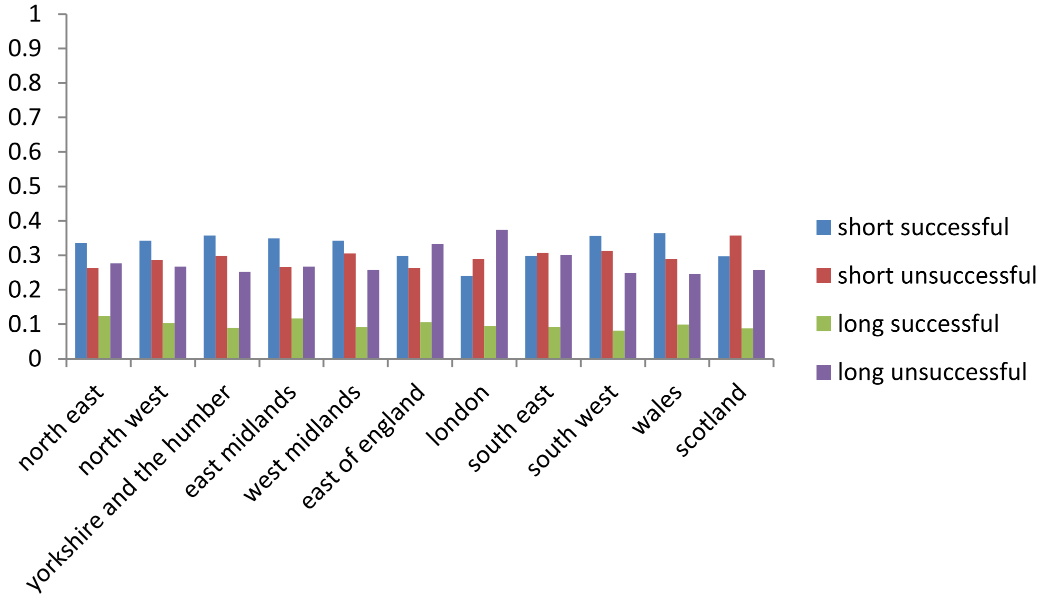


Figure 11 shows that London has the highest probability of having long unsuccessful sequences when compared to other GORs, followed by East of England. The probability of having short unsuccessful sequences is the highest in Scotland.

Figure 11: Multilevel modelling: Types of sequences by GOR



Discussion

This paper used sequence analysis to analyse call record data using Understanding Society data in the UK. Sequence analysis proved useful when the whole call sequence is of interest. This methodological approach helped to better understand the complex patterns of interviewer calls to housing units. Sequence analysis was used to identify the best way of summarising large number of different sequences into a manageable number of groups for further analysis. This analysis helped to identify long unsuccessful sequences which might be of primary interest for survey practitioners. Sequence analysis and especially plots can be employed as a useful tool for survey practitioners to spot problems during and after the fieldworks. Ideally it would be very beneficial if the problem spotting could become an automated procedure which would flag problems automatically.

Multilevel modelling helped to identify the main characteristics which are associated with the probability of households to have a certain types of call sequences. These results might be useful for improving survey management process.

Compared to call record sequences in other countries (see Kreuter & Kohler 2009 for examples from the ESS) we find that the UK seems to have longer call records This may have important implications for survey practice. Further research should be conducted to address the question why this is happening in the UK.

The findings from this research will ultimately inform survey practitioners about interviewer calling behaviours and response processes, overall and within subgroups of both interviewers and respondents.

The main limitation of the work is related to the limitations of the R software package and its inability to obtain a distance matrix using all of the 47,843 sequences. The maximum number of sequences possible to use is around 35,000. Alternative software packages such as STATA were considered but had similar limitations. It was, therefore, impossible to conduct MDS using the whole analytical sample. However, the results of the MDS employed on a smaller sample are in agreement with the results obtained by other studies (e.g. Kreuter & Kohler 2009).

Another limitation of the work comes from the ambiguity of the code 'any other status' which contains too many possible outcomes. This influences the interpretations of the results of the analysis.

It is not possible to control for number of people in household during the modelling process as the survey does not have this information recorded for the ineligible household or refusals.

Another limitation of the work is related to the potential subjectivity of the interviewers' observations which might lead to imperfect representation of the real situation (Sinibaldi *et al.* 2013). However, this limitation is not unique to these data.

Short unsuccessful sequences might be problematic due to the reason that an interviewer might terminate the process earlier than necessary. Therefore, in reality some of the short unsuccessful sequences might belong to the group of long unsuccessful sequences. The arbitrary decision about definition of short versus long sequences might also introduce a bias into the analysis.

Previous research has focussed primarily on cross-sectional surveys. However, a better understanding of call patterns may be most beneficial for longitudinal surveys where information from the current and previous waves as well as a wealth of information for selected housing units are available. The call record data analysed here comes from a new large-scale longitudinal survey in the UK, Understanding Society. The next step in the analysis is to repeat and expand the same work for subsequent waves of the Understanding Society data. Another avenue for further work is to use sequence analysis broken down by subgroups (e.g. interviewers) to inform survey practice. This analysis can be useful when interviewer performance is of interest. The findings from this research will ultimately inform survey practitioners about interviewer calling behaviours and response processes overall and within subgroups of both interviewers and respondents.

Kreuter & Kohler (2009) find that the UK has longer call sequences when compared to other countries they used for the analysis. Our findings are in agreement with the Kreuter & Kohler (2009) findings and also report longer sequences. Another avenue for further work would be to identify the reasons why the UK has longer call record sequences and these reasons can have important implications for the survey management practices.

References

Abbott, A. and A. Hrycak (1990). "Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers." American journal of sociology: 144-185.

Abbott, A. and A. Tsay (2000). "Sequence analysis and optimal matching methods in sociology review and prospect." Sociological Methods & Research **29**(1): 3-33.

Bartholomew, D., Steele, F., Galbraith, J. and I. Moustaki (2008). Analysis of multivariate social science data, Routledge.

Brick, J. M., Allen, B. and P. Cunningham (1996) "Outcomes of a calling protocol in a telephone survey." Proceedings of the Survey Research Methods Section of the American Statistical Association: 142–149.

Brzinsky-Fay, C., Kohler, U. and M. Luniak (2006). "Sequence analysis with Stata." Stata Journal **6**(4): 435.

Couper, M. P. (1998) "Measuring survey quality in a CASIC environment." Proceedings of the Survey Research Methods Section, American Statistical Association **48**: 743–772.

Durrant, G.B. and F. Steele (2009). "Multilevel Modelling of Refusal and Noncontact Nonresponse in Household Surveys: Evidence from Six UK Government Surveys." Journal of the Royal Statistical Society, Series A **172** (2): 361-381.

Durrant, G.B., D'Arrigo, J. and F. Steele (2011). "Using Field Process Data to Predict Best Times of Contact Conditioning on Household and Interviewer Influences." Journal of the Royal Statistical Society, Series A **174** (4): 1029-1049.

Gabadinho, A., Ritschard, G., Mueller N.S. and M. Studer (2011). "Analyzing and visualizing state sequences in R with TraMineR."

Greenberg, B. S. and S. L. Stokes (1990). "Developing an optimal call scheduling strategy for a telephone survey." Journal of Official Statistics **6**(4): 421-435.

Groves, R. M. and M. P. Couper (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Groves, R. M. and S. G. Heeringa (2006). "Responsive design for household surveys: Tools for actively controlling survey errors and costs." Journal of the Royal Statistical Society: Series A (Statistics in Society) **169**(3): 439-457.

Hanly, M., Clarke, P. and F. Steel (2013) "Sequence analysis of call record data: Are sequence typologies useful for nonresponse adjustment?" 5th ESRA conference, Ljubljana, Slovenia.

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Kreuter, F. and U. Kohler (2009). "Analyzing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey." Journal of Official Statistics **25**(2): 203.

Kruskal, J. and M. Wish (1978). Multidimensional scaling. Beverly Hills, CA.

Kulka, R. A. and M. F. Weeks (1988). "Toward the development of optimal calling protocols for telephone surveys: a conditional probabilities approach." Journal of Official Statistics **4**(4): 319-332.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. Soviet physics doklady.

Lin, I.-F. and N. C. Schaeffer (1995). "Using survey participants to estimate the impact of nonparticipation." Public Opinion Quarterly **59**(2): 236-258.

McFall, S. L. (ed.) (2012). *Understanding Society* –UK Household Longitudinal Study: Wave 1-2, 2009-2011, User Manual. Colchester: University of Essex .

O'Muircheartaigh, C. and P. Campanelli (1999). "A multilevel exploration of the role of interviewers in survey non-response." Journal of the Royal Statistical Society: Series A (Statistics in Society) **162**(3): 437-446.

Purdon, S., Campanelli, P. and P. Sturgis (1999), "Interviewer's calling strategies on face-to-face interview surveys." Journal of Official Statistics **15**(2): 199-216.

Rasbash, J., Steele, F., Browne, W.J. and H. Goldstein (2012). A user's guide to MlwiN, v. 2.26. Centre for Multilevel Modelling, University of Bristol.

Sankoff, D. and J. B. Kruskal (1983). Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Reading: Addison-Wesley Publication.

Sinibaldi, J., Durrant, G.B. and F. Kreuter (2013). „Evaluating the Measurement Error of Interviewer Observed Paradata." *Public Opinion Quarterly*, Special issue: Topics in Survey Measurement and Public Opinion **77** (1): 173-193.

Studer, M. (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers **24**: 1-33.

Weeks, M. F., Jones, B. L., Folsom, R. E. and C. H. Benrud (1980). "Optimal times to contact sample households." Public Opinion Quarterly **44**(1): 101-114.

Weeks, M. F., Kulka, R. A. and S. A. Pierson (1987). "Optimal call scheduling for a telephone survey." Public Opinion Quarterly **51**(4): 540-549.

Wood, A. M., White, I. R. and M. Hotopf (2006). "Using number of failed contact attempts to adjust for non-ignorable non-response." Journal of the Royal Statistical Society: Series A (Statistics in Society) **169**(3): 525-542.