Tracking And Rotation – The Next Steps

Jo Edwards, Office for National Statistics

1. Introduction

As stated in Setchfield, 2007, in order to meet new requirements for the European Union's Survey of Income and Living Conditions (EU-SILC), the General Household Survey became longitudinal in 2005. It is now called the General Lifestyle Survey (GLF). The result was the implementation of a rotating four-year panel design. Whilst longitudinal social surveys were not new for ONS, this was the first one that followed all adult persons with the original sampled household for four years, unless they died, moved to an institution or moved abroad. Other characteristics include interviewing adult household members if they join after wave 1 as long as the live with an original sampled household member.

In 2006, a new method of rotating data, and tracking household members was developed. This method catered to the addition and removal of household members, including moving non-residents into new households for either interviewing in the case of movers or processing in the case of those who became ineligible.

Since this time, ONS has taken on two new longitudinal surveys – the Household Assets Survey (HAS) and the Lifestyle Opportunity Survey (LOS). Like GLF, both of these surveys follow all adult members of the original sampled household. HAS follows up eligible respondents after 2 years. However, if respondents show certain characteristics, they are followed up after 1 year for a Household Debtors Survey, which is conducted by telephone. LOS follows households after 1 year.

A Keeping-in-Touch Exercise (KITE) survey is run about 3-4 months prior to the face-to-face survey for all the surveys. The KITE survey checks for movers, update address and contact information, and in the case of HAS checks for joiners and collects basic information, and in the case of LOS will additionally check for onset of disabilities. In the case of LOS, if onset of disability is recorded, that household becomes eligible for a face-to-face interview at the next wave.

ONS is reviewing current processes and practices in relation to its longitudinal social surveys. Several gaps and improvements have been identified, and work is ongoing to define the issues more thoroughly so that appropriate solutions can be implemented. This paper will primarily concentrate on the GLF and will discuss some of these gaps and solutions, although at the time of writing, they are yet to be finalised and implemented.

2. Current Methods.

GLF and HAS both currently use the same methods for tracking and rotation. The method was detailed in Setchfield, 2007. The four main areas this paper will consider are: use of an external database to hold previous waves responses; filter block; concertinaing of remaining household members; and the Keeping-In-Touch-Exercise (KITE) survey. Current practices for these methods are summarised below.

2.1 Use of external database

The method of having the previous wave's survey data in an external file was chosen over rotating directly into the fields. This allowed several other options to be considered. Firstly, to allow consideration of having an option to choose people from different addresses, that is, to allow for merged households. Secondly, rotation could be processed within the datamodel. And lastly, in most cases the information was fed forward into question text or used as checks rather than pre-filling answers.

Subsequent analysis suggested merged households would be rare, so the option of allowing individuals to be picked from different addresses was dropped.

The external file holds information at both the household and person level.

A Keeping-in-Touch Exercise (KITE) survey is run about 3-4 months prior to the face-to-face survey. See section 2.4 for more information about the KITE survey. To alleviate the need to merge the KITE database with the previous wave's database, the previous wave's database is rotated into KITE. Thus the resulting KITE database is turned into the external file for the next wave. The data is rotated into a rotate block, which also executes some rules that the survey has on person level rotation. Then questions that use rotated data refer to this block.

2.2 The Filter Block and Splitting Households in the Field

A filter block was developed to ascertain the eligibility of all household members who were present in the previous wave, prior to the interview taking place. For those members who are no longer present, extra information such as date they moved, and contact details if they were still eligible (that is, moved within UK to another private household) or date of death are collected.

A summary screen then alerts the interviewer to whom will be interviewed in the current household, as well as alerting them to the extra households they need to create for the movers/ineligibles. The filter block is repeated on the newly created households. Currently the interviewer has to fill in the filter block on the additional households. We are looking to see if it can be automatically pre-filled correctly. That is, if the newly created household is eligible for interview, then the residents in the newly created household will have a status of resident – whereas in the original household they will have the status of mover. An interviewer cannot transmit back a case if they haven't created the required number of households.

The method of opening up extra households for those no longer resident/eligible, was used to process people from the last interview into similar categories. The categories were those who have moved locally, which the interviewer would go and interview, those who moved within UK, which would get reallocated out to an interviewer in that area, and those no longer eligible, such as those who moved abroad, moved to an institution or died.

2.3 Concertinaing the Household for Interview

Once those no longer resident in the household are removed, all remaining members are put together for the interview. Within the rotate block, there is code that matches the interviewee position number with their position number in the Filter grid. This ensures that the correct information is rotated forward for the interviewee. It also assists the interviewer visually. By grouping together those who are to be interviewed, no spaces are left in the answer pane. This is particularly useful for large share houses where there can be a lot of movement in and out of the household.

2.4 KITE

About three months prior to the main interview, and as part of the Keeping in Touch Exercise (KITE), a telephone survey is conducted to check on respondents. The main purpose is to check for movers, so that interviewers can go to the correct location at the main survey. Due to the current telephone unit systems, households are not split during the interview if household members move or become ineligible at the KITE survey. Instead, movers are flagged and put into new households when the sample is created for the next main survey. Currently ineligibles are ignored at KITE and remain in their current household. In the case of HAS, the KITE survey also gets information about joiners, as well as some basic demographic information. In the case of LOS, the KITE survey will also check for onset disability in households where members don't have any disabilities to check whether an interview is required at the next wave.

In terms of tracking respondents, the KITE survey is an integral part of the process. However, current processes mean that KITE facilitates sample loss. As new households are created for movers, this means they get a new serial number. However, this is done manually in the sample system and not in the Blaise questionnaire or the Blaise database. To implement the new ID calculations, we need to address the mover process at KITE. In addition we are restricted in developing new systems unless they are within the Blaise environment.

3. Gaps

As part of the development of LOS, current methods and processes for longitudinal social surveys were informally reviewed. LOS were looking at both GLF and HAS to see what processes they needed to set up or develop. As part of this, several issues came to light.

For example, in order to prepare data for EU-SILC, several manual processes were introduced on the GLF. For instance, recreating the original household to show movers, matching households and people between waves if they were movers, and removing households, which were created due to their ineligibility at the filter screen are mostly done manually. Not only are these manual processes very time consuming, but they also allow errors to creep in.

The main issues, together with suggested solutions are described in the following sections.

3.1 Tracking for Weighting

Every year on GLF, a yearly file is prepared and sent to Methodology so that they can produce weights. In order to do this work, there are several pieces of information that they require.

Firstly, information on the status at the current wave of everyone who was in the household in the previous wave is required. This enables attrition and movements to be looked at.

Secondly, the members who have moved should be recorded both in the household where they were resident at the previous wave and the household where they are now resident. That is, they should appear in two households.

Thirdly, all persons from the previous wave must be recorded whether they are contacted/respond or not. This means that unallocated quotas, deaths, and moves to institutions/abroad need to be included in the file.

Lastly, all persons who have been lost (i.e. we don't know whether they have moved to a private household in UK) or become ineligible (i.e. moved abroad, died, moved to institution) should remain in the same household, but be coded at a person level as lost or ineligible. That is, a new household should not be created for these persons, as this suggests they have moved to a private household in the UK then later became ineligible.

To implement these requirements we need to make some changes to our processes. Firstly we need to modify whom we create new households for. We only need to do this for movers to eligible households in the UK. That is, for those that die, move to an institution or move, but location unknown, we do not create a new household but leave them where they are.

For the actual interview, we have two options for leaving in these ineligible household members. We could follow our Labour Force Survey (LFS) method that leaves these people in the same position in the household grid, but just blanks them out for the purposes of interview. Instead of a name, they put a description in, like "Moved to Institution". This can get unwieldy, particularly in group households that can change dramatically year to year. It should be noted that for LFS, it is repeated every quarter for 5 quarters so generally doesn't experience as much change to household membership like the yearly (and in the case of HAS, two-yearly) surveys do. The other, and preferred option at this stage is to keep the interview grid as it is currently – concertinaing the household members such that only eligible members appear for the interview. Some of the problems this creates can be overcome by better use of the IDs, as discussed in the next section.

Lastly, we have been looking at creating a tracking block that contains information on the membership of the household in all waves. We are currently concentrating on requirements for GLF, which has 4 waves. The block will be an array of 4, each representing a wave. The information captured is likely to be HHNO (see section 3.2 for details), Household ID, Split Number (see section 3.2 for more details), then an array of household members containing Name, Age, Sex, Status at that wave, date of death/move if appropriate, Person No, Person ID. Ideally if they have moved to a new household, which is still eligible, it would be great to have the household ID of their new household.

3.2 ID's

A big problem area is identification of members and households. Currently we use a serial number made up of an area code, an address number and a household number as the identifier for households. For wave 1 the household number is always '01'. The area code is created during the sample creation and relates to a physical location within the UK. The address number is a sequential number from 1 to the maximum number of cases an interviewer will carry out in a particular month. Due to the large sample size of HAS, a particular area may be in the sample twice during a 2 year period. Hence within a wave, serial number is not always unique. However as it is unique within a month, we include the last two digits of year (e.g. '09' for 2009) and numeric month (e.g. '06' for June).

For example, if Area code = 05000, Address = 10, and household = 01, and the survey takes place in June 2009, the serial number will be 0906050001001.

In each subsequent main wave survey, new household can be created for movers in the field by the interviewer. When a new household is created, the household number is incremented by 1, and thus the serial number changes.

For example, if there is a mover from the household represented by Area code = 05000, Address = 10 and household = 01 in June 2009, then a new household is created which will have household = 02. The serial number becomes 0906050001002.

When the sample is created for the next interview – either KITE or main survey, the serial numbers are looked at. Due to the sample and allocation system, household number must equal 1 prior to the case being allocated and scattered to the interview. If a person moves to a different area, then they will receive a new area and address numbers, with household number set back to 1. If a person moves to a house within the same area, the area code stays the same but they get a new address number and household number is set back to 1.

Rather than keeping the serial number from wave1, it is recreated at each interview (both KITE and main interviews). The sample file contains current serial number and "lastserial", which we can use in matching. "Lastserial" refers to the serial number at the last interview. As we use serial number as a proxy for household ID, the recreation of serial number each time makes it is difficult to match between waves, particularly for movers. Using Manipula, we currently match from a wave back to the KITE then back to the wave and so on. Then manually, any households that were dropped from the sample, either because they were refusals, or ineligibles or movers with unknown contact details are added back in. Once this is done, we can use the resulting file as look-up across all waves so that wave 4 can be matched to responses in wave 1.

With the introduction of LOS, we have taken the opportunity to develop a different method, which will be used on all longitudinal surveys. Rather than using serial as a proxy for household ID, and recreating it each wave, we will create a new variable called HHNO in wave 1 that will remain with the household (and those that split off) through the life that they are in the survey. In addition we will implement the 'split number' concept as per EU-SILC guidelines to identify those who have split from the original household. The household ID will then be made up of HHNO and the split number. The split number for the original sampled household is '00'. Rather than using a random number start to create HHNO, we will still utilise the last two digits of year, the numeric month, the area code and address number in wave 1 HHNO. HHNO will stay with all original sample households whilst they are in the survey regardless of whether they move or not.

Example 1: household at Wave 1, Surveyed June 2007

Area = 05000			
Address = 10			
Household = 01			
Name	**Date of Birth**	**Sex**	**Status**
Bob	10/03/1963	Male	Resident
Mary	24/09/1965	Female	Resident
Tom	15/10/1987	Male	Resident

So in example 1 HHNO = 07060500010, and the split number = 00, as it is wave 1. Hence household ID = 0706050001000.

The second issue is the split number. This is proving more difficult to do automatically in the field. The main complication is the KITE survey and the fact we manually split out movers at this point. Initially in Wave 1 the split number is 00. Theoretically, every time a split occurs, the split number increases. As KITE is not done in the casebook environment, new households are not created at point of interview when movers are identified. Rather they are flagged and during the sample creation process new households are created for the movers. It is at this point that the split number will need to be updated for new households. Due to the current sample creation process, this will need to be done manually. For movers at the main wave surveys, it initially looked like we could create the split number automatically at point of interview when new households are created in the field. However, we quickly realised that if the original household had another mover and the already split out household had a mover, they would end up with the same split number. So again, this means households with movers will need to be identified and split numbers calculated once the cases have been returned to HQ. In the short term, the calculation of split numbers is likely to be done manually, although it is preferred that an automatic solution will be developed.

Example 2: household at Wave 2, Surveyed June 2008. Following on from example 1, Tom moves out. The household now looks like:

Area = 05000			
Address = 10			
Household = 01			
Name	**Date of Birth**	**Sex**	**Status**
Bob	10/03/1963	Male	Resident
Mary	24/09/1965	Female	Resident
Tom	15/10/1987	Male	Moved out

HHNO in the first household remains the same: 07060500010, and the split number is still 00. Household ID = 0706050001000

A new household is created for Tom.

Area = 05000			
Address = 10			
Household = 02			
Name	**Date of Birth**	**Sex**	**Status**
Tom	15/10/1987	Male	Resident

HHNO in the second household is the same as the first household: 07060500010. However the split number changes to 01. Household ID = 0706050001001.

Lastly there is the issue of person number and person ID. Person ID was made up of household ID and person number. Again by default, we have been using position number in the interview grid to act as a person number. This is unique in a wave, but can be used by several different people across waves. Surveys are mainly interested in Person ID rather than person number so this hasn't been much a problem. In addition due to fact serial number was recreated at each wave, personal IDs for those joining a household always had a different year and month that those who were already in the household, so personal IDs were never duplicated even though person number was. However, changing the way household ID is created and re-looking at EU-SILC rules, we realised that we were not calculating person number correctly. Person number should not be reused within a household. Also, movers take with them, their person number. Joiners to households get the highest, unused person number. For more information see the EU-SILC paper referenced at the end of this paper. Along with changing household ID, we will implement the creation of person number and person ID at the same time. Person ID will still be made up of household ID and person number. Person number, in wave 1, will be the position number in the interview. Then in subsequent waves, person number for joiners will be the highest unused person number.

Example 3: Using the household in example 1. Person number is equal to position number in the first wave.

HHNO = 07060500010				
Household ID = 0706050001000				
Name	**Date of Birth**	**Sex**	**Status**	**Person No**
Bob	10/03/1963	Male	Resident	1
Mary	24/09/1965	Female	Resident	2
Tom	15/10/1987	Male	Resident	3

So Person ID of Bob will be 070605000100001.

In example 2, Tom had split out into a new household. Let's say he is joined by Jane.

Tom keeps his person number from the first time he was interviewed, so it is 3. Jane takes the highest unused person number in the household, which is 1 in this case.

HHNO = 07060500010				
Household ID = 0706050001001				
Name	**Date of Birth**	**Sex**	**Status**	**Person No.**
Tom	15/10/1987	Male	Resident	3
Jane	05/04/1988	Female	Resident	1

Tom's person number was assigned to him in Wave 1. So his person ID is 07060500100003. Jane's person ID is 070605000100101.

3.3 Rotation – Next Evolution

As mentioned in Setchfield, 2007 and summarised in section 2, the current rotation method uses an external file, which contains all the cases from the same reference month in the previous wave. This method is used for both HAS and GLF. Due to data confidentiality concerns as well as size concerns, we have begun to look into other methods.

As part of the Integrated Household Survey project, Manipula was used to rotate cases and data into those cases in HQ prior to the interview. This meant that the cases would go out with their data only. The method was developed on the Labour Force survey, which has different rules for rotating data. However, we are confident that we can use the same approach on GLF, HAS and LOS.

The main difference is that the process for which households to rotate occurs in the sample creation process. That means we only need to write a Manipula script for which variables to rotate data for. The data will be rotated into a holding block. This means we can still utilise the same filter and rotate blocks as we currently do. The only difference is that instead of referencing an external file, the holding block will be referenced. It is hoped that this new method will be implemented on HAS within a few months, then GLF and will be implemented for wave 2 of LOS which occurs early 2010.

Initially, it will be implemented in the main face-to-face survey case creation process. Unfortunately, at this stage the case creation process between the face-to-face survey and the telephone unit survey and also between surveys are not necessarily the same, so we have had to prioritise.

4. Next Steps

In addition to the solutions mentioned above, consideration is also being given to creating a tracking database. From this database, it is envisioned sample files can be created for KITE surveys and main wave surveys, data files can be created which go to methodology for the creation of weights, and updates can be made to household information, such as new addresses which are notified to ONS via change of address cards or other keeping in touch exercises can be applied. All sampled households will be loaded to the database. This is still in the initial discussion phase and it is not clear how urgent the need will be after the solutions discussed above are implemented.

The issue of updating household information as it is notified to HQ outside of surveys – such as receipt of change of address cards, individual refusals via the public enquiry line, notifications of deaths – is one outstanding issue that needs to be addressed. These notifications are generally small in number. A short-term solution is required, and the current idea is to implement a block that only comes online when the case is opened in HQ. To do this, we would utilise the edit version of the questionnaire. This block would be automatically pre-filled with the household information as at the end of the last interview, but would allow HQ staff to update fields such as status or contact information prior to the next survey. It is hoped a trial version of the block will be made available to HAS in the next few months to test.

5. References

Setchfield, C. (2007): Coping with People who just won't stay put: The Use of Blaise in Longitudinal Panel Surveys. Presented at 11th International Blaise Users Conference, 2007, Annapolis, September 2007

Eurostat (2008): Description of Target Variables: Cross-sectional and Longitudinal.