Understanding Society

Working Paper Series

No. 2011 – 04

June 2011

Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey

Peter Lynn

Institute for Social and Economic Research, University of Essex

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey

Peter Lynn

Abstract

This paper describes the problem of maintaining cross-sectional representativeness in a longitudinal survey of a changing population. The extent and nature of the problem is outlined and potential solutions are described. The procedures adopted on *Understanding Society* are described. The main challenges are to correctly identify members of the initial sample who leave the population – through death or emigration – and to periodically add appropriate samples of people who join the population – through birth or immigration. The samples to be added should have known selection probabilities and should strike an appropriate balance between precision of estimation and cost-efficiency of fieldwork.

Key words: births, deaths, emigration, flow sampling, immigration, sample design, weighting

JEL classifications: C81, C83

Author contact details: plynn@essex.ac.uk

Acknowledgements: The work reported in this paper is part of the *Understanding Society* programme, funded by the UK Economic and Social Research Council (ESRC). The author is grateful to Nicole Watson for stimulating discussions and for analysis of the extent of identification of deaths in the BHPS sample.

Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey

Peter Lynn

Non-Technical Summary

Understanding Society is designed to follow a representative sample of persons over a period of years, collecting data from them at annual intervals. The sample is designed to represent the entire resident household population of the United Kingdom at the time of wave 1 of the survey. The data will permit study of the changing circumstances of the population.

However, over the lifetime of the study the population of the United Kingdom will also itself change. People will leave the population, through death or emigration, and people will join the population, through birth or immigration. It is desirable that the *Understanding Society* sample should remain representative of the United Kingdom population. To achieve that, suitable procedures are needed to deal with the various types of changes in the population.

This paper first describes the various types of population change and provides some information on the extent of those changes in the UK. It then sets out procedures that are being used – or could be used – to keep the *Understanding Society* sample representative of the current UK population.

Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey

Peter Lynn

1. Introduction: Sampling for Longitudinal Surveys

Longitudinal surveys are surveys which collect data at several points in time from the same set of study units (Lynn 2009). A key issue in sample design for longitudinal surveys is that the population of interest must be defined in time as well as in space and in terms of other characteristics (typically age, residency status, etc). A range of different sample designs are used for longitudinal surveys (Elliot et al 2009) and in this paper we explore the issue of how sample design can maintain cross-sectional representativeness over time. To understand why this is important, we need to first consider the basic objectives of longitudinal surveys. The primary aim is typically to be able to study micro-level change of one kind or another. However, although the change takes place, and is measured, at the level of the individual unit, the researcher desires to make inferences about the extent and nature of change at the population level. This begs the question of exactly what population the sample represents.

For example, the following research question could be posed: Amongst working-age adults in the UK, are transitions from employment to unemployment over a period of time more common in one region than another? To answer the question it is necessary first to assemble a sample of working age adults in the UK and to measure activity status transitions of those persons over a period of time. Let's assume that we do that by carrying out a longitudinal survey. Then, for each sample member we can construct a measure of how many relevant transitions they experienced over a particular period of time. This is our micro-level measure of change. We can then calculate the mean number of transitions amongst sample members in each region and we can calculate differences between regions in these means. These are our sample-level statistics, which we use as estimates of population-level parameters. But what population does our sample represent? The answer, it turns out, depends on how the sample is selected and on how we use the data to construct our estimates.

Suppose we initially select a representative sample of working age adults in the UK and then interview them regularly over a period of five years. Over that period of time, some sample members will reach retirement age, some may die and some may emigrate. As a result, the survey will obtain data relating to the full 5-year period for some sample members, but for other sample members the data will relate to only a shorter period. So, as we want estimates of mean numbers of transitions in a 5-year period, one option is to use only the data for sample members who were observed for the full 5-year period. If we do this, then our sample is representative only of *people who were of working age and in the UK throughout the 5-year period*. The trouble with this approach is that this gives only a partial picture of all the transitions taking place in the UK labour market. Some of those sample members who reached retirement age, died or emigrated during the 5-year period may also have made some transitions from employment to unemployment at some point during the period. We could of course include those people in our estimates, but then our sample is representative of a different population, namely all *people who were of working age and in the UK at the start of the 5-year period*.

And similarly, some transitions may have been made by people who stood no chance of being selected in the initial sample, because they only reached working age or entered the UK during the 5-year period. Maybe we would ideally like our estimates to be representative of all *transitions from employment to unemployment made in the UK labour market during the 5-year period*. In that case, we additionally need a sample of people who reached working age or entered the UK during the 5-year period.

Effectively, if we want to study the totality of a particular phenomenon (such as transitions from employment to unemployment) taking place in society over a particular period of time, then we need a sample of observations that represent all the people who could have experienced that phenomenon at any time during the period. In other words, our population of interest is all people who were ever a member of the cross-sectional population at any point during the survey period. Such people cannot all be identified at the outset of the survey, because at that point of time we do not know who will enter the population during the survey period, so we need to find a way of sampling flows into the population and adding them to the initial sample of the population stock.

2. The *Understanding Society* Sample

The main component of the *Understanding Society* sample is a general population sample, representative of all people resident in households in the United Kingdom at the time of the wave 1 fieldwork (January 2009 to December 2010). Details of the design and selection of this sample can be found in Lynn (2009b). Attempts will be made to re-interview members of this sample at annual intervals. Our focus in this paper will be on this sample. The *Understanding Society* sample also includes other components, to which the same principles could be applied.

Eligibility for inclusion in the general population sample is defined by being a member of a resident household in the UK at the time of wave 1 fieldwork, regardless of age, nationality or citizenship. People living in institutions such as nursing homes, prisons and military bases are excluded. We therefore consider ways of maintaining the sample representative over time of the cross-sectional population defined in this way. There are three ways in which people can enter or leave the population defined in this way:

- Birth or death;
- Immigration or emigration from the UK;
- Moving between an institution and a household.

In the next sections of this paper we consider the implications of each in turn.

3. Potential Under-Coverage

Under-coverage is potentially caused by the groups of people who enter the population subsequent to selection of the initial sample at wave 1. In this section we consider each of those three groups in turn. Common to all three, however, is that the excluded group differs from the included population only in the recency of the event that defines their group membership. So, while immigrants subsequent to wave 1 are excluded from the initial sample, immigrants in the period immediately before wave 1 – and indeed all prior periods – are included. Similarly, people who left an institution in the period prior to wave 1 will be represented in the initial sample, as will those born recently prior to wave 1.

3.1 Births

There were approximately 790,000 births in the UK in 2009[1]. In 2009, people born in the previous twelve months accounted for 1.27% of the resident population (Office for National Statistics, 2010b). Thus, at wave 2 of *Understanding Society* the initial sample (the sample selected at wave 1) would suffer from undercoverage of 1.27% of the population. This undercoverage would of course increase over time in a roughly linear fashion. For example, by wave 6 the sample would not include anyone aged under 5, a group which, in 2010, accounted for around 6.11% of the population.

The level of undercoverage would not itself be a problem if it occurred at random. But of course, the youngest members of the population are by no means a random subset of the total. Any estimates of characteristics of the full population (of all ages), or of any sub-population that includes the youngest members of the population, would be biased. For example, a researcher using data from the first 19 waves to study transitions from education to the labour market would find that persons aged under 18 at wave 19 were not represented in the sample. The observed transitions would therefore be restricted to older cohorts and potentially not representative of all transitions.

3.2 Immigration

Immigrants to the UK subsequent to wave 1 will of course not be represented in the initial sample. It is not straightforward to estimate the proportion of the population represented by such immigrants as this is a complex function of the immigration rate, the re-emigration rate and the interval between survey waves. For example, at wave 2 of *Understanding Society* the undercoverage would relate to all persons who had entered the UK since wave 1 and were still resident at the time of wave 2. As the interval between waves is only twelve months, we can imagine that a sizeable proportion of immigrants entering the country over the course of twelve months will still be resident at the end of that period, but we do not know the exact size of that proportion. Note also that this proportion will be higher than the proportion of immigrants who remain in the country twelve months after immigration, as the undercoverage includes anyone who entered the country at any time

[1] 706,250 in England and Wales (Office for National Statistics 2010a), 59,050 in Scotland (General Register Office for Scotland, 2010) and 24,900 in Northern Ireland (Northern Ireland Statistics and Research Agency, 2010)

between one day and twelve months before wave 2. In 2009, approximately 701,200 immigrants to the UK were granted entry visas (Home Office 2010[2]). The total number of immigrants may be greater than this as entry visas are not required by citizens of the European Economic Area (EEA) or Swiss nationals. The Office for National Statistics estimate that 567,000 long-term migrants settled in the UK during 2009 (Office for National Statistics, 2010c). A long-term migrant is defined as someone who changes his or her country of usual residence for a period of at least a year. This is perhaps our best estimate of the extent of under-coverage at wave 2. 567,000 corresponds to approximately 0.92% of the resident population.

The extent of undercoverage after *n* waves (*n* > 1) would be difficult to estimate, even if we had good knowledge of the extent of undercoverage after one wave. This is because the rate of re-emigration is very unlikely to be constant over time. Re-emigration rates are likely to be greatest in the first year or two and to reduce thereafter, though the exact extent of re-emigration is unknown. The Office for National Statistics produce estimates of the number of emigrants each year, but the length of time that each emigrant had spent in the UK is unknown. We do know, however, that around 200,000 non-British citizens emigrate from the UK per annum (Office for National Statistics 2011), accounting for almost two-thirds of all emigrants from the UK (own calculation). These are only the ones who stayed long enough to establish permanent residency. The total number of non-British citizens leaving the UK after living here for a period will be higher than this.

As with births, the problematic feature of excluding immigrants from the sample is not the extent of under-coverage, but rather the fact that the excluded subgroup is non-random. Immigrants are likely to have rather different characteristics and experiences to other members of the population in terms of many of the social dimensions of interest to *Understanding Society* – e.g. housing, education, fertility, health, employment and so on (e.g. Hickman et al 2008; Robinson et al 2007). Consequently, the undercoverage will cause bias in the estimation of population parameters.

[2] The figure includes only those granted employment-related visas, study visas, leave to join family members in the UK, or granted asylum. All visitor visas are excluded.

3.3 Leaving Institutions

Persons resident in institutions will be excluded from the initial *Understanding Society* sample. Some of these persons will subsequently move (back) to a resident household and will therefore (re-)enter the study population. These will include people leaving military establishments (and, in some cases, their families), prison, residential care, hostels and residential educational establishments (though attempts are made to include in the sample students who have a permanent address in the UK to which they return out of term-time).

Though approximately 1.8% (Office for National Statistics, 2005) of the UK population reside in institutions, the proportion of these who re-enter a residential household in any given year is likely to be very low, considering that a majority of the institutionalised population are in long-term nursing and care homes.

The ex-institutionalised population is likely to be rather more heterogeneous than either new births or immigrants, so the nature of any under-coverage bias is hard to predict. The excluded groups include some with more education than average (students) and some with less education than average (prisoners, military personnel); some who are predominantly young (students) and some who are predominantly older (care home residents).

4. Potential Over-Coverage

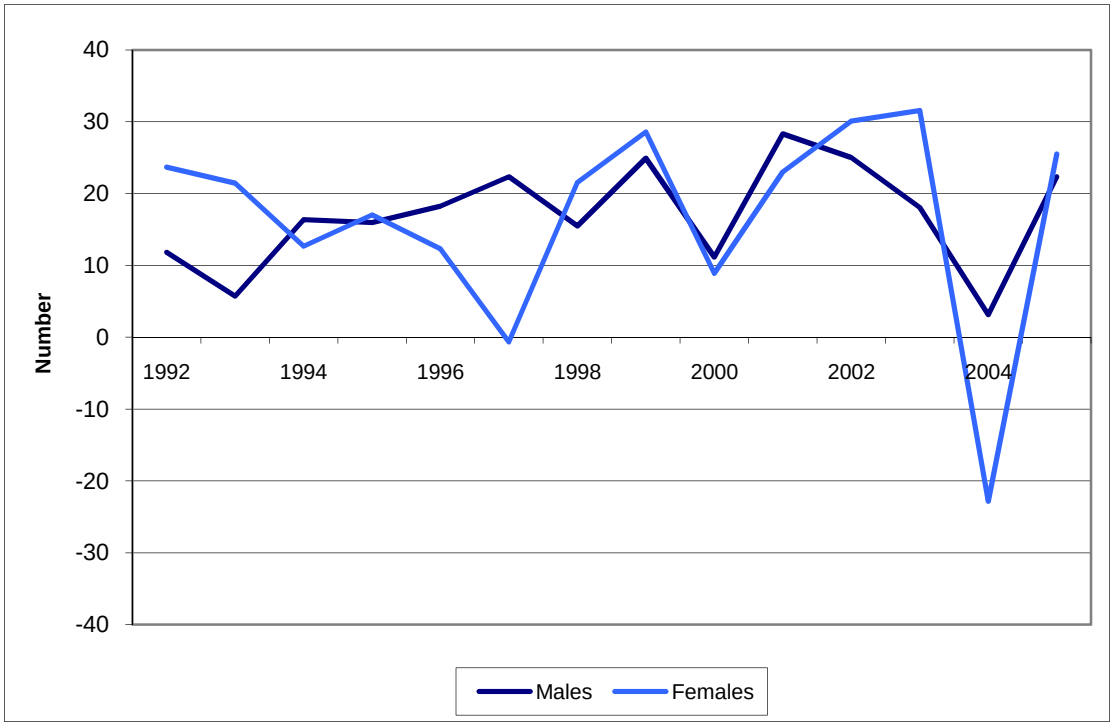
Over-coverage is potentially caused by including in the sample people who have in fact left the population subsequent to wave 1. In the case of *Understanding Society* it is unlikely that such people would be interviewed in error. This obviously cannot happen in the case of death and is very unlikely to happen in the case of either emigration or institutionalisation, for so long as field work is carried out face-to-face. (We would note that with telephone or web interviewing there would be a greater risk of failing to identify such moves out of the population as movers may still be contactable at the same email address or same telephone number. It would be necessary to ask an explicit question to check that the sample member is still resident in a household in the UK.) The problem of over-coverage can therefore reasonably be assumed to relate only to sample members who are not interviewed at a particular wave, i.e. non-respondents. This is a problem if the survey weighting procedures are based on an assumption that such sample members still belong

to the population of interest. This could distort the (weighted) sample if such people – those who are assumed to be members of the study population when in fact they have left the population – are not randomly distributed across categories of the variables used in the weighting. For example, if these people include a large number in the oldest age group, the remaining respondents in this group may be assigned a weight that is too large – because we mistakenly think that there are many non-respondents in this group. This would result in over-representation of the age group and consequent bias in estimates of any parameters that are associated with age. For this reason it is important to avoid over-coverage amongst survey non-respondents.

4.1 Deaths

There are approximately 500,000 deaths in England and Wales each year, equivalent to around 0.9% of the population (Office for National Statistics, 2010d). The deaths are of course predominantly in the oldest age groups. For example, in 2009 4.7% of the population were aged 80 or over while 53.2% of deaths were of people aged 80 or over. Overcoverage related to death is therefore likely to result in bias towards older people and towards characteristics associated with older people. Overcoverage will arise if *Understanding Society* sample members who die subsequent to wave 1 are not identified as having died and are therefore treated as still being eligible members of the study population. This is most likely to occur when contact is lost with the sample member's household, something which often happens when the deceased sample member lived alone and is very likely to happen if they have lived alone since wave 1 (so there are no other household members who are sample members), particularly if they did not provide any contact details of people who would know their circumstances or if the contact details are no longer correct. Analysis of the BHPS sample assumed to be eligible - i.e. after removing known deaths but retaining those with whom contact was lost – indicates a clear net under-identification of deaths (see figure 1).

Figure 1: Difference between expected and observed number of deaths in BHPS original England and Wales sample (expected-observed), 1992 to 2005



Source: Analysis carried out by Nicole Watson while visiting researcher at ISER

4.2 Emigration

Around 400,000 people emigrate from the UK each year (400,000 in 2006, 340,000 in 2007, 430,000 in 2008, and 370,000 in 2009 – Office for National Statistics, 2010c). This number is equivalent to approximately 0.8% of the UK population and may serve as a maximum estimate of the proportion of the *Understanding Society* wave 1 sample that would have emigrated by the time of wave 2, one year later. It is a maximum because some of those emigrants may be people who had only entered the country within the previous twelve months and were therefore not part of the population sampled to form the initial sample. As noted above in section 3.2, the re-emigration rate over a particular period of time is hard to estimate, but we should note that the number of emigrants over the course of a year who had only immigrated to the UK within that year is not equal to the number of immigrants who emigrate within a year of arrival. In fact, immigrants over the course of a year will only have to have stayed in the country for an average of six months in order to form part of the wave 2 cross-sectional population.

Over time, the proportion of the initial population who have emigrated from the UK will increase, but not linearly.

The likelihood of failing to identify that an *Understanding Society* sample member has emigrated is greater in the case of a whole-household emigration than when other sample members from the same household remain in the UK, simply because the latter case provides a greater chance of contact being made with someone who knows that the sample member has emigrated. Any bias caused by over-coverage of emigrants is therefore likely to result in over-representation of groups with a relatively high propensity to emigrate and, particularly, of the types of people who emigrate as part of a whole-household emigration.

4.3 Institutionalisation

The number of people who move from a residential household to an institution in the course of twelve months and are still in an institution at the end of the twelve-month period is unknown. For some types of institutions, e.g. prisons, statistics exist on the number of people entering in a year. However, it is not known either what proportion remain in prison at the end of the period, nor what proportion were already in an institution at the start of the period. For other types of institutions, e.g. nursing homes, it may be reasonable to assume that very few entrants would have returned to a resident household by the end of the period. The majority will either remain in an institution at the end of the period or will have died. However, there is little or no information about the number of entrants from households in the course of a year. There exists partial information about number of entrants, but this does not indicate the type of accommodation from which people have moved, nor does it enable double-counting to be avoided (as some people may have entered more than one institution during the course of a year).

As with the other categories of potential over-coverage, a failure to identify a move into an institution is most likely to occur when no contact is made with the previous household of the sample member, which in turn is most likely to occur when the sample member was previously a single-person household (e.g. a single elderly person who moves into a care home, or a single person who enters prison or moves to a military base).

Understanding Society attempts to follow and interview sample members who move into institutions, so implicitly the cross-sectional population of interest to the survey at each wave n ($n > 1$) includes all persons who have moved from the household population to an institution since the time of wave 1.

5. Sampling Births

Understanding Society takes advantage of the natural process of household evolution to add new births to the sample at each wave. Whenever a female sample member has given birth since the previous wave, that child, or those children, become sample members. This procedure gives all new births an equal probability of selection, given that each child has one mother. Furthermore, the probability of selection is equal to that of the mother, ensuring that the overall sampling fraction of new births is the same as that of mothers. The procedure additionally aids the study of household dynamics, given that all sample births are in existing sample households. This method of sampling new births is similar to that used by most other household panel surveys, though some include births when either the mother or father is an existing sample member, as was done on *Understanding Society*'s predecessor the British Household Panel Survey (Lynn 2006).

This sampling method provides an efficient probability sample of children born to mothers who were themselves a member of the cross-sectional population at the time of wave 1. However, it does not give any chance of inclusion to births in the UK since wave 1 to mothers who were themselves not resident in the UK at the time of wave 1 (i.e. recent immigrants) or who were resident in an institution at that time. To include such births, a separate sampling method would be needed. We propose that an effective way to do this would be to select samples of recent immigrants and recent movers from an institution to a household and then simultaneously to include any children born to female members of those samples since wave 1. Sections 6 and 7 below, respectively, propose how this might be done.

6. Sampling Immigrants

Understanding Society has considered the possibility of adding samples of recent immigrants to the survey but has not yet adopted a procedure for doing so. Adding samples of recent immigrants would be costly and these costs would have to be justified in terms of the added scientific value to the study. In this section we propose a relatively cost-efficient procedure that could be used for identifying appropriate samples of recent immigrants.

The general population sample component of *Understanding Society* has a clustered design and is based upon an initial sample of 18 addresses in each of 2,880 postal sectors. Each cluster of 18 addresses is allocated to a specific month in the 24-month field period. This provides for efficient fieldwork loads, as each point is generally worked by one interviewer, who therefore only has to travel to the sample area once on each day that she is working on the survey. Travel between sample addresses during the working day involves only short distances. However, after wave 1 some sample members will move home. Sometimes one or more sample members will leave the initially-sampled addresses while one or more will remain at that address (for example when adult children leave the parental home or when one partner in a couple leaves). In other instances, all sample members will leave the initial address (for example when a whole household moves home). These whole-household moves are central to our proposal for sampling immigrants.

The initial *Understanding Society* sample of addresses is a representative sample of all addresses in the UK. Consequently, the residents of those addresses are a representative sample of all residents of the UK. The same is true of any sub-population, so recent immigrants at those addresses are a representative sample of all recent immigrants. Our proposal, then, is to identify all recent immigrants – i.e. people who have become resident in the UK since wave 1 – at this sample of addresses. This is something that could be attempted once every n waves ($n > 1$) and would be relatively efficient. It would require screening of every initial sample address. But a majority of these addresses would already require a visit from an interviewer in order to interview existing sample members (those who have not moved since wave 1). Indeed, the *Understanding Society* sample procedures already involve identifying all current members of the household. A recent immigrant who had joined the household would become a “temporary sample member” under the current procedures and would be interviewed for so long as he or she remained in the same household as at least one original sample member. Under our proposal, such people would instead become a full sample member.

Additionally, screening would need to be carried out at the remaining initial sample addresses – those where the whole household has moved away and those where the household at the time of wave 1 became non-respondents (as some of those may have moved away subsequently). This screening would be relatively efficient as:

- it would take place in the same set of postal sectors that interviewers already have to visit to carry out the main fieldwork;

- it could take place at the same time as the main fieldwork;
- it would involve a relatively small number of addresses in each postal sector – we estimate an average of 7 or 8 addresses per sector if the screening were to first take place at wave 6;
- it would involve a relatively simple screening question, to identify whether there is anyone resident at the address who was not resident in the UK at the time of wave 1 (a specific month could be specified, e.g. January 2009 if the address is part of the January year 1 sample).

Of course, this approach assumes that the population of residential addresses remains unchanged over time. In practices, some new properties will have been built or converted to residential use since the time of wave 1. Recent immigrants at such addresses would not be included by this screening procedure, resulting in some under-coverage. This could be remedied by adding to the screening sample a set of such new addresses. This could be achieved by comparing, for each sample postal sector, the addresses listed on the Postcode Address File (the sampling frame from which the initial sample of addresses was selected) at the time the initial sample was selected with the addresses listed currently. For example, if a sample of recent immigrants were to be added at wave 6, this might involve comparing the versions of the Postcode Address File current in November 2008 and November 2013. Any addresses present in the latter but not the former could then be sampled using the same sampling fraction as for the original sample. These addresses would then be screened in exactly the same way as the initial sample addresses that no longer contain a resident sample member.

7. Sampling People Leaving Institutions

To identify a sample of people in resident households who were living in an institution at the time of wave 1, a method similar to that proposed above for immigrants could, in principle, be used. There may be greater sensitivities around the screening questions, however, as it may be more challenging to explain why the questions were being asked. Screening for this particular subgroup could be combined with screening for recent immigrants into a single exercise. Inclusion of this group in the screening exercise would not affect the sample of addresses to be screened, only the nature of the screening questions, and would result in a larger screened-in sample.

8. Identifying Deaths, Emigrants and People Entering Institutions

It is an on-going task at every wave of *Understanding Society* to attempt to identify sample members who have died, moved abroad or moved into an institution since wave 1. However, the action taken in the event of identifying a sample member in that situation differs in the three cases:

- Sample members identified as having died are treated as no longer being eligible for inclusion in the survey and no further attempts are made to contact them or other members of their family or household (unless, of course, those people are sample members in their own right). (This is unlike the practice on the English Longitudinal Study of Ageing, where an “exit interview” is attempted with a relative or carer of the deceased sample member - Scholes et al, 2008).

- Sample members identified as having moved abroad are noted as such but, should it later be discovered that they have returned to the UK, they will be treated as continuing sample members and attempts will be made to interview them and retain them in the sample thereafter.

- Sample members identified as having moved into an institution are treated no differently to those who move to another residential address. Attempts continue to interview them at each wave and to retain them in the sample. Thus, over time and notwithstanding non-response, the sample will gradually become representative of the institutionalised population. It will become fully representative at the point where no members of the institutionalised population have been institutionalised for longer than the elapsed period of time since wave 1.

Information on whether any of these three types of transitions have occurred is obtained primarily during the course of the normal survey procedures to contact sample households. Often other household members are able to inform the interviewer of a death or a move abroad or into an institution. If there are no other household members, then information may be obtained from the new resident(s) of the last-known address of the sample member. Failing that, if no information about the whereabouts of the sample member is forthcoming from the field, the case will enter the “tracking” system, which begins with survey office staff attempting to contact the sample member by alternative means such as email or work telephone, or to contact people who had previously been

named by the sample member as “stable contacts” or related sample members in other households (this latter situation becomes more common over time as initial sample households split and generate multiple households). These contacts, who are often close relatives, may be able to inform the survey staff that the sample member has died, moved abroad or moved to an institution. The final resort of the tracking process is to use the services of a commercial search agency, Capscan (www.capscan.co.uk), who search a number of databases. If the tracking process fails to identify the whereabouts of a sample member, the outcome for that wave will be recorded in the data as “failure to track”. A proportion of such cases may in fact have died and therefore ceased to be sample members. As discussed in section 4.1, this is perhaps most likely to occur when the sample member had lived alone, and particularly if they had not provided any stable contact details. The system used on *Understanding Society* to identify deaths and moves away from the UK household population is therefore likely to be imperfect, though it probably identifies the majority of such transitions.

A procedure used on some surveys is to match sample members with death registers via the National Health Service Central Register. For example, this is done on the *Health Survey for England* and the *English Longitudinal Study of Ageing* (Hussey et al, 2010). This has the potential to identify all deaths, though this is subject to the success of the matching process, which is itself imperfect (Jabine & SCheuren, 1986). *Understanding Society* plans to link survey records to death registers for sample members who have given permission for the linkage of administrative health data.

9. Concluding Remarks

It is important for data users to be aware of the implications for sample representativeness of the procedures adopted on *Understanding Society*. Current procedures in principle deal appropriately with births, deaths, emigrants and moves into institutions. In practice, there are a couple of caveats:

- There is likely to be some under-identification of deaths, emigrants and moves into institutions, which may lead to over-representation in the weighted sample of people with similar characteristics;
- There is no reason to doubt the effectiveness of the procedures for sampling births to initial sample members, but it should be recognised that these procedures

exclude a small proportion of UK births, namely those to mothers who entered the UK household population since the time of wave 1.

Current procedures do not attempt to deal with immigrants or moves from institutions to households. People who enter the UK household population via either of these routes subsequent to wave 1 are not included in the sample.

The net result is that the *Understanding Society* sample will become, over time, increasingly representative of the UK institutionalised population (who are not represented at all in the initial sample) and increasingly unrepresentative of recent immigrants and recent movers from institutions to households (who are represented in the initial sample but not subsequently added to the sample). The sample will also become increasingly unrepresentative of recent births, though this effect will be much weaker and more gradual (as the majority of recent births are represented by new additions to the sample; only a small, but increasing, minority are excluded).

Careful evaluation of the merits of periodically adding samples of recent UK immigrants and recent movers from institutions to households may be worthwhile.

References

Berthoud, R., Fumagalli, L., Lynn, P. & Platt, L. (2009) Design of the *Understanding Society* Ethnic Minority Boost Sample. *Understanding Society Working Paper* 2009-02, Colchester: ISER, University of Essex. Available at http://research.understandingsociety.org.uk/publications/working-paper/2009-02.

General Register Office for Scotland (2010) Births time series data. Available at http://www.gro-scotland.gov.uk/statistics/theme/vital-events/births/time-series.html

Hickman, M., Crowley, H. & Mai, N. (2008) *Immigration and Social Cohesion in the UK*. York, Joseph Rowntree Foundation.

Home Office (2010) *Control Of Immigration: Quarterly Statistical Summary*, United Kingdom, July – September 2010. Available at http://rds.homeoffice.gov.uk/rds/immigration-asylum-stats.html

Hussey, D., Lessof, C., Ward, K. & Wood, N. (2010) Methodology, chapter 10 in *Financial Circumstances, Health and Well-being of the Older Population in England: The 2008 English Longitudinal Study of Ageing (wave 4)*, ed.s Banks, J., Lessof, C., Nazroo, J., Rogers, N., Stafford, M. & Steptoe, A. London, Institute for Fiscal Studies.

Jabine, T.B. & Scheuren, F.J. (1986) Record Linkages for statistical purposes: methodological issues, *Journal of Official Statistics* 2 255-277.

Lavallée, P. (2007), *Indirect Sampling*, Springer.

Lynn, P. (ed.) (2006) Quality Profile: British Household Panel Survey, version 2.0: waves 1 to 13: 1991 - 2003. Colchester: ISER, University of Essex. Available at http://www.iser.essex.ac.uk/survey/bhps/quality-profile

Lynn, P. (2009a) Methods for longitudinal surveys, chapter 1 of *Methodology of Longitudinal Surveys* (ed. P. Lynn). Chichester, Wiley.

Lynn, P. (2009b) Sample Design for *Understanding Society. Understanding Society Working Paper* 2009-02, Colchester: ISER, University of Essex. Available at http://research.understandingsociety.org.uk/publications/working-paper/2009-01.

Northern Ireland Statistics and Research Agency (2010) Births and deaths press release. Available at http://www.nisra.gov.uk/demography/default.asp23.htm

Office for National Statistics (2005) Focus on Families. London: Office for National Statistics. Available at http://www.statistics.gov.uk/downloads/theme_compendia/fof2005/families.pdf

Office for National Statistics (2010a) Live births: fertility rates fall. Available at http://www.statistics.gov.uk/cci/nugget.asp?id=369

Office for National Statistics (2010b) Population estimates for UK, England and Wales, Scotland and Northern Ireland - current datasets. Available at http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15106

Office for National Statistics (2010c) Long-Term International Migration (LTIM) tables: 1991 – latest. Available at http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=15053

Office for National Statistics (2010d) Mortality Statistics: Deaths Registered in 2009. Available at http://www.statistics.gov.uk/downloads/theme_health/dr2009/dr-09.pdf

Office for National Statistics (2011) Migration Statistics Quarterly Report No. 9: May 2011. Available at http://www.statistics.gov.uk/pdfdir/mig0511.pdf

Robinson, D., Reeve, K. & Casey, R. (2007) *The Housing Pathways of New Immigrants*. York, Joseph Rowntree Foundation.

Scholes, S., Taylor, R., Cheshire, H., Cox, K. & Lessof, C. (2008) Technical report (wave 2): retirement, health and relationships of the older population in England: the 2004 English Longitudinal Study of Ageing. London, National Centre for Social Research.