Sample design for longitudinal surveys

Dave Elliot, Peter Lynn and Paul Smith

In this chapter we review the considerations that influence sample design for longitudinal surveys, the range of sample design options available, and research evidence regarding some of the key design issues. We attempt to provide practical advice on which techniques to consider for different types of longitudinal surveys, where they are most likely to be effective, and what pitfalls to avoid.

1. Types of longitudinal sample design

Surveys that collect data from the same units on multiple occasions vary greatly in terms of the nature of the information collected, the nature of the population being studied, and the primary objectives (Binder, 1998). They also face a variety of practical constraints, from the level of financial resources available to regulations on respondent burden. Consequently, there is also considerable variation in the choice of appropriate sample design. We identify five broad types of longitudinal survey design:

- **Fixed panel.** This involves attempting to collect survey data from the same units on multiple occasions. After the initial sample selection, no additions to the sample are made. In principle, the only loss to the eligible sample is through “deaths”. The 1970 British Cohort Study (BCS70) is an example of a fixed panel design.

- **Fixed panel plus “births”**. Like a fixed panel, except that regular samples of recent “births” to the population are added. Typically, at each wave of data collection a sample of units “born” since the previous wave are added. This may be preferred to a fixed panel if there are non-trivial numbers of births in the population during the life of a panel and there is a desire to represent the cross-sectional population at the time of each wave as well as the longitudinal population of wave 1 “survivors”. Most household panel surveys have this design, as a sample of “births” into the eligible age range is added at each wave.

- **Repeated panel**. This design involves a series of panel surveys, which may or may not overlap in time. Typically, each panel is designed to represent an equivalent population, i.e. the same population definition applied at a different point in time. The England and Wales Youth Cohort Study (YCS) is an example of such a design: the panels consists of samples of the age 16-17 one-year age cohort selected in different years, each panel involving at least three waves over at least three years.

- **Rotating panel**. Pre-determined proportions of sample units are replaced at each fieldwork occasion. Typically, each unit will remain in the sample for the same number of waves. A wide variety of rotation patterns have been used. As Kalton & Citro (1993) note, a rotating panel is in fact a special case of a repeated panel with overlap. It is special because the overlap pattern is fixed, and is typically balanced, and because each panel that is ‘live’ at one point in time is designed to represent the same population, allowing combination of the panels for cross-sectional estimation. Rotating panel designs are often used when the main objectives are cross-sectional estimates and short-term estimates of net and gross change. Labour Force Surveys have a rotating panel design in many countries (Steel 1997).

- **Split panel**. This involves a combination of cross-sectional and panel samples at each fieldwork occasion. A common design described by Kish (1987, p.181-183) involves one panel sample from which data are collected on each occasion, plus a
supplemental cross-section sample on each occasion. A series of cross-sectional surveys in which a proportion of sample elements are deliberately retained in the sample for consecutive surveys – referred to by Kalton & Citro (1993) as an *overlapping survey* – can also be thought of as a type of split panel. A variation is a series of repeated cross-sectional surveys, but with a \( P=1 \) stratum, so that in practice a subset of the sample units are included on every occasion. In this case the objective is a series of cross-sectional estimates, but longitudinal estimates are serendipitously possible for the \( P=1 \) stratum. Many statutory business surveys have this characteristic.

This is of course only a broad typology and it does not fully describe the range of possible designs. For example, each panel in a repeated panel design may or may not include additional regular samples of births. In general, we would expect to use the fixed panel designs where we will select a single sample and follow them to provide detailed information on changes in their characteristics. The repeated and rotating panels are useful for assessing changes in longitudinal estimates since they cover multiple panels.

One motivation for considering a panel design is to be able to construct measures derived from data collected at several time points, where these data could not reliably be collected retrospectively at one time point (e.g. totals or averages of quantities that may change frequently such as income or physical activity, or indicators of whether certain low-salience events were ever experienced). But perhaps the prime motivation is to be able to measure gross change at the element level. This is useful for the study of transitions and stability in many fields. To study phenomena that change only slowly, or that may have long-term effects, long panels are needed. Where change is more rapid and only short- or medium-term effects are of interest, short panels or rotating panels may be preferred.

Often, cross-sectional estimates are required in additional to longitudinal ones. In principle, a fixed panel plus births can achieve this, but in many situations adding regular samples of births to a panel is a very complex and expensive task. A more efficient way to achieve cross-sectional representativeness can be to select a fresh cross-sectional sample. A split panel design may be used in this case or, if the gross change of interest need be observed only over relatively short periods, a rotating panel design.

Repeated panels are often used where the population of interest is naturally thought of as a series of cohorts defined in time, each of which may be expected to have distinct experiences or where policy interventions are likely to apply at the cohort level. Examples might include cohorts defined by age, year of graduation, period of immigration, year of recruitment, etc.

2. Fundamental Aspects of Sample Design

2.1 Defining the longitudinal population

As in any survey it is necessary to define both the target population about which we will want to make inferences, and the frame population of units which are available to be sampled. Ideally these populations will coincide, but in many situations they are not completely congruent, and in these cases some appropriate weighting system may be necessary to adjust (approximately) the sample (from the frame population) to the target population.
However, population definition is complicated additionally in longitudinal surveys because the populations are dynamic. Some method of defining a fixed population on which to base a design is needed, and there are three broad approaches.

i) A static population based on the population at the time the first wave sample is selected.

There can be no “births” but “deaths” are allowed in this population as members either emigrate, change their defining characteristics or cease to exist. For practical purposes we can consider the population size to be fixed, but with units moving to “dead”, from which it may be possible or not to return. Therefore the accuracy of more interesting transitions will reduce with sample sizes as sample units also move to “death”. The big advantage of this approach is that the population information is fixed for all waves, which makes weighting and analysis of subsequent waves relatively straightforward.

A fixed panel design is the most natural design to study a population defined in this way. Such a population may become increasingly less relevant to policy decisions over time as it ages. For example inferences from the early waves of the BCS70 in the 1970s are unlikely to be good predictors of behaviour in the 2000s. There is a danger that users may interpret results as applying to population definition iii) below, in respect of which estimates of change will obviously be biased if new units are more or less likely to change than established units, or if there is a relationship between the variable being measured and the probability of becoming a “death”.

ii) The population is defined as the intersection of the cross-sectional populations at each wave.

With this approach, both births and deaths over the life of the study are excluded from the study population. This creates the same problems as the last case and has the added disadvantage that auxiliary population totals that are needed to calibrate sample estimates may be very difficult to construct. The intersection also changes when a new wave is added.

A variant of this approach is to select the sample from the intersection, but then to provide some appropriate weighting to compensate for differences between the sample and the target population. This latter approach is used on the longitudinal dataset from the UK Labour Force Survey. Note that it only gives unbiased estimates for the target population if there is no relationship between the probability of being in the intersection and the survey variable(s).

iii) The population is defined as the union of the cross-sectional populations at each wave.

This approach seems best suited to a genuinely dynamic population but obviously requires a correspondingly dynamic sample design to introduce births and so avoid bias. This approach is used on the British Household Panel Survey. There can be some difficulty in saying what a sample actually represents (“the population of all people in a household in GB at any time during a five-year period”, for example). Also, when a further wave is added, the union changes and may require revision of previous estimates.

In general therefore option (iii) seems to be the design most likely to fulfil general requirements for longitudinal information, but has difficulties of both methodology and interpretation. In using (i) or (ii) in its place, there is a clear need to explain the limitations of what can be inferred to users of the information – noting however that there is a risk that such limitations may be ignored by users, particularly if no other information is available.
2.2 Target variables

Optimal sample design can not be determined in the abstract, but only relative to some estimate(s). Typically, sample designers identify a small set of key survey variables and attempt to optimise sample design for estimates of means, or possibly other functions, of those variables. Choice of strata, allocation to strata, and the degree of sample clustering will be chosen to minimise the variance of those estimates. In ONS, the key estimates are typically means in the case of cross-sectional surveys, but for rotating panel designs there is a move towards using estimates of change in means. For a longitudinal survey, the key estimates will typically be rates of change, levels of change, durations, hazard rates or other dynamic parameters. Estimates of variance of these quantities are therefore needed (see section 2.7) and knowledge of how the quantities are related to strata or PSUs. Such information can be difficult to estimate as, unlike the case with cross-sectional surveys, there is often no similar recent survey to draw upon.

2.3 Sample size

It is necessary to plan for the eventual sample size required for longitudinal analysis by taking account of any population births and deaths that are reflected in the sample but also specifically taking account of the expected attrition during the life of the survey. Such attrition may arise for several reasons. Sample members who respond at one wave of the survey may not be contacted in the next, either because they have changed their address or because they are temporarily absent when an interview attempt is made (see Calderwood et al, chapter X in this book); or they may refuse to take part in the survey (see Watson and Wooden, chapter Y in this book). Some movers may of course be relocated at their new address, where this is known but some loss of sample numbers can usually be expected from this cause. The treatment of movers in longitudinal surveys is considered further below.

If the proportions of the wave $i-1$ responding sample who are expected to respond at wave $i$ can be estimated as $p_i$ (with $p_1$ indicating the proportion of the initial sample responding at wave 1), then the initial sample size $n_0$, needed to achieve (in expectation) a target number, $n_k$, after $k$ interview waves can be simply calculated. With a fixed panel design, we have $n_0 = n_k / \prod_{i=1}^{k} p_i$. More generally if response at each occasion is independent, then $n_k \sim B\left(n_0, \prod_{i=1}^{k} p_i\right)$, and using a Normal approximation for the confidence interval (under the reasonable assumption that $n_0$ is large) allows us to solve for the value of $n_0$ sufficient to be 95% confident that $n_k$ will be achieved,

$$n_0 = \frac{1}{\prod_{i=1}^{k} p_i} \left[n_k + 2 \left(1 - \prod_{i=1}^{k} p_i\right) + 2 \sqrt{\left(\prod_{i=1}^{k} p_i\right)^2 - (2 + n_k) \prod_{i=1}^{k} p_i (1 + n_k)}\right].$$

In a rotating panel design with $k$ waves and an initial sample size of $n_0$ in each panel, we have $n_0 = n_k / \sum_{j=1}^{k} \prod_{i=1}^{j} p_i$ (in expectation), where $n_k$ is the target responding sample at each survey occasion, consisting of the sum of the respondents from each panel. Note that these all assume that once a unit is lost to the panel it is never recontacted. In a situation where units may be missed on one occasion but subsequently participate again, we would expect to achieve at least $n_k$ interviews.
Experience of net attrition rates in the UK, in both quarterly and annual panels suggests that the greatest loss occurs between the first and second waves. For example on the UK Labour Force Survey which uses a 5-wave quarterly rotating address panel, \( p^T = \{0.728, 0.878, 0.963, 0.936, 0.956\} \) in Q1 2006\(^1\).

These approximations assume the same attrition across the whole population, but this is unlikely to be true in practice, so instead we would expect to calculate \( n_0 \) separately within groups of relative homogeneity. In a sound sample based on good auxiliary information, this will mean that the most difficult cases will be heavily over-represented at the first wave, to allow for later attrition, which is likely to make the survey more expensive. Cutting back on this expense risks having estimates with high variances and biases at later waves.

### 2.4 Clustering

Decisions about the nature and extent of sample clustering should depend on slightly different criteria in the case of a longitudinal survey. Specifically, the relationship between clustering and data collection cost may change over the waves of a survey due to mobility of sample units and sample attrition. The impact of clustering on variance of estimates (design effects) may also change. To take this into account at the sample design stage requires knowledge of the number and timing of waves, and estimation of the changing effects of clustering over those waves.

For example, the General Household Survey in Britain has, since 2005, used a 4-wave rotating panel design with annual interviews. Unit cost is reduced by clustering the initial sample within a sample of postal sectors. But unit cost increases at each subsequent wave as attrition and mobility in the sample reduces the number of interviews in each postal sector. One way to partly ameliorate this would be to retain the same postal sectors in the sample and perform the rotation of addresses within these sectors, but that could lead to biases in cross-sectional estimates over time. A careful compromise is therefore needed between selecting a sample size per sector that is large enough at wave 1 to still constrain unit costs at wave 4 and controlling the variance of estimates by limiting the wave 1 sample size per sector.

Identification of the optimal level of sample clustering is even more difficult with long-term panels, especially if the eventual life-span of the panel is unknown at the time of initial design. This was the case with the British Household Panel Survey, which was conceptualised as a survey with an indefinite number of annual waves, but with initial funding only for five years. The decision to use a clustered design was based on crude estimates of the impact over the first few years of the survey. After many years of the survey, the unit cost of data collection is only marginally less than would be the case with a completely unclustered design, while design effects are still considerably greater than with an unclustered design. Therefore for a long-running longitudinal survey it is better to have an unclustered sample, as the accuracy in later years will offset the additional cost in the earlier years.

### 2.5 Treatment of Movers

Movers create two types of problems in longitudinal surveys. First, there is a need to collect new contact information and an associated increased risk of failure to contact the sample unit. Second, there are additional costs. Most social surveys that employ face-to-

face interviewing use some form of cluster sampling or multi-stage sampling to control costs, so following a mover to a new address may incur considerable extra cost if the new address is not in one of the original sample areas. Also, there is a risk that no interviewer may be available to visit a mover if the move is only discovered during the field period, so procedures need to be put in place to track and record moves among the sample members. Problems are fewer on business surveys, as they tend to use remote methods of data collection, but there is still the challenge of identifying movers before the survey, because of lags in the administrative systems which are often used to create the frame.

However, omitting movers may create an obvious bias in some surveys. One way of bypassing this problem is to define sampling units which do not move. In the UK LFS, the ultimate sampling unit is an address, so that movers out of a sample address during the study period cease to be part of the sample but are balanced by movers in. It is important in such a design to retain vacant addresses identified in any wave as part of the sample to be issued to interviewers in subsequent waves.

2.6 Stratification

Ideally, sampling strata should be defined that are relatively homogeneous in terms of a survey’s target variables. In practice, of course, the ideal strata often can not be determined in advance; cannot be identified in the sampling frame; or are quite different for different target variables. In longitudinal surveys, the main targets for inference will often be measures of individual or gross change, both overall and for domains, for which the standard cross-sectional stratifiers may be of little use. Special studies may be required to identify suitable candidates, by analysing the relationships between survey-based estimates of change and potential stratification variables. Even obtaining survey-based estimates of change may be difficult if there is no previous longitudinal data to work with.

2.7 Variances and design effects

There is a substantial literature on estimating changes in totals and means from longitudinal and particularly rotating samples. Kish (1965) was one of the first to discuss the problem. His analysis applies in a static population when the fpc can be ignored. Holmes & Skinner (2000) also give useful general results for variances of post-stratified and calibrated estimators of the change in means and totals for this case. In surveys with large sampling fractions the finite population correction is both important and open to a variety of different definitions - see Tam (1984), Laniel (1988), Nordberg (2000) and Berger (2004), where the analysis is extended to dynamic populations and cases where the fpc cannot be ignored.

Design effects for complex statistics such as these are often found to be less than those for simple linear statistics from a single sample (Kish & Frankel 1974) but Skinner & Vieira (2005) found unusually high design effects for regression coefficients based on a longitudinal sample, with the design effects growing with the number of occasions that were included in the analysis.

2.8 Selection probabilities

Any longitudinal survey that includes a mechanism for sampling births must also allow the estimation of the selection probabilities of units that are added to the sample subsequent to the initial selection. This is not always simple. For example, in the case of household panel surveys a common sampling mechanism is to add to the sample the
members of households into which members of initial sample households move. The selection probabilities of such people are a complex function of their household membership history over the life of the panel, but fortunately, it is not required to know these probabilities. Instead, the ‘weight share’ method (Ernst, 1989; Lavallée 1995) can be used to produce unbiased estimates so long as people not in the initial households are not followed if they later leave the households of a person from the initial sample.

3. Other Aspects of Design and Implementation

3.1 Choice of rotation period and pattern

As with most aspects of sample design, the rotation pattern is a compromise between how frequently and how many times we are prepared to resurvey a sample unit, and which characteristics we most want to analyse. Where changes over short periods are important, as often in business surveys, we will typically use a pattern in which a sample unit is revisited on every survey occasion (for example monthly) for \( T \) occasions. Steel (1997, p20) characterised these designs as “in-for-\( T \)”, and contrasted them with designs where changes over longer periods than the frequency of the survey are also important, typically annual changes within a survey with a subannual period. These types of designs usually have a period in sample, a period out of sample and then a further period in sample, and are denoted \( T-O-T(n) \) with \( n \) giving the number of periods in sample – as in the US Current Population Survey (CPS) which has a 4-8-4 design (a sampled household is interviewed for four consecutive months, left out for eight months, and then interviewed for four months again). This type of design, ensures some annual overlap, and therefore takes advantage of sample autocorrelation in estimating annual flows.

The number of occasions on which a unit can be sampled is typically a balancing act between the length of time over which changes are to be estimated, the number of occasions on which it is reasonable to expect a sample unit to contribute to a survey, and operational challenges. In the case of business surveys periods are often quite long to ensure some annual overlap. Many ONS monthly business surveys had an in-for-15 design in the mid-1990s to help support the management of response burden while allowing some annual overlap. However, respondents who are newly included in the sample require more effort to obtain a usable response (that is, the start-up costs are high relative to the ongoing costs), and replacing \( \frac{1}{15} \) of the sample on each occasion proved relatively expensive, so in several surveys the rotation pattern has lengthened to 27 months, giving an in-for-27 design, although this has a much greater impact on individual respondents (Smith, Pont & Jones 2003). Statistically the gain is small, since the monthly overlap was already high, and has gone from 93% to 96%. There is much more impact on annual overlap which has gone from 20% to 56%, but this is not the main focus of design for short-period business surveys. The big change is in the number of new respondents each month, which is almost halved.

In contrast the costs in longitudinal surveys of households are often concentrated in the follow-up because of the need to follow people moving house (that is, start-up costs are small relative to ongoing costs), in which case the operational and burden constraints act together to keep the number of follow-ups small.

From this discussion it is clear that the objective of the survey has an impact on the choice of rotation period. The longer the period over which changes are to be analysed, the longer the period of surveys required. Over relatively long periods, however, the respondents are more likely to become jaded and drop out of the survey.
One strategy for reducing the impact of a large number of visits while maintaining relatively high sample overlaps for fixed periods is to have a discontinuous pattern, such as the 4-8-4 of the US CPS. This gives 75% overlap between consecutive months and 50% overlap between months a year apart, but low overlaps for other changes relative to an in-for-$T$ design.

So a stylised version of the decision tree for deciding on the rotation pattern for a longitudinal survey might be:

### 3.2 Dealing with births (and deaths)

Most populations are not fixed, and are affected by births (immigration) and deaths (emigration). How to deal with these in a longitudinal study is important because of the effect on the validity and interpretation of the estimates. First let’s take the easier challenge – how to deal with deaths. As long as these can be identified and distinguished from non-response, they are easily incorporated in analyses by using a code for dead units (and zeros for any numeric variables). If they are not identifiable, the challenge is greater. Imputation based on the valid responses will give biases in flows between states, underestimating flows to deaths and overestimating other flows. But from a design perspective, future deaths are already present in the sample, and so no special action is required.

Births, however, are another matter. With any non-static population, new births will occur continuously, and if the sample is not updated, it will quickly become deficient in this kind of observation. This means having a mechanism for identifying births to the
population, and sampling from them; appropriate weighting will be needed if the sampling fraction is different from the initial sample or differs from occasion to occasion if births are added to the sample periodically.

Where there is a good frame for the population that is kept up-to-date, as in many business surveys, the problem of keeping the sample up to date while retaining the longitudinal linkage is facilitated by using permanent random numbers (PRNs) as the basis of sampling (Ohlsson 1995, Smith, Pont & Jones 2003). Here, each unit on the register (including new births) is assigned a PRN when it is first added to the register. To sample, units in the population are ordered by their PRN and then a range of the available PRNs is used to define the sample; new births are included in the sample when their assigned random number falls in the same sample range. Varieties of PRN sampling can be used to manage rotating panel designs as well as for strictly longitudinal surveys.

3.3 Sample overlap

Where longitudinal information is produced as a by-product rather than the main purpose of a panel or rotating panel design, there will be the need occasionally to redesign or reclassify the main survey. In this instance we have an interest in maximising the overlap between the two designs, and there are methods for doing this – see Kish & Scott (1971) and Hidiroglou et al. (1991, section 2.2.4).

3.4 Stability of units and hierarchies

There are two forms of instability over time that can affect sample design for longitudinal surveys. The first is inherent instability in the sample units. In some cases, the sample units (e.g. people) are well-defined and unlikely to change identification during the period of the survey. But in other cases (e.g. businesses), the units are not stable, and can split or merge with other units. In these cases it is necessary to define clearly at the start what the approach will be – to follow up all parts or only certain parts, and how they will be treated in analysis.

The options of which parts of a sample to follow up are, from most restrictive (and cheapest) to least (and most expensive):

i) Only the entities initially sampled, eg households, so any change results in the unit dropping out of the survey. This is so restrictive that it would only really be useful in the study of a characteristic which is of the household, and not of the individuals within it, and we don’t know of any examples of this kind of longitudinal study;

ii) all the components of the initial sample, wherever they end up, for example following all people in initially sampled households;

iii) all components of the initial sample, and any components with which they join, for example all households which contain any people from the initially sampled households;

iv) all components of the initial sample, and any components with which they have joined since the start of the study, for example all households which contain any people who have lived in a household with one of the initially sampled people. This last example is really a version of network (snowball) sampling (Berg 1996), and has the potential to lead to very large samples for later waves, and so is not normally adopted.
Options ii) and iii) are therefore the most practical. There is little in travel costs between them, since the same households must be visited in either case; the main difference is therefore whether a whole household is interviewed or only part of it. Whole households may be best where some context of the new household will be a useful explanatory in analysing the longitudinal outcomes, whereas for more personal data (for example medical history), it may be sufficient to follow up the original sample people only.

Another possibility is that the units themselves may be stable, but the classification of units may change over time. In other words, the characteristic that determines the classification is potentially transient. An example might be persons with low income. This can be problematic if the intention is to restrict the sample to units in a particular category of the classification or to sample different categories at different rates. The first approach (restricting the sample) is only appropriate if the population of interest is genuinely a static one (e.g. people who had a low income on a particular date). The second approach (unequal sampling fractions) may improve the precision of (some) estimates initially but may reduce precision for equivalent estimates after several waves, depending on the speed at which changes in classification occur and the length of the panel. These factors can be difficult to predict in advance, making it difficult to determine the optimum allocation to strata (categories of the classification).

4 Conclusion

The range of issues for consideration when constructing a longitudinal sample design is large; they often need the complexity of cross-sectional designs with additional constraints and considerations for the longitudinal element. All these elements must be traded off to arrive at a final design achieving the best balance of properties within financial constraints.

The major steps are then to:

• define the units to be followed, and any rules for changes to those units;
• define the (longitudinal) population of units;
• ensure that the sample is representative (measurable) for the population (including any cross-sectional populations) required;
• make use of design procedures (stratification, clustering, etc, for longitudinal outcomes) to optimise the use of the available resources;
• implement the design.

References


BINDER, D.A. (1998) Longitudinal surveys: why are these surveys different from all other surveys?. Survey Methodology 24, 101-108.


