Some Issues in the Design and Analysis of Longitudinal Surveys

Graham Kalton*
Westat, Rockville, Maryland, USA. grahamkalton@westat.com

Abstract

Longitudinal studies in survey research include panel surveys, rotating panel surveys, and cohort studies. This paper reviews a selection of issues arising in the design and analysis of such studies, supported by illustrations from existing surveys. Panel and rotating panel studies are widely analyzed both longitudinally and cross-sectionally, often as repeated cross-sections. For cross-sectional analysis, the longitudinal sample needs to be augmented by samples of new entrants to the population. It may also become necessary to augment the sample for panel losses. Many sets of weights may be needed for the various types of analyses. Measurement errors and imputed values can be particularly harmful for estimates of gross change. Panel conditioning may seriously affect the results of analyses that make comparisons over time. Changing modes of data collection and changes in respondents across waves of a panel can also affect comparability, as can item nonresponse and imputation for it. A commonly encountered design issue is whether to oversample certain domains; here it is important to distinguish between domains defined in terms of static characteristics and those defined in terms of transient characteristics. The geographical spread of the initial sample of primary sampling units needs to take account of subsequent population mobility.

Keywords: cohort design, panel attrition, panel conditioning, rotating panel design.

1. Introduction

I have chosen the topic of longitudinal surveys for my paper in this memorial session for Gad Nathan because Gad made many important contributions in this area and because he worked on two projects related to longitudinal studies on extended visits he made to Westat in the 1990s, projects with which I was associated. My coverage focuses mainly on issues of design and basic analysis. It does not extend to statistical methods for analyzing longitudinal surveys, an area in which Gad also made important contributions (see, for example, Nathan, 2009).

The topic of panel surveys covers a wide range of designs and methodological issues. Rather than attempt a general overview, I have chosen a set of design and analytic issues that I have encountered recently with a number of panel designs of human populations. Broader treatments can be found in the books edited by Kasprzyk et al. (1989) and Lynn (2009), and additional references are given by Kalton (2009). For space reasons, references are not provided for specific studies, but all have readily accessible websites from which further details can be obtained.

There are various forms of longitudinal surveys, developed to satisfy different analytic objectives. To set the scene for the design issues taken up in Section 3, Section 2 provides a short overview of some alternative longitudinal designs. Section 4 presents some concluding remarks.

2. Panel Designs

It is useful to define three types of longitudinal designs for the purpose of this paper:

- *Cohort studies*. These studies are a special type of panel survey that follows a sample of a defined subgroup of the population over time. A common form of this type of study is a birth cohort, which starts with a sample of births in a given time period. Birth cohort studies have been conducted in many countries, and some have followed sample members over many years, even following the offspring of sample members. The U.S. National Children's Study (NCS), currently under development, is one example of a birth cohort study. The U.S. National Longitudinal Surveys (NLS) are also cohort studies. They follow several cohorts defined by sex and starting age group (e.g., youth aged 12-17, women aged 35-44).

- *Panel studies*. Panel studies in general follow a sample of members of a population over time, sometimes long periods of time. Often the initial sample is selected from some general population rather than from a more narrowly defined cohort. Some studies select persons directly; others are household panel surveys that sample households, take all persons in the sampled households, follow those persons in the panel, and sometimes collect survey data for all those living with the sampled persons at each wave of data collection. The Panel Study of Income Dynamics (PSID) is perhaps the earliest version of a household panel survey.

- *Rotating panel surveys*. Rotating panel surveys are a series of overlapping panel surveys of fixed duration. At each wave of data collection, one panel is terminated and a new panel begins. Many labor force surveys employ rotating panel designs.

All these designs collect longitudinal data on the sampled persons. They thus provide the data needed for analyses of gross change over time (although only for a limited period of time for a rotating panel). In essence, gross change is the driving analytic focus of cohort and panel studies. However, these designs are also widely used for analyses of net change and for cross-sectional analyses. Rotating panel surveys are primarily designed to produce cross-sectional estimates and efficient estimates of net change.

The several waves of data collection in longitudinal studies provide the opportunity to collect data on a much wider range of variables than is possible in a cross-sectional survey. The combination of longitudinal data and the collection of data on a wide range of subject matter presents special challenges for deciding design priorities for satisfying a longitudinal study's multiple potential analytic objectives.

3. Some Design and Analysis Issues

This section contains a partial set of issues in designing a longitudinal data collection. They are organized into three broad areas: sampling, sample losses, and measurement.

3.1. Sampling

In addition to the general sample design issues for a cross-sectional survey, at least three additional issues arise with a cohort or panel design:

- *General sampling considerations for a panel study*

A general point not always fully acknowledged in panel design is that the benefits of a high-quality initial sample design accrue across the full period of the study. It is therefore

wise to invest heavily in developing and implementing a good design at the outset because the costs can be amortized across the full study.

Many longitudinal studies start with multistage area probability designs because the concentration of the sample in selected areas facilitates the initial data collection. However, the study designers need to also consider the data collection procedures to be applied in subsequent waves when sizable proportions of the sampled persons have moved to new locations. If in-person contacts are required at subsequent waves, the use of the same interviewer field force is facilitated by the selection of a large initial sample of primary sampling units (PSUs) that provides a widespread geographical coverage.

- *Domain oversampling*

Oversampling is used in many cross-sectional surveys to provide adequate sample sizes for certain smaller domains of special interest, for example for race/ethnicity, poverty, and geographic domains. It is also frequently considered for panel and cohort studies. However, given the usual multiple objectives for such studies, as well as changes in objectives over time, oversampling should be applied judiciously, if applied at all. In designing the sample for a panel or cohort study, careful consideration should be given to the fact that oversampling one set of domains will adversely affect the precision of estimates for other domains and for the total sample.

It is important in longitudinal studies to distinguish between domains that are defined by permanent characteristics (e.g., race) and those that are defined by transient characteristics (e.g., poverty status). Oversampling by a permanent characteristic can work well provided the survey objectives do not change over time; however, if the objectives change later on in a way that no longer requires estimates by that characteristic, the oversampling causes an undesirable loss of precision. The loss can be substantial if the oversampling rate is high. For that reason, it is safer to keep the oversampling rates within reasonable bounds. In the early planning stages, the NCS considered oversampling by characteristics related to health disparities and exposure to certain environmental contaminants, but decided against oversampling given the potentially harmful effects for the study's aim to serve multiple objectives.

The use of oversampling is more much problematic for transient domains. For example, consider oversampling persons in poverty at wave 1. Some of those oversampled will no longer be in poverty at wave 2, whereas some of those not oversampled will have moved into poverty at wave 2. Oversampling transient domains of critical interest may be dictated by the survey objectives, but the effects of this oversampling over time need to be carefully assessed. In the Population Assessment of Tobacco and Health (PATH) study, oversampling of tobacco users is planned, but attention has been paid to controlling the rate of oversampling. In the National Health and Aging Trends Study (NHATS), which is a panel study of functioning in persons aged 65 and over, oversampling is employed to increase the sample sizes for the oldest age groups. This oversampling was needed to satisfy an important analytic need for the oldest old domain, and particularly for those aged 90 and over; however, over time, many of these persons will die and the sample size of those entering this domain will be much smaller. This feature was recognized at the design stage, and the consequences accepted.

- *Freshening the sample for new entrants*

The focus of cohort studies is generally restricted to the population that defined the cohort at the time of the initial sample selection. With that definition, there are no new

entrants to the population over time. However the addition of new entrants, such as recent immigrants and persons who reach the lower age limit for study eligibility, does need to be considered for panel and rotating panel surveys. New entrants may not contribute to longitudinal analyses that start from the beginning of the panel, but they can be important for cross-sectional analyses at later waves and for longitudinal analyses that start later in the life of the panel. The need for supplementary samples to give representation to new entrants increases as the duration of the panel lengthens.

Freshening a panel sample to cover new entrants can sometimes be straightforward, as with NHATS where the sample is selected from a frame of Medicare recipients that covers a high proportion of the U.S. population aged 65 and over. In contrast, lacking such a frame, the Health and Retirement Survey, a household panel study with a target population of persons aged over the age of 50, needs a large screening survey to identify those who age into the population at later waves. As another example, the Panel Study of Income Dynamics conducted a large screening survey in 1997/1999 to freshen its sample with immigrants to the United States who arrived after the study started in 1968. The target population for the PATH study starts with youth aged 12 and above. The technique for freshening the PATH sample is to follow children aged 9 to 11 in sampled households and add them into the panel as they reach age 12 in the three subsequent waves; in the fourth wave another screening technique will be needed.

In studies like NHATS and PATH, new entrants can be added at each wave to maintain full representation of the target population on an ongoing basis. However, while that is possible for PATH, for NHATS the special nature and contents of the enrollment interview and of the subsequent interviews would make the addition of new entrants at each wave extremely difficult. For this reason the sample will be freshened less frequently.

3.2. Sample Losses

All longitudinal studies are subject to sample losses at the second and subsequent waves. Some sample members may miss one or more waves, whereas others drop out on one wave, never to return to the panel. A study's following rules define whether any data collection should be attempted at a given wave with nonrespondents at the previous wave and, if so, which nonrespondents. For example, the following rules may specify that sample members are to be dropped from the panel if they were adamant refusals or untraceable at the previous wave, if they have failed to respond to two previous waves, or if they have now moved to locations inaccessible for data collection.

Panel sample losses complicate analyses and often lead to a multitude of sets of weights to enable analysts to retain as large a sample as possible for their particular analyses. The considerable literature on weighting and/or wave imputation for longitudinal data sets is outside the scope of this paper (see, for example, Pfeffermann and Nathan, 2001, and references therein). The key issue is whether the reweighted sample provides a representative sample of the population, an issue that Nathan (1999) examined in detail for three long-term longitudinal studies in Britain. Compensating for potential nonresponse bias from wave nonresponse is challenging; see, for example, Eideh and Nathan (2009) for an approach for estimating gross flows from a rotating panel that is subject to wave nonresponse.

Another concern about sample loss is the reduction in sample size that it creates. Sometimes substitution is used in surveys to control sample size (see, for example,

Nathan, 1980). In the case of panel nonresponse after the initial wave, the drop in sample size may be addressed by sample replenishment, adding new sample at later waves in compensation. This may be done in conjunction with sample freshening for new entrants, but it is more complex. The aim is to select a substitute sample that replaces the attrition cases with clones of them. One approach is to identify the distribution of demographic characteristics of the sample losses and select the replenishment sample to mirror that distribution. The replenishment sample serves the need for cross-sectional analyses at later waves and for longitudinal analyses from that point on, but it does not readily help with longitudinal analyses that start with data from waves before the replenishment wave.

3.3. Measurement Issues

The raison d'étre for longitudinal studies lies in their ability to measure gross changes over time. Like repeated surveys with independent samples, they can also produce estimates of net change and time trends, provided that issues of freshening the sample and sample losses are resolved. They can in fact produce more precise estimates of net change, taking advantage of a generally positive correlation in the responses of sampled persons on successive waves. However, like repeated surveys, they risk collecting non-comparable data over waves due to changes in any aspect of the survey methods. Indeed, longitudinal studies are particularly vulnerable to changes in data collection methods, for example changing from face-to-face interviewing at the first wave to some other mode, such as telephone interviewing or web-based data collection, at the second and subsequent rounds. Also, context effects may create non-comparability in the responses to core items since different question modules may be asked in different waves. Another possible source of non-comparability is panel conditioning, whereby a sample member's response on one wave may influence the response to the same question on a later wave.

At each wave of a panel an individual's survey response can be represented by a true value, an individual bias, and a random error with mean of zero. The concern for analyses of net change and trends is that differential biases in responses across waves result in differences in the means of the individual biases across the waves. Random errors are a much lesser concern since they do not lead to a bias in the net change estimate, although they do lower the estimate's precision. However, random errors are a serious problem for measures based on gross change since they attenuate associations over time. This attenuation can be corrected if the reliability of the survey measures is estimated. When gross change is to be analyzed, it is therefore very valuable to conduct a reinterview study to provide measures of the reliabilities of the key survey variables. When conducted, such studies are often carried out as part of a pilot study. However, it is preferable to conduct a reliability study in conjunction with the main survey data collection to obtain reliability measures under the actual survey conditions.

Rotating panel surveys provide the data needed to investigate comparability of cross-sectional estimates across the different waves of the panel, and there is often clear evidence of what is termed rotation group bias. It is hard to disentangle the extent to which this bias is due to changes in methods, panel conditioning, and sample losses. On one of his visits to Westat, Gad Nathan led an investigation of this issue for a survey in which half the sample was reinterviewed from the previous year while the other half was a fresh sample. Despite extensive analyses, no clear explanation was found to account for the large differences in the estimates produced by the two samples. Rotating panel studies can also provide data for analyzing gross changes in some cases. See Kantorowitz and Nathan (1987) for an investigation of the extent and pattern of changing responses for unchanging characteristics across waves of a labor force panel. Changing from self-

response to proxy response across waves is often an important source in the overestimation of gross changes.

A particular case of measurement error occurs when imputation is used to assign values for item nonresponses. Imputation can attenuate correlations of the item being imputed with other items unless those other items are included in the imputation model. In panel surveys, to produce a data set for immediate analysis, the imputations are mostly conducted at the completion of each wave, incorporating variables from that wave and previous waves in the imputation model. However, when the next wave's data become available, the pool of potential data for use in the imputation model increases. In particular, the same variable measured on the subsequent wave will likely be an important variable to include in the imputation model to avoid distorting the gross change measure. The British Household Panel Study has applied an approach that uses data from the previous, current, and following wave in the imputation model.

4. Concluding Remarks

Gad Nathan has made many important contributions in a number of different areas related to the design and analysis of longitudinal studies. My scope here cannot address all these areas. Instead, I have restricted my coverage to a selected set of topics that I consider to be important in the design and conduct of longitudinal studies.

References

Eideh, A. H., and Nathan, G. (2009). Joint treatment of nonignorable dropout and informative sampling for longitudinal survey data. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 251-264). Chichester, U.K.: Wiley.

Kalton, G. (2009). Designs for surveys over time. In D. Pfeffermann and C.R. Rao (Eds.), *Handbook of statistics: Vol. 29A. Sample surveys: Design, methods, and applications* (pp. 89-108). Amsterdam: Elsevier.

Kantorowitz, M., and Nathan, G. (1987). The estimation of response error micro-effects from repeated surveys for invariant characteristics. *Proceedings of the Third Annual Research Conference*, 359- 390. Washington, D.C.: U.S. Bureau of the Census.

Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P. (Eds.). (1989). *Panel Surveys.* New York: Wiley.

Lynn, P. (Ed.). (2009). *Methodology of longitudinal surveys.* Chichester, U.K.: Wiley.

Nathan, G. (1980). Substitution for non-response as a means to control sample size. *Sankhya C, 42*, 50-55.

Nathan, G. (1999). *A review of sample attrition and representativeness in three longitudinal surveys.* GSS Methodology Series No. 13. Newport, U.K.: Office of National Statistics.

Nathan, G. (2009). The analysis of longitudinal surveys. In D. Pfeffermann and C.R. Rao (Eds.), *Handbook of statistics: Vol. 29B. Sample surveys: Inference and analysis* (pp. 315-327). Amsterdam: Elsevier.

Pfeffermann, D., and Nathan, G. (2001). Imputation for wave nonresponse – existing methods and a time series approach. In R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (Eds.), *Survey nonresponse* (pp. 417-429). New York: Wiley.