Chapter 1

Introduction: Longitudinal research design and analysis

Scott Menard

1 Longitudinal and cross-sectional designs for research

As described in Menard (2002), longitudinal research designs can best be understood by contrasting them with cross-sectional research designs. In a purely cross-sectional design, data are collected on one or more variables for a single time period. In longitudinal research, data are collected on one or more variables for two or more time periods, thus allowing at least measurement of change and possibly explanation of change. There are some designs which do not fall neatly under the definition of pure cross-sectional research or longitudinal research. One example is research in which data are collected for different times for different cases, but only once for each variable, and the time dimension is ignored. This design may be used, for example, when data are not all available at the same time, as in Ahluwalia's (1974; 1976) study of economic development and income inequality. Although the data come from more than one time period, the design for any given case, and also the analysis, is cross-sectional in nature. The danger here lies in assuming that relationships are constant over time; the alternative is that any bivariate relationship may reflect not the relationship one would obtain if all of the data were measured for a single period, but may instead be contaminated by changes in that relationship over time.

Another possibility is a time-ordered cross-sectional design, in which each variable is measured only once, but variables are, by design, measured at different times. An example of this is the study by Tolnay and Christenson (1984), who deliberately selected variables which were measured at different times for use in a causal path analysis of fertility, family planning, and development. Each variable was measured at the same time for all countries, but different variables were measured at different times, in order to match the temporal order of measurement with the causal order in the path model. Although measurement occurred at different times for different variables, each variable is measured only once for each case, and the data cannot be used to perform even the simplest true longitudinal analysis (e.g., measuring change in a variable from one period to another). Once again, the design and the analysis are essentially cross-sectional in nature. Had Tolnay and Christenson chosen to postulate instantaneous effects, the analysis could have been performed just as well with purely cross-sectional data.

For the purposes of the analysis (evaluating direct and indirect effects of family planning effort and development on fertility), this design is reasonable, and may have an advantage over models in which causal order in the path model and temporal order of measurement are not the same (Menard and Elliott 1990a). The use of time-ordered cross-sectional data, as in Tolnay and Christenson (1984), is desirable once temporal order has been established, but as described in Menard (2002), it is insufficient to insure that one does not "predict" a cause from its effect. With a true longitudinal design and analysis, it might be possible to ascertain the true causal direction in the relationship between X and Y. With cross-sectional data, even time-ordered cross-sectional data, we run the risk of undetectable misspecification because of incorrect causal ordering in the model being estimated. With longitudinal data, incorrect causal ordering is more likely be detected, and the model can be corrected.

2 Designs for longitudinal research

Menard (2002) describes four basic designs for longitudinal research: total population designs, repeated cross-sectional designs, revolving panel designs, and longitudinal panel designs. These designs are illustrated in Figure 1.1, and examples of each are provided in Chapters 2–6 of Section I in this volume. In Figure 1.1, the horizontal dimension represents the period (a month, year, or decade) for which data are collected, and the vertical dimension represents the cases (population or sample) for which data are collected. Moving from left to right, vertical lines on the left indicate entry into the population or sample being analyzed, and vertical lines on the right indicate exit, as indicated in the first part of Figure 1.1.

In a *total population design*, the total population is surveyed or measured in each period of the study. Because some individuals die and others are born from one period to the next, the cases are not identical from one period to the next, but if the periods are short, the overwhelming majority of cases may be the same from one period to the next. As one example, the decennial census of the United States attempts to collect data on age, sex, ethnicity, and residence of the total population of the United States every ten years, and does so with an accuracy estimated at 95–99% (Hogan and Robinson 2000; Robey 1989). With somewhat lower, but still substantial accuracy and completeness of coverage, the Federal Bureau of Investigation's *Uniform Crime Reports* attempt to collect data on arrests for specific offenses and, for a limited set of offenses, crimes known to the police, plus the age, sex, race, and residence (urban, suburban, or rural) of arrestees for all police jurisdictions in the United States. In Chapter 2 of this volume, Margo Anderson illustrates the use of total population data, specifically census data, for longitudinal research. To the extent that individual data across time are recoverable from the total population data, the total population design permits the use of all possible methods of longitudinal data analysis, but total population designs are most commonly used in aggregate rather than individual level research, and more often involve analytical techniques such as those in Chapters 13–14 (analyzing developmental and historical change) and Section VII (time series analysis and deterministic dynamic models), rather than techniques better adapted to analysis of change at the individual level. In addition to this type of analysis, which focuses on *changes in the values of variables* (e.g., changes in per capita gross national product or changes in homicide rates) over time, this type of design is also well suited to the analysis of *changes in relationships among variables* (e.g., the correlation between ethnicity and political affiliation, or between education and income) over time.

Each of the other three longitudinal designs in Figure 1.1 involves a sample drawn from the total population, and is thus a subset of the total population design. The three designs differ in the extent to which the same or comparable cases are studied from one period to the next.

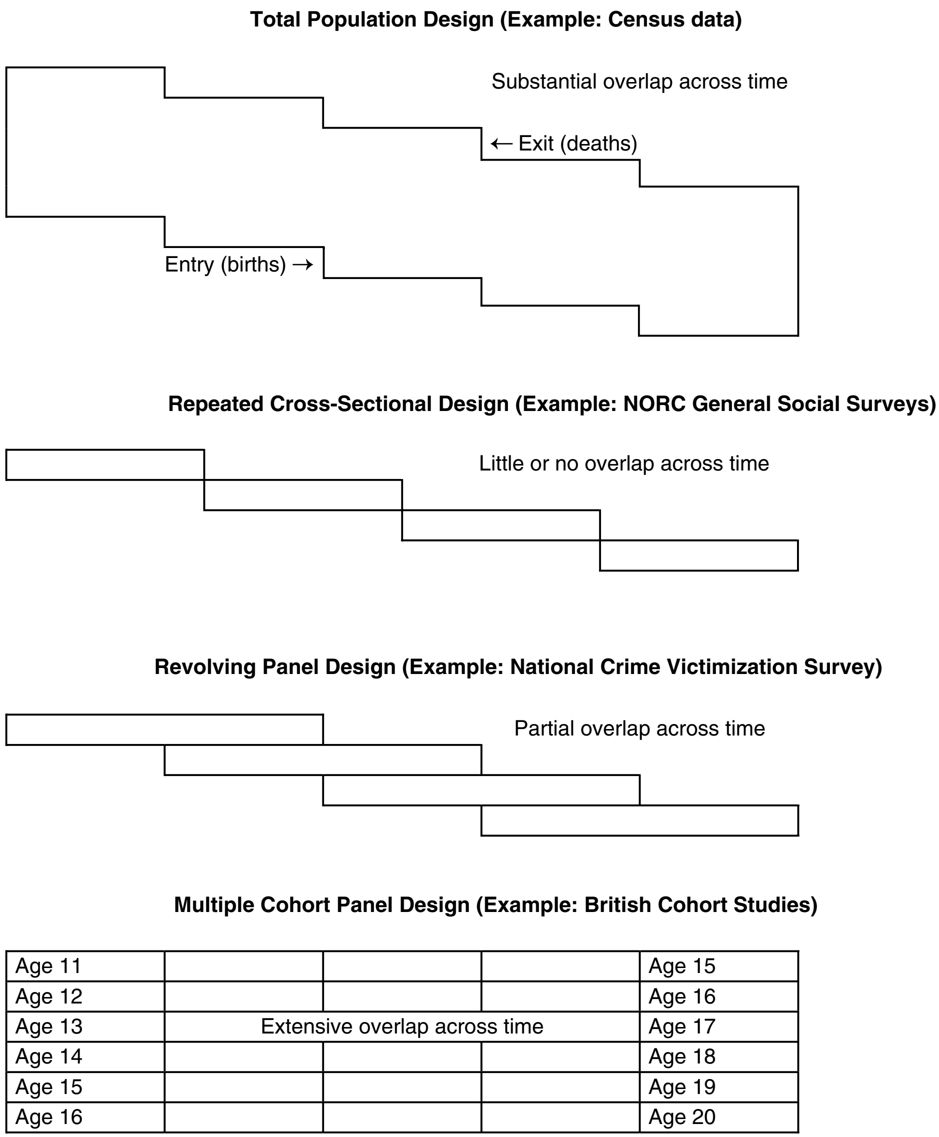

Figure 1.1 Longitudinal designs for data collection

This distinction has important implications for which types of longitudinal analysis are possible with each design. In the *repeated cross-sectional design*, the researcher typically draws independent probability samples at each measurement period. These samples will typically contain entirely different sets of cases for each period, or the overlap will be so small as to be considered negligible, but the cases should be as comparable from one period to another as would be the case in a total population design. An example of the repeated cross-sectional design is the General Social Surveys (GSS), which include an annual general population sample survey conducted by the National Opinion Research Center, which covers a wide range of topics, and emphasizes exact replication of questions to permit comparisons across time (Davis and Smith 1992). Thomas W. Smith, in Chapter 3, describes the GSS, including the methods used to collect the data, and gives an overview of the types of research that have been done using this extensive dataset. Much of the research involving repeated cross-sectional data is cross-sectional in nature, and even more than is the case with total population data, the analysis of change in repeated cross-sectional data may involve aggregate level research; and like the total population design, the repeated cross-sectional design is well suited to examine changes in values of variables and in relationships among variables over time.

Revolving panel designs collect data on a sample of cases either retrospectively or prospectively for some sequence of measurement periods, then drop some subjects and replace them with new subjects. The revolving panel design may reduce problems of panel mortality and repeated measurement in prospective studies (to be discussed in Section II), or problems of extended recall periods in retrospective studies. Retention of a particular set of cases over several measurement periods allows short-term measurement of change on the individual or case level, short-term analysis of intracohort developmental change, and panel analysis. Replacement of the subsample which is dropped in a measurement period with a new but comparable subsample of cases permits analysis of long-term patterns of aggregate change, similar to the analyses possible with total population and repeated cross-sectional designs. If the time lag between cause and effect is smaller than the time (periods) for which cases are retained in the sample, analysis of temporal and causal order is possible. The combination of longitudinal data involving repeated measurement on some cases with data which do not involve repeated measurement on others may permit comparisons which can indicate whether repeated measurement is producing any bias in the data (e.g., increased or decreased willingness to report events after either building up some level of trust or finding out that reporting leads to long and tedious follow-up questions). A good example of a revolving panel design is the National Crime Victimization Survey, whose use in longitudinal research is described by Lawrence Hotchkiss and Ronet Bachman in Chapter 4.

In a longitudinal panel design, the same set of cases is used in each period. In practice, there may be some variation from one period to another as a result of missing data. For example, when cases are individuals, some of those individuals may die between one measurement period and the next, others may not agree to cooperate, and others may move to new locations and not be found by the researcher. All of these are sources of *panel attrition*, and apply primarily to *prospective* panel designs, in which measurement or data collection occurs *during* more than one period as well as *for* more than one period. The combination of measurement during more than one period and for more than one period represents, for some scholars, the only true longitudinal design, the only design that allows the measurement and analysis of intraindividual changes in cognitive and behavioral characteristics of individuals. The prospective panel design is here illustrated by Heather Joshi, using examples drawn from British longitudinal cohort studies, in Chapter 5. For this design, the techniques presented in Sections III–VI of this volume, but not Section VII, are generally appropriate.

The analytical methods in Sections III–VI are also appropriate for the analysis of *retrospective* panel designs, in which data collection may occur only once, at a single period, but the data are collected for two or more periods (prior to or during the period in which the data are being collected). In retrospective panel designs, there may be sampling bias as a result of excluding respondents who have died by the last period *for* which the data are collected (or by the time *at* which the data are collected), or from whom data would have otherwise been available for earlier periods but not for the last period. In both retrospective and prospective panel designs, missing data may result from failure of the respondent to remember past events, behaviors, or attitudes, or from unwillingness by the respondent to divulge some information, and also from inability of the researcher to locate or obtain cooperation from some respondents. In principle, there need be no difference in the quality of the data obtained in prospective and retrospective panel designs, although such differences have often been observed in practice. An example of a retrospective panel design with extensive attention to potential issues of data quality is presented in Chapter 6 by Karl Ulrich Mayer, using the German Life History Study (GLHS).

As noted in Menard (2002), the designs diagrammed in Figure 1.1 are not the only possible designs for longitudinal research. It is possible, for example, to have a revolving sample in which subsamples may be dropped for one period, then re-included in the sample in a subsequent period. It is also possible to have a panel design in which cases are dropped, without replacement, after they meet some criterion (e.g., age 21). This latter design would result in a monotonically decreasing sample size which could pose problems for analysis of data from later years of the study (unless the design were further modified by replenishing the sample with new respondents from younger cohorts). The general considerations associated with the various designs for data collection do not change, however, with modifications of the four designs presented in Figure 1.1, and variations on these basic designs must be evaluated in terms of their adequacy for describing short- and long-term historical trends (period effects), intercohort and intracohort developmental changes (age effects), separating age, period, and cohort effects, and ascertaining not only the strength but also the direction of causal influences. Total population designs can, in principle, be used for practically any type of longitudinal analysis, given a sufficient number of cases and measurement periods. Other designs are more limited, and their appropriateness must be judged in the context of a particular research problem.

3 Measurement issues in longitudinal research

Longitudinal research is subject to all of the concerns about measurement that arise in cross-sectional research, plus some issues with particular relevance to longitudinal research. Put another way, longitudinal research has all of the problems of cross-sectional research, plus a few more. In the second section of this handbook, the focus is on those issues most specifically relevant to longitudinal research. Skipping ahead for a moment, in Chapter 9 Toon W. Taris discusses reliability issues in longitudinal research. Taris examines issues of distinguishing unreliability from true change, and raises (not for the last time in this volume) the issue of the reliability of change scores as measures of change. This is followed in Chapter 10 by Patterson's discussion of one of the challenging issues in long-term longitudinal research on individual change, the possibility that it may be appropriate to operationalize the same concept in different ways across the life course. The issue here is that, on one hand, whenever we change the way we measure a concept in longitudinal research, if there appears to be a change, we cannot be certain whether the change results from change in the concept we are trying to measure, or change in the measurement of the concept. Yet for research on individuals over the life course, the same measurement at different stages of the life course may not be validly measuring the same concept because different measures are appropriate at different ages. In much of longitudinal research, there is an emphasis on consistency of measurement, avoiding changes in how a concept is measured because otherwise we cannot tell whether an apparent change represents a true change in the underlying concept or merely in the measurement itself. As Patterson explains, however, using the same operationalization of the same concept over the life course may not always be the best approach, and the same underlying concept may manifest itself, and thus need to be measured, in different ways at different stages of the life course. Taken together, the chapters by Taris and Peterson address the issue of distinguishing true change and stability from measurement effects that mimic change in longitudinal research.

The chapters by Taris and Patterson apply to longitudinal research in general, whether measurement is done prospectively or retrospectively. The remaining chapters deal with issues more specific to different types of longitudinal research. Jennifer Grotpeter in Chapter 7 provides a general conceptual framework for understanding long-term retrospective recall, and examines the results of studies of recall as it is related to the length of the recall period. On this topic, see also Chapter 6 on the (retrospective) German Life History Study in the previous section, in which Karl Ulrich Mayer describes the techniques (and their results) used to enhance recall in a major retrospective panel study. Chapter 8 by David Cantor examines an issue specific to prospective longitudinal research, the effect of panel conditioning in panel research. Panel conditioning potentially occurs when respondents react to previous experience of participating in the study by changing their behavior or answers, possibly in response to their perceptions of what the researcher is seeking, or possibly to reduce their own burden as respondents. Consideration of issues specific to prospective longitudinal panel research continues in Chapter 11 by Heather Laurie, who discusses procedures for minimizing panel attrition in longitudinal samples. Despite our best attempts to minimize panel attrition, however, circumstances beyond our control (and sometimes beyond the control of our respondents) may result in missing data in longitudinal designs. In Chapter 12, E. Michael Foster and Anna Krivelyova present a brief discussion of different types of missing data, along with an example of how to handle nonignorable nonresponse in longitudinal research designs.

4 Descriptive and causal analysis in longitudinal research

The first stage in the process of analyzing longitudinal data is to provide a basic description of the data. The chapters in Section III present issues and techniques which cut across different types of longitudinal research designs. In Chapter 13, Garrett Fitzmaurice describes graphical techniques for presenting longitudinal data. Fitzmaurice shows how exploratory graphical techniques in longitudinal research help in providing insights prior to estimation of the model, and are also useful in the post-estimation diagnostic phase for examining residuals. In Chapter 14, I review the distinction between historical and developmental change and the issues involved in separating the two, with special attention to the disentangling of age, period, and cohort effects. In Chapter 15, John L. Worrall provides an introduction to pooling cross-sectional and time series data, a topic which will recur in other chapters in this handbook. In Chapter 16, Ronald Schoenberg describes the consequences of dynamic misspecification in the use of cross-sectional data to model dynamic processes. Schoenberg's chapter indicates the conditions under which cross-sectional data may be adequate to model dynamic processes, and indicates the consequences of using cross-sectional data when those conditions are not met. In Chapter 17, David Greenberg reviews attempts to draw causal inferences from nonexperimental panel data, tracing the evolution of causal inference in longitudinal research from some of the earliest methodological attempts to more contemporary approaches. Jos W. R. Twisk in Chapter 18 provides a parallel consideration of techniques for drawing causal inferences in longitudinal experimental research.

5 Description and measurement of qualitative change

The definitions of qualitative data and qualitative change may be approached from different perspectives, including how the data were collected, and at what level of measurement (nominal or at most ordinal for qualitative data). While consideration of qualitative data is not excluded from Section III, the focus is on techniques for the presentation and analysis of quantitative data. Section IV begins with Chapter 19, in which Johnny Saldaña describes an approach to the description and measurement of qualitative change in qualitative observational research. Saldaña offers a systematic approach to organizing and analyzing data from qualitative research with an emphasis on tracing patterns of change in qualitative data. Turning from qualitative defined in terms of method to qualitative defined in terms of level of measurement, Alexander von Eye and Eun Young Mun in Chapter 20 describe the use of configural frequency analysis for describing and analyzing qualitative change in longitudinal data. In configural frequency analysis, the emphasis is on tracing change in nominal variables across multiple measurement periods to identify normative and exceptional patterns of change. In Chapter 21, Catrien C. J. H. Bijleveld describes the use of optimal scaling techniques, typically calculated using alternating least squares (ALS) estimation, as a way of "quantifying" qualitative variables, and the applications of optimal scaling to the study of change in longitudinal research. The approaches described by Saldaña, von Eye and Mun, and Bijleveld are perhaps less well known, and typically less well covered, than other techniques for longitudinal data analysis. More widespread at present, at least in the social and behavioral sciences, is the use of latent class analysis to identify different qualitative "types" of individuals or of patterns of behavioral or attitudinal change over time. In a companion pair of chapters, C. Mitchell Dayton in Chapter 22 provides an introduction to latent class analysis, and Jeroen Vermunt, Bac Tran, and Jay Magidson in Chapter 23 describe the application of latent class models in longitudinal research. Taken together, the chapters in Section III offer an array of options for the analysis of data that are qualitative in terms of the research design, the level of measurement, and the assignment of cases to latent qualitative classes in longitudinal research.

6 Timing of qualitative change: event history analysis

Event history analysis is not so much a single technique as a set of related techniques for describing, analyzing, and predicting the timing of qualitative change (including whether it occurs at all). Section V begins with Chapter 24 by C.M. Suchindran, in which the most basic models for event history analysis, life table models for change, are described. These models make minimal distributional assumptions, and hence can be described as distribution free or nonparametric methods. In Chapter 25, Janet M. Box-Steffensmeier and Lyndsey Stanfill describe the Cox proportional hazards model, a semiparametric technique for event history analysis. Parametric event history analysis is briefly described and illustrated by Hee-Jong Joo in Chapter 26. The proportional hazards and parametric event history analysis models both assume that measurements occur fairly continuously in time. This, however, is not the case in much social science research, which may consist of measurements separated by a year or more. For these longer measurement intervals, discrete time event history analysis, as described by Margaret K. Keiley, Nina C. Martin, Janet Canino, Judith D. Singer, and John B. Willett, allows for the occurrence of many events within a single discrete time period. Like parametric event history analysis, discrete time event history analysis makes certain distributional assumptions regarding the parameters in the model. In contrast to the continuous time parametric and semiparametric approaches, discrete time event history analysis works more easily with time-varying covariates and with multiple events occurring in a single time interval, and it can be implemented using ordinary logistic regression or related (e.g., complementary log-log regression) techniques.

7 Panel analysis, structural equation models, and multilevel models

The statistical techniques in Section VI are techniques primarily oriented to the analysis of longitudinal panel data, and would probably be considered by some to be the most mainstream longitudinal analysis methods. The section begins with a discussion by Joseph M. Hilbe and James W. Hardin in Chapter 28 of the generalized estimating equation (GEE) approach to the analysis of longitudinal data. The use of GEE involves the estimation of parameters and standard errors that avoids unrealistic assumptions of independence of observations in longitudinal analysis and adjusts for the dependencies in the data. In Chapter 29, Steven E. Finkel describes approaches to linear panel analysis with quantitative (interval and ratio scaled) outcome variables, and in the following chapter, Chapter 30, I describe the use of linear panel analysis for the analysis of categorical (dichotomous, polytomous nominal, and polytomous ordinal) dependent variables, including the critical issue of how to measure and model change in categorical variables in linear panel models. Taken together with the chapters by Worrall (15), Greenberg (17), Twisk (18), and Hilbe and Hardin (28), these chapters provide an overview of the analysis of short-term quantitative and qualitative change and causal inferences, in which the specific nature of the trajectories or patterns of change is typically not itself being modeled.

The next three chapters turn to the modeling of trajectories of change, usually over the relatively short term, but potentially involving long-term trajectories as well. Michael Stoolmiller in Chapter 31 describes the latent growth curve modeling technique, based on structural equation modeling techniques. Latent growth curve models view trajectories or patterns of change over time as unobserved variables to be treated as latent variables in structural equation modeling. In contrast, in multilevel growth curve analysis of quantitative outcomes, as described by Douglas A. Luke in Chapter 32, one typically attempts to fit a manifest (not latent) polynomial or other function to the data to describe the trajectory of individual cases over time, and to explain variations in those trajectories using a combination of time-invariant individual case characteristics and time-varying covariates. When the dependent variable in the multilevel analysis is categorical rather than quantitative, it may be appropriate to speak not of "growth" curve analysis, but of multilevel change analysis. My focus in Chapter 33 is on showing the application of the logistic regression framework to the multilevel analysis of change, and on highlighting some of the contrasts of multilevel change analysis for categorical dependent variables from multilevel growth curve models for quantitative dependent variables, from event history analysis, and from linear and logistic regression panel analysis.

8 Time series analysis and deterministic dynamic models

Time series analysis stands out from the other methods of analysis in this handbook in the number of cases and the number of time periods. Most often, time series analysis is applied to aggregated data for a single case (a nation, city, corporation, or other aggregate entity, not an individual), or perhaps a handful of such cases, typically analyzed separately rather than together, as in other methods covered in this handbook, and the number of time periods is typically large, often over 100. In Chapter 34 I provide a brief introduction to time series analysis from the perspective of longitudinal research, which is a little different from the perspective out of which time series analysis itself has grown. Here, time series analysis is viewed as one tool for longitudinal research, with more of a focus on description and explanation and less of a focus on forecasting than is typical in the mainstream time series analysis literature. In Chapter 35, William W. S. Wei provides an introduction to spectral analysis, the most mathematically demanding of the time series analysis approaches. In Chapter 36, David Sanders and Hugh Ward provide further details on alternative approaches to time series analysis with one or more predictors included in the model, and offer a useful comparison of the different approaches to time series analysis to the empirical study of public opinion in political science.

The final two chapters in Section VII also involve a higher level of mathematical sophistication than most of the other chapters in this handbook. Steven M. Boker in Chapter 37 describes the application and estimation of differential equation models in longitudinal research using a latent variable structural equation modeling approach to estimate the parameters of the differential equation model. Finally, Courtney Brown in Chapter 38 provides a brief introduction to the application of nonlinear dynamics, chaos, and catastrophe theory to the study of change.

9 Conclusion

One of the goals of this handbook is to make the reader aware of the richness and breadth of research design and analytical techniques available for longitudinal research. The first section of this handbook begins with strong examples of each of the major types of longitudinal research design. Section II focuses on measurement issues that arise in longitudinal research generally, and also more specifically in particular types of longitudinal research designs. With each of these designs, the number of cases and periods may vary, and as a result of this variation, different methods of analysis may be appropriate. The number of cases is in principle independent of the type of design. In a total population design, for example, at the individual level, the total population of a tribal society may number fewer than 100. In aggregate analysis, a cohort or a population, rather than its individual members, may be the unit of analysis, and the number of these aggregate units may be small. At the other end of the continuum, the revolving sample in the National Crime Victimization Survey includes over 100,000 individuals from 60,000 households. All of these possible combinations of type of design and number of cases are included within the broad category of longitudinal research.

The number of cases and the number of time periods, in turn, drives the choice of analytical methods. With no more than a handful of cases but many time periods, the time series and deterministic dynamic models in Section VII are most appropriate. With no more than a handful of time periods but many cases, panel analytic techniques described in Section VI may be best, and as the number of time periods increases up to ten or so, techniques such as latent and multilevel growth curve and change models in Section VI, event history analysis in Section V, and the techniques for qualitative data analysis in Section IV become increasingly feasible. In the best of all possible worlds for longitudinal research, many cases and many time periods, event history and multilevel growth curve and change models seem at present to offer the best options. It is hoped that, by presenting in some detail the different designs for longitudinal research, issues in longitudinal research design, and techniques of analysis for longitudinal data, all in a single sourcebook, readers will be increasingly aware of and better able to make informed selections among the different options available to best capitalize on the strengths of longitudinal research.

References

Ahluwalia, M. S. (1974). Income inequality: Some dimensions of the problem. In H. B. Chenery, M. S. Ahluwalia, C. L. G. Bell, J. H. Duloy and R. Jolly (eds), *Redistribution with Growth: An Approach to Policy* pp. 3–37. Oxford, UK: Oxford University Press.

Ahluwalia, M. S. (1976). Inequality, poverty, and development. *Journal of Development Economics*, 2: 307–342.

Davis, J. A. and Smith, T. W. (1992). *The NORC General Social Survey: A User's Guide*. Newbury Park, CA: Sage.

Hogan, H. and Robinson, G. (2000). *What the Census Bureau's Coverage Evaluation Programs Tell Us About the Differential Undercount*. Washington, DC: US Bureau of the Census.

Menard, S. (2002). *Longitudinal Research*. Thousand Oaks, CA: Sage.

Menard, S. and Elliott, D. S. (1990). Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency. *Justice Quarterly*, 2: 11–55.

Robey, B. (1989). Two hundred years and counting: The 1990 census. *Population Bulletin*, 4(1): 1–43.

Tolnay, S. E. and Christenson, R. L. (1984). The effects of social setting and family planning programs on recent fertility declines in developing countries: A reassessment. *Sociology and Social Research*, 62: 72–89.