

GESIS Survey Guidelines

Why Panel Surveys?

Klaus Pforr & Jette Schröder

Abstract

This paper provides an overview of the methodological advantages and disadvantages of panel surveys and of the fundamental considerations when deciding for or against a panel design.

Citation

Pfarr, K., & Schröder, J. (2016). Why Panel Surveys? *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_008

This work is licensed under a Creative Commons Attribution – NonCommercial 4.0 International License (CC BY-NC).



1. Advantages of Panel Surveys

As a survey design in which the units of observation are interviewed on several occasions over time, panel surveys have a long tradition in the social sciences (Lazarsfeld, Berelson, & Gaudet, 1944; Lazarsfeld & Fiske, 1938). Panel surveys can therefore yield longitudinal individual-level data, which, compared to cross-sectional data, have great advantages for data analysis (Allison, 2009; Brüderl, 2010; Giesselmann & Windzio, 2014; Halaby, 2004).

When it comes to answering social science questions, longitudinal individual-level data have several strengths. Based on the objective of hypothesis testing, a necessary precondition for the presence of a causal relationship is that the independent variable must predate the dependent variable. However, when conducting an analysis with cross-sectional data, it is not possible to empirically check whether this condition is met. The temporal structure between the independent and dependent variables must therefore be assumed or theoretically established (Opp, 2010). Longitudinal data, by contrast, can ensure that the assumed temporal structure between these variables actually exists. One special potential of panel data for social science analysis comes into play when respondents are repeatedly asked the same questions over time. At the descriptive level, this enables changes and processes of development in the units of observation to be represented over time. For example, if one uses cross-sectional data to represent women's participation in the labour force by age, it remains unclear whether the fact that female labour force participation decreases with age is due to a cohort effect – in other words, whether the labour force participation of the older cohorts of women is, in principle, lower than that of the younger cohorts – or whether it is a life-cycle effect – in other words, whether female labour force participation decreases with age (Andreß, Golsch, & Schmidt, 2013, p. 5).

However, decisive advantages of panel data arise, in particular, when it comes to testing hypotheses by means of methods of multivariate analysis. The results of analyses with cross-sectional data are biased when factors that correlate with the independent variable and the error term are not controlled. Because, all of these factors are not usually measured – or cannot usually be measured – there is a great risk that cross-sectional analysis will suffer from unobserved heterogeneity or from endogeneity (Arellano, 2003; Baltagi, 2005; Cameron & Trivedi, 2009; Cameron & Trivedi, 2010; Hsiao, 2003; Mátyás & Sevestre, 1996; Wooldridge, 2010). For example, if cross-sectional analyses reveal that children have a negative effect on women's income, this link may be due to the fact that women who have children may be less career-oriented anyway than those who do not. If career orientation is not controlled in the cross-sectional analysis, the negative effect of children on income will be overestimated. However, by using longitudinal data to investigate this effect, it is possible to at least consistently estimate it by means of fixed effects regression if the unobserved factors are constant over time. In contrast to the cross-sectional analysis, one does not compare women who have children with women who do not (between-subjects comparison), but rather one compares the income of the individual women before and after the birth of children (within-subjects comparison).

Although, longitudinal data can, in principle, be collected in cross-sectional surveys by means of retrospective questions, this is possible only to a limited extent, because retrospective questions are beset by serious measurement errors. Even when information on past events is collected, the quality of the data suffers to a varying extent depending on the subject matter. This is because, for example, respondents suppress the recollection of undesirable events, forget or incorrectly date subjectively unimportant or rare events, or use current circumstances as a biasing anchor (Belli, Stafford, & Alwin, 2009). In the case of subjective characteristics, such as attitudes, competencies, or personality traits, this measurement error is so large that measurement by means of retrospective questions is impossible.

Examples of German panel surveys that are aimed at improving the estimation of causal effects by collecting suitable longitudinal data in the manner presented above include the German Socio-Economic Panel (SOEP), the German Family Panel (pairfam; Panel Analysis of Intimate Relationships and Family Dynamics), the Survey of Health, Ageing and Retirement (SHARE), the Panel Study Labour Market and Social Security (PASS), the National Educational Panel Study (NEPS), the Microcensus Panel, and the National Cohorts. A central feature of these studies is the repeated measurement of the same variables or the measurement of variables that predate the dependent variables.

Besides the better analytical potential compared to cross-sectional studies, there are, however, other reasons for choosing a panel design. One possible reason might be that the amount of different information required from the same individuals is so large that it could not be gathered in a single survey interview. To reduce the burden on respondents, the content can be distributed over several survey waves. However, careful thought should be given in advance to whether the fact that some variables are collected at an earlier point in time than others is problematic for the envisaged (cross-sectional) analyses. This applies, in particular, in the case of time-variant variables.

Another reason for using a panel design is to enable researchers to have quick and cost-effective access to respondents. It can be cheaper to interview a sample of persons repeatedly over time instead of having to repeatedly draw new samples from the same population. In addition, an existing panel enables researchers to react much faster to current political and social issues than would be the case if they first had to draw a sample of respondents before conducting the survey. The main focus of panels that are established for these reasons is on the collection of cross-sectional data rather than on the repeated measurement of the same variables for the purpose of improving causal analysis. Studies that adopt this approach are often known as access panels (Engel, Bartsch, Schnabel, & Vehre, 2012).

2. Disadvantages of Panel Surveys

In addition to the above-mentioned advantages of longitudinal data, panel surveys also present a number of methodological problems. On the one hand, panel attrition may occur selectively – that is, over time, the sample becomes less representative of the population (Kasprzyk, Duncan, Kalton, & Singh, 1989; Lynn, 2009; Menard, 2008). Depending on the analysis plan, this problem may be more or less serious. It can be reduced, first, by measures taken while the survey is being conducted, for example panel care measures, interviewer training, contact- and refusal-conversion procedures, and incentives. Second, the problem can be addressed after the survey has been conducted. Examples of possible measures include weighting, Heckman models that explicitly model attrition, and imputation procedures. However, these measures presuppose that the influencing variables that reflect the selective attrition are also collected.

A further distinctive methodological feature of panel studies is the fact that interviews in earlier waves may influence interviews in subsequent waves (panel conditioning; Warren & Halpern-Manners, 2012). On the one hand, the actual value of a variable may change as a result of an interview because the behaviour or the attitude of the respondent is influenced by the interview. Another variant of panel conditioning is that, although the true value of a variable does not change, response behaviour does. Research has shown that, depending on the subject matter and the question, panel conditioning may, but need not, occur. There are few findings and hypotheses relating to the type of questions associated with panel conditioning; established approaches to solving this problem are lacking.

3. Deciding for or against a Panel Design

As outlined above, apart from experiments, panel surveys are, as a rule, superior to cross-sectional surveys when it comes to causal analysis. If, on the other hand, an analysis requires only cross-sectional data, as is the case, for example, with descriptive analyses of the distribution of certain characteristics in the population, cross-sectional surveys should be given preference over panel surveys. In principle, the same applies to trend analyses in which changes in the distribution of characteristics over time are examined. In this case, trend data from several cross-sectional surveys should be preferred over panel data from panel surveys. In both cases, panel data do not yield any analytical gains, but, as mentioned above, they are associated with possible losses of data quality through panel attrition and panel conditioning. Furthermore, in the case of long-running panels, a sample that had the same composition as the population at the time it was drawn may no longer reflect the current composition of the population – as a result of migration, for example.

In the light of the foregoing, whether the advantages of a panel design outweigh the disadvantages, or vice versa, depends on the focus of the panel survey. If the aim of the panel is to achieve an improved analytical potential by collecting longitudinal individual-level data, the advantages and disadvantages of this design should be compared with those of a (possibly retrospective) cross-sectional survey or a trend survey that collects data for the same analytical purpose. Because of the additional costs of repeated interviews and panel care measures, a panel survey is more expensive and involves additional time expenditure and waiting time. On the other hand, however, it yields better data for testing causal relationships or descriptively representing individual developments over time. Both the advantages in terms of data analysis associated with the panel design and the costs or feasibility of a panel survey vary greatly, depending on the research question. Accordingly, the advantages and disadvantages compared to alternative survey plans must always be weighed up on a case-by-case basis. However, if – to take another extreme example – one is planning an access panel to collect different cross-sectional data in repeated interviews, this design should be compared with the alternative of conducting repeated cross-sectional surveys and drawing a separate sample each time. Compared to repeated cross-sectional surveys, cost savings may possibly be achieved in the case of an access panel because the sample is drawn only once. Moreover, an access panel enables researchers to react faster to current social and political topics and to conduct a survey at short notice. This must be weighed up against possible impairments of data quality due to selective panel attrition or panel conditioning. Here, too, the outcome of these deliberations depends on the individual case.

References

- Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles: SAGE.
- Andreß, H.-J., Golsch, K., & Schmidt, A. S. (2013). *Applied panel data analysis for economic and social surveys*. Berlin, London: Springer.
- Arellano, M. (2003). *Panel data econometrics*. Oxford, New York, NY: Oxford University Press.
- Baltagi, B. H. (2005). *Econometric analysis of panel data* (3rd ed.). Chichester, Hoboken, NJ: John Wiley & Sons.
- Belli, R. F., Stafford, F. P., & Alwin, D. (Eds.). (2009). *Calendar and time diary methods in life course research*. Thousand Oaks, CA: SAGE.
- Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 963–994). Wiesbaden: VS Verlag.

- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics: Methods and applications*. Cambridge, New York, NY: Cambridge University Press.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata*. College Station, TX: Stata Press.
- Engel, U., Bartsch, S., Schnabel, C., & Vehre, H. (2012). *Wissenschaftliche Umfragen: Methoden und Fehlerquellen*. Frankfurt am Main: Campus.
- Giesselmann, M., & Windzio, M. (2014). Paneldaten in der Soziologie: Fixed Effects Paradigma und empirische Praxis in Panelregression und Ereignisanalyse. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1), 95–113. doi:10.1007/s11577-013-0248-z
- Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30, 507–544.
- Hsiao, C. (2003). *Analysis of panel data* (2nd ed.). Cambridge, New York, NY: Cambridge University Press.
- Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M. P. (Eds.). (1989). *Panel surveys*. New York, NY: Wiley.
- Lazarsfeld, P., & Fiske, M. (1938). The "panel" as a new tool for measuring opinion. *Public Opinion Quarterly*, 2(4), 596–612.
- Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential campaign*. New York, NY: Duell, Sloan and Pearce.
- Lynn, P. (Ed.). (2009). *Methodology of longitudinal surveys*. Chichester: Wiley.
- Mátyás, L., & Sevestre, P. (Eds.). (1996). *The econometrics of panel data: Handbook of theory and applications*. Dordrecht: Kluwer Academic.
- Menard, S. (Ed.). (2008). *Handbook of longitudinal research: Design, measurement, and analysis*. Amsterdam, Boston, MA: Elsevier.
- Opp, K.-D. (2010). Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 9–38). Wiesbaden: VS Verlag.
- Warren, J. R., & Halpern-Manners, A. (2012). Panel Conditioning in Longitudinal Social Science Surveys. *Sociological Methods & Research*, 41(4), 491–534. doi:10.1177/0049124112460374
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.