



UN/POP/EGM-CPD49/2015/05

ENGLISH ONLY

UNITED NATIONS EXPERT GROUP MEETING ON STRENGTHENING THE DEMOGRAPHIC EVIDENCE
BASE FOR THE POST-2015 DEVELOPMENT AGENDA
Population Division
Department of Economic and Social Affairs
United Nations Secretariat
New York
5-6 October 2015

INTERNATIONAL LONGITUDINAL DATA COLLECTION¹

*William G. Axinn**

¹The opinions expressed in this paper are those of the author and do not necessarily reflect those of the United Nations or its Member States. This paper is reproduced as submitted by the author. This paper is being reproduced without formal editing.

* Institute for Social Research, University of Michigan.

International Longitudinal Data Collection

William G. Axinn, Institute for Social Research, University of Michigan

A. INTRODUCTION

This note addresses the potential of international longitudinal data to assess development needs and evaluate programmes aimed at meeting high priority development targets. It briefly describes some of the key benefits of individual-level longitudinal data. It will also describe key issues in achieving high quality longitudinal data. These include standardized quality control, minimizing attrition, and use of responsive design to maximize efficiency.

Individual-level longitudinal data have proven to be extremely powerful resources for science and policy related to human needs, but they are expensive to collect. Because of the effort and expense involved in longitudinal studies it makes good sense to use them carefully, to accomplish clear and specific goals, and to carefully integrate longitudinal studies with other forms of data including censuses, household surveys, registration systems and social media data.

B. BENEFITS OF INDIVIDUAL-LEVEL LONGITUDINAL DATA

Longitudinal studies following the same individuals over significant periods of time have proved so valuable to the construction of effective public policy and design of effective public programs that many wealthy governments have invested heavily in the creation of longitudinal studies across many substantive areas. It is likely governing organizations, such as the United Nations, will have demand for similar studies of poor countries—countries unable to afford such data resources themselves—for the construction of policies and programs to effectively attain key development goals. Longitudinal studies are now widespread throughout the globe, including in poor country settings with weak infrastructure.

1. *Spells of adversity*

The United States of America *Panel Study of Income Dynamics* (PSID) exemplifies the tremendous scientific and public policy value of individual-level and household-level longitudinal data. The PSID was launched in the 1960s to try to better understand the dynamics of poverty and income processes over time in both poor households and a nationally representative sample of all United States households. By following the same individuals and household over time, the PSID was able to document that poverty was not simply a condition of a person or a family, but rather a temporary situation through which many persons or families passed, entering into or exit in poverty, or remains poor. These descriptive findings were so important to the construction of successful public policies and programmes that a number of other rich countries launched their own versions of the PSID, including Canada, Germany, United Kingdom, Switzerland, Korea and Australia.

Of course, the descriptive trajectory of a condition or adverse situation was so powerful and important that governments have invested in such studies across many different topics. In the United States of America, examples of long-term or repeated longitudinal studies come from many topic areas including education, labour force participation, adolescent health, elderly health and well-being, families and households, and substance use. Just as with exit and entry into poverty, understanding entry and exit into depression, anxiety or alcohol use are necessary to design effective programmes and policies.

2. Consequences of change

Longitudinal studies of individuals have become one of the most powerful tools in the social sciences for evaluating cause and consequence. Whenever possible, social scientists generally mount random control trials (RCTs) to evaluate cause and consequence. Such trials immediately imply longitudinal data collection—measures of the same individual both before and after the intervention being evaluated. But in many situations randomization is either impossible or not easily feasible. Even then, longitudinal measurement that tracks the timing of changes in individuals relative to the timing of changes in conditions within which those individuals live their lives have become the most powerful tool for assessing the potential of cause and consequence (Axinn and Pearce, 2006). Perhaps most important for the SDG agenda, longitudinal data are frequently used to assess relative consequences of multiple interventions.

Individual-level longitudinal studies have been used for this purpose worldwide, including in rural Asia, Africa and Latin America. For example, data from the Chitwan Valley Family Study (CVFS) in Nepal reveal many important consequences of programmes designed to help poor people, including the independence and relative size of these consequences. CVFS data have been used to document that placement of new schools is more strongly associated with change in contraceptive use to avert births than placement of new health services that provide contraception (Axinn and Barber, 2001). In another example, CVFS data have been used to document that child immunization programs are strongly associated with the subsequent use of contraception in a way that is independent of associations with the availability of contraceptive methods (Brauner-Otto, Axinn and Ghimire, 2007). Even from this single longitudinal study there are dozens of other examples of results that document the likely social and economic consequences of specific interventions—knowledge needed to create effective programs to achieve development goals.

3. Understanding the role of Migration

Finally, longitudinal studies are particularly effective tools for understanding the role of migration in producing change as well as the consequences of ignoring migration for the creation of bias in local area estimate. By following individuals over time, even when they move from one place to another and interviewing them about the changes in their life we can better describe the changes individuals experience when they migrate. Likewise, by interviewing family members over time, both before and after an individual in the family migrates, we can better describe the changes those who stay in one place experience when a member of the family migrates.

C. ACHIEVING HIGH QUALITY LONGITUDINAL DATA THROUGH STANDARDIZED QUALITY CONTROL

A crucial issue in the success of longitudinal studies is the creation of high quality data in the baseline study used to launch the longitudinal study and standardization of this quality control across time and across sites. Step one in the creation of high quality baseline data is careful scientific selection of individuals, households, or other cases for the study. The science of statistical sampling provides many tools to apply to this step—tools that can be useful no matter what type of data one intends to collect.

For interview-based data—census, surveys, or semi-structured interviewing—computer assisted interviewing (CAI) has revolutionized the science of quality control. The CAI approach has many advantages, especially in speeding up access to data and enhancing its quality. Conducting multiple data collections per country over time requires enormous effort without CAI in terms of manual data entry, data cleaning operations, verification and reconciliation with respondents, creating data sets and harmonizing data. By automating these steps through CAI, those collecting data can ‘frontload’ some of

this effort into programming CAI instruments to perform consistency checks during the interviews. Moreover, data from the interviews can immediately flow to central servers, giving analysts immediate access to new fieldwork data.

In international data collection, these qualities of CAI can be particularly important. Simultaneous, uniform quality control across multiple sites is one of the most efficient ways to create comparable data records from multiple sites. Electronic interview records also greatly facilitate re-interview of individuals who move, making this tool especially important for longitudinal study designed to track migrants.

Finally, the “paradata” (data about the data collection processes) these tools create are highly valuable for both quality control and implementation of responsive survey design tools. The availability and analysis of paradata has revolutionized the science of quality control in survey data collection (Couper and Lyberg, 2005; Couper, 2009). These paradata also provide the key means of exercising responsive survey design to increase the quality and efficiency of survey data collection. Responsive survey design is described in more detail below.

To provide uniform quality control across multiple data collection sites and rapid harmonization and analysis of the data, use a high quality CAI software, such as Blaise—the most flexible, robust and safe CAI software currently available. Michigan has used Blaise successfully worldwide for more than 15 years (including China, Ghana, Nepal and Saudi Arabia). Producers and users of this software are constantly innovating to increase the ability to collect high quality longitudinal measures across multiple family members and substantive domains spanning behaviors, beliefs and feelings.

D. CREATING HIGH QUALITY LONGITUDINAL DATA BY MINIMIZING ATTRITION

Attrition out of a longitudinal study is a major threat to success of such a study (Groves and Couper, 1998; Lepkowski and Couper, 2002; Groves, 2006; Couper and Ofstedal, 2009; Cobben and Bethlehem, 2009; Schoeni and others, 2013; Schouten, Trappmann, Gramlich and Mosthaf, 2015). Individuals who either cannot be relocated or who refuse to participate are rarely selected through pure randomization. Typically, those who are lost are selected on criteria associated with one or more of the objectives of the study. Migration is the most obvious example—if those who move are lost from the study it becomes impossible to provide unbiased estimates of associations that are somehow shaped by migration.

1. *Excellent Re-contact Information*

Collection of excellent re-contact information during the baseline study is a crucial tool for minimizing attrition from a longitudinal study. A small number of key principles help to provide strong re-contact information:

- Multiple modes of contact information (address, phone number, email, other);
- Multiple people likely to maintain contact (family members best; closest possible and multiple);
- Multiple sources of unique identification (name, ID number, employer, or other; biometrics likely to be most successful).

2. *Steps for Minimizing Attrition.*

Successful longitudinal studies demonstrate specific steps that can be used to minimize attrition. They include:

- Controlling the length of interviews to reduce burden;
- Re-contact at short intervals, between rounds (include information about how the study is used);
- Re-contact with multiple family members;
- Change mode of contact (e.g. face-to-face to phone);
- Continuing contact with those temporarily away no matter where;
- Track across long distances, including borders (representative subsamples of movers, if necessary).

These steps can greatly reduce attrition across all different types of longitudinal surveys. Some special situations require special steps. The literature on reducing attrition in longitudinal studies offers many specialized steps for special situations. Because armed conflict is a major disruption in efforts toward SDG outcomes, situations of armed conflict are an important example (Axinn, Ghimire and Williams, 2012).

E. TOOLS FOR STANDARDIZED QUALITY CONTROL AND DATA MANAGEMENT

Computerization of survey data collection unleashed a series of technological breakthroughs that make a new science of data collection possible. As early as the 1990s, the use of computers allowed both face-to-face and telephone interview survey questionnaires to move from paper held by an interviewer to software on a computer used by an interviewer. A decade or so later, internet technologies supported the construction of computerized management tools—tools that though residing at a central location could use the internet to reach inside the computers being used by interviewers, even interviewers long distances away, to keep track of the interviewers’ work. These tools, sometimes described as “sample management systems,” created data about the data collection process, or “paradata” (Couper, 2005; Couper, 2009; Couper and Lyberg, 2005).

Creation and analysis of paradata provided a boon to data collection quality control. Analysis of interviewer key strokes and time stamps through the questionnaire provides the means for quick detection of fabrication, identification of interviewer driven errors, and identification in questionnaire problems (Couper, 2009; Kreuter, Couper and Lyberg, 2010). Paradata also gave the field of survey methodology the raw material to create Responsive Survey Design.

F. RESPONSIVE SURVEY DESIGN

Groves and Heeringa (2006) were the first to outline a conceptual framework for responsive survey design, and Wagner and others (2012) outlined responsive design features that have been successfully implemented in the United States National Survey of Family Growth (NSFG). Briefly, surveys employing responsive designs attempt to decrease the costs, errors and uncertainty associated with various aspects of survey data collection by carefully monitoring selected information summarized from

paradata, and making design decisions during data collection based on patterns evident in the paradata. Groves and Heeringa (2006) defined five steps that responsive survey designs generally follow:

1. Pre-identify a set of design features potentially affecting costs and errors of survey statistics (e.g., number of calls made to a sampled unit, or over-sampling of certain ethnic groups);
2. Identify a set of indicators of the cost and error properties of those features (e.g., costs per call attempt, response rates as a function of number of calls attempted, or response rates over time for various ethnic groups);
3. Monitor those indicators in initial phases of data collection;
4. Alter the active features of the survey in subsequent phases based on cost/error tradeoff decision rules (e.g., ask interviewers to increase their efforts for a particular ethnic group, or ensure that all cases have been called a certain number of times); and
5. Combine data from the separate design phases into a single estimator.

Responsive survey designs rely on carefully designed collection of paradata that are relevant for a given survey (i.e., predictive of key survey outcomes, including response to the survey request). Using these paradata, statistical models can be used to both evaluate the success of data collection up to any given point and predict the actions most likely to produce effective data collection going forward. The collection and analysis of paradata is an extremely active area of research in survey methodology (Kreuter, 2013; Luiten and Schouten, 2013; Lundquist and Särndal, 2011; Peytchev, Baxter and Carley-Baxter, 2009; Schouten and others, 2012; Schouten, Calinescu and Luiten; 2013; Wagner, 2008) and models for optimal use of paradata to improve efficiency are evolving rapidly.

The successful implementation of responsive survey design strategies in the United States National Survey of Family Growth (NSFG) demonstrates the promise of this approach. Initial aspects of the responsive design framework were successfully implemented in the sixth cycle of the NSFG (2002-2003) (Axinn and others, 2011; Lepkowski and others, 2006). Learning from the Cycle 6 experiences, the NSFG implementation team further incorporated responsive design ideas into the seventh cycle of the NSFG, resulting in substantial cost savings and increased data collection efficiency relative to Cycle 6 (Kirgis and Lepkowski, 2013; Lepkowski and others 2013; Wagner and others, 2012). By switching NSFG from a “one-shot” interviewing design to a continuous interviewing design and then incorporating responsive survey design on a continuous basis NSFG was able to increase the production of interviews by 80 per cent in Cycle 7 at a lower cost than Cycle 6. These same principles have also been applied successfully in small scale local area surveys.

G. DATA DISSEMINATION AND LINKING ACROSS TYPES OF DATA

Although successful longitudinal studies require substantial effort, they can be harnessed efficiently as components of a linked data collection design. While national census, administrative records, or cross-sectional surveys have an important value for characterizing a national population, more intensive studies of sub-populations can provide important, complementary benefits. Longitudinal studies are a key example. Nationally representative data from another source can be used to choose an appropriate sub-

sample of the population for more intensive longitudinal study. Such designs offer the complementary strengths of breadth and depth.

1. *Creating synergy among census, administrative records, repeated cross-sections, and longitudinal data*

To create high complementarity across multiple data sources, repeating at least some subset of measures across data types is a high priority. It is now standard practice to repeat census-style demographic measures of age, sex and race in cross-sectional surveys to assess the extent to which any specific sample matches the full census population. In a similar way, it is important to replicate key substantive measures from administrative records or national sample surveys in longitudinal studies of sub-populations to assess the extent of match across data sources.

2. *Linking dissemination of data*

The use of data collection management tools as described above as the source for paradata, can also greatly improve linking across data sources and dissemination of those data to users. In the domain of interview based data, computer assisted interviewing can create same-day analytic availability if these data collection management systems are used carefully. Well-designed data management systems not only transport data securely, they also maintain those data in carefully organized records so that they can be replicated exactly at the destination. These properties make them an ideal tool for linking data from multiple sources together. This approach holds especially high promise for linking administrative records of various sorts to either cross-sectional surveys or longitudinal studies. This type of computational system based linking can be used to speed delivery of linked data to users. Virtual data enclaves, secure data enclaves and similar tools are improving. These tools can protect the confidentiality of data and protect human subjects from disclosure risks even as sophisticated linking systems speed the availability of linked data resources to analysts.

REFERENCES

- Axinn, W.G. and J.S. Barber (2001). Mass education and fertility transition. *American Sociological Review*, vol.66, No. 4, pp. 481-505.
- Axinn, W.G., C.F. Link and R.M. Groves (2011). Responsive survey design, demographic data collection, and models of demographic behavior. *Demography*, vol. 48, No. 3, pp. 1127-1149.
- Axinn, W.G., D.J. Ghimire and N. Williams (2012). Collecting survey data during armed conflict. *Journal of Official Statistics*, vol. 28, No. 2, pp. 153-171.
- Axinn, W.G. and L.D. Pearce (2006). *Mixed Method Data Collection Strategies*. New York: Cambridge University Press.
- Brauner-Otto, S.R., W.G. Axinn and D.J. Ghimire (2007). The spread of health services and fertility transition. *Demography*, vol. 44, No. 4, pp. 747-770.
- Brick, J. M. and D. Williams (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, vol.645, No. 1, pp. 36-59.
- Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review*, vol. 23, No. 4, pp. 486-501.
- Couper, M. P. (2009). The role of paradata in measuring and reducing measurement error in surveys. *NCRM Network for Methodological Innovation*.
- Couper, M. P. and L. Lyberg (2005). The use of paradata in survey research. Proceedings of the 55th session of the International Statistical Institute, Sydney, Australia.
- Couper, M.P. and M.B. Ofstedal (2009). Keeping in contact with mobile sample members. In *Methodology of longitudinal surveys*. P. Lynn, ed. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9780470743874.ch11
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, vol. 70, No. 5, pp. 646-675.
- Groves, R.M. and M.P. Couper (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley and Sons.
- Groves, R. M. and S. G. Heeringa (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 169, No. 3, pp. 439-457.
- Groves, R.M. and E. Peytcheva (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, vol. 72, No. 2, pp. 167-189.
- Kirgis, N.G. and J.M. Lepkowski (2013). Design and management strategies for paradata-driven responsive design: Illustrations from the 2006-2010 National Survey of Family Growth. Chapter 6. In *Improving Surveys with Paradata: Analytic Uses of Process Information*. Wiley.
- Kreuter, F., ed. (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Wiley.

- Kreuter, F., M. Couper and L. Lyberg (2010). The use of paradata to monitor and manage survey data collection. In *Proceedings of the Joint Statistical Meetings, American statistical association* (pp. 282-296).
- Lepkowski, J.M. and M.P. Couper (2002). Nonresponse in the second wave of longitudinal household surveys. *Survey Nonresponse*, pp. 259-272.
- Lepkowski, J.M. and others (2006). National Survey of Family Growth, cycle 6: Sample design, weighting, imputation, and variance estimation. *Vital and Health Statistics*, vol. 2, No. 142. National Center for Health Statistics.
- _____ (2013). Responsive design, weighting, and variance estimation in the 2006-2010 National Survey of Family Growth. *Vital and Health Statistics*, vol. 2, No. 158. National Center for Health Statistics.
- Luiten, A. and B. Schouten (2013). Tailored fieldwork design to increase representative household survey response: an experiment in the Survey of Consumer Satisfaction. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, vol. 176, No. 1, pp. 169-189.
- Lundquist, P. and C.E. Särndal (2011). Aspects of responsive design for the Swedish Living Conditions Survey. *Research and Development: Methodology Reports from Statistics Sweden*. Statistics Sweden.
- Peytchev, A., R. K. Baxter and L. R. Carley-Baxter (2009). Not all survey effort is equal: reduction of nonresponse bias and nonresponse error. *Public Opinion Quarterly*, vol. 73, No. 4, pp. 785-806.
- Schoeni, R.F. and others (2013). Response rates in national panel surveys. *The ANNALS of the American Academy of Political and Social Science*, vol. 645, No. 1, pp. 60-87.
- Schouten, B., F. Cobben, and J. G. Bethlehem (2009). Indicators for the representativeness of survey response. *Survey Methodology*, vol. 35, No. 1, pp. 101-113.
- Schouten, B. and others (2012). Evaluating, comparing, monitoring and improving representativeness of survey response through R-indicators and partial R-indicators. *International Statistical Review*, vol. 80, No. 3, pp. 382-399.
- Schouten, B., M. Calinescu and A. Luiten (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology*, vol. 39, No. 1, pp. 29-58.
- Trappmann, M., T. Gramlich and A. Mosthaf (2015). The effect of events between waves on panel attrition. *Survey Research Methods*, vol. 9, No. 1, pp. 31-43.
- Wagner, J. R. (2008). Adaptive survey design to reduce nonresponse bias. Ann Arbor: ProQuest.
- Wagner, J. and others (2012). Use of paradata in a responsive design framework to manage a field data collection. *Journal of Official Statistics*, vol. 28, No. 4, pp. 477-499.