

## Made in Heaven: Secondary Data Analysis and Interdisciplinary Collaborators

Greg J. Duncan  
Institute for Social Research  
University of Michigan

Secondary data analysis and interdisciplinary collaboration have a great deal to offer developmentalists. The research traditions of economists and quantitative sociologists have produced large-scale longitudinal studies using national samples, useful statistical techniques for approximating experimental conditions and interesting theoretical insights.

The research traditions of psychologists and economists interested in human development and behavior could not be more different. Graduate students in developmental psychology are taught to collect their own data. However, students in quantitative economics, demography, and sociology learn how to purify and then, in effect, sanctify data collected by others. Secondary analysis is so valued by economics journals that some editors give explicit priority to replication studies, whereas others require that authors deposit their data in national archives. The very thought of needing to make a case for secondary data analysis is as alien to economists as is the possibility of having to collect primary data themselves.

Their research tradition has led quantitative economists and sociologists to develop a number of research strategies that complement those typically used by developmental psychologists. In this commentary, I discuss three. First, concerning data, professional support and resources for primary data collection have focused on a few large-scale longitudinal studies using national samples. One of these studies, the National Longitudinal Survey of Youth (NLSY), is highlighted in Chase-Lansdale, Mott, Brooks-Gunn, and Phillips (1991). Another, the Panel Study of Income Dynamics (PSID; Duncan et al., 1984; Hill, in press) has provided stimulation, frustration, and, above all, steady employment for me for the past two decades. I have devoted the first portion of my comments to expanding on the list of reasons provided by Brooks-Gunn, Phelps, and Elder (1991) on why developmental psychologists might consider secondary analysis of these kinds of data sets.

Second, regarding the statistical side, economists' long struggle with data collected for other purposes has led them and quantitative sociologists to approximate experimental treatments. The second portion of my comments focuses on them.

Finally, from a theoretical perspective, I argue that economists have useful insights into developmental issues, especially into understanding how parents cope with the problem of allocating time and money resources to their children.

### Data Sets

Although Chase-Lansdale et al. (1991) conducted secondary analysis of NLSY data, they probably would have preferred to use experimental data in which mothers were randomly assigned to reemployment strategies following childbirth and their children randomly assigned to different child-care arrangements. However, practical and ethical constraints prevented them and other researchers from gathering the experimental data we would most like to use.

Most developmentalists perceive the next-best alternative to running experiments themselves to be gathering data from small, purposive samples, using an instrument explicitly tailored to the research goals. With luck, the measures used will prove up to the task, but doubts will still persist about the representativeness of the sample and, for many longitudinal studies, the typically short duration of measurement. The costs of identifying target populations may preclude altogether prospective studies of rare events, such as divorce.

To quantitative economists and sociologists, the typical next-best alternative to running the experiments themselves is analyzing the data from long-term longitudinal surveys that are based on national samples. Because the multimillion-dollar cost of these studies means that limited survey time must be shared by a number of competing research goals, the data collected in them are almost never ideal for any particular analysis. However, this drawback may well be outweighed by a formidable list of benefits, some of which were spelled out by Brooks-Gunn, Phelps, and Elder (1991). To their list I would add several items.

First, if coupled with benign nonresponse, a properly drawn sample provides a number of important benefits. Not only does it ensure broad population representativeness, but it also provides insurance against the kinds of selection bias that may accompany recruitment techniques often used in purposive samples. The large samples typically drawn for national studies like the NLSY or PSID usually provide sufficient sample sizes for prospective studies of the effects of infrequent events such as divorce, unemployment, or falling into poverty. Data on these events, when combined with other data typically gathered in these surveys, constitute an unexpectedly rich and accurately

---

Correspondence concerning this article should be addressed to Greg J. Duncan, Room 3260, Institute for Social Research, University of Michigan, 426 Thompson, 1200 ISR, Ann Arbor, Michigan 48109.

measured set of covariates that can be included in the analysis of interest.

A second and largely unanticipated benefit is the ability of these data sets to investigate the effects of neighborhood contexts. Neighborhood influences are part of ecological models (Bronfenbrenner, 1989) that view individuals in the context of a series of ecosystems—nuclear family, extended family, peer group, neighborhood, community, and institutions such as the school or the work place. And yet the small and geographically clustered nature of most developmental data precludes analysis of all but the most proximal environments. The recent coupling of neighborhood-level data (e.g., extent of poverty, public assistance receipt, and middle-class neighbors) to the PSID and NLSY makes it possible to investigate the ways in which two important ecosystems—the neighborhood and the family—affect the development of children and adolescents. Brooks-Gunn, Duncan, Kato, and Sealand (1991) used the PSID and found relatively powerful neighborhood effects, particularly for the presence of affluent neighbors and single-parent families, on problem behavior of adolescent girls. Neighborhood effects on preschoolers drawn in the Infant Health and Development Program (IHDP, 1990) were smaller and more specialized.

Secondary analysis has the third advantage of providing the opportunity of replication and extension of original research findings. How much better it would have been if Cyril Burt's falsified data on sibling IQs could have been exposed at the time it was first published. More than a few dramatic research results in economics have been shown to be either erroneous or not robust when subjected to secondary analysis. Indeed, I would assert that the value of replication and extension argues for public release of all data as soon as practically possible, although few of my colleagues would go this far.

A fourth important advantage of several of the longitudinal studies is that they provide data on children, both when they are living with their parents as well as years later when they are adults in their own households. For example, the PSID contains several thousand observations on individuals who were adolescents in the early waves (during the late 1960s and early 1970s) and adults in their 30s and even 40s in the most recent waves. A great deal of information was gathered from the parents of these children during the course of the 25 years of interviews. At the same time rich information was gathered from the children once they left the parental nest. The combined matched information is well-suited for a variety of intergenerational studies.

### Statistical Considerations

Economists and quantitative sociologists have traditionally relied in their research on secondary analysis of individual- and family-based data. As explained in Brooks-Gunn, Phelps, and Elder (1991), this reliance has fostered the development of statistical methods, some of which are probably familiar to many quantitative developmentalists, that seek to extract estimates of structural parameters (i.e., causal effects) from the secondary data. The large-scale longitudinal surveys launched nearly a quarter of a century ago have stimulated great interest in fixed effect (e.g., England, Farkas, Kilbourne, & Dou, 1988) and event-history (e.g., Tuma & Hannan, 1984) models.

Not stressed in Brooks-Gunn, Phelps, and Elder (1991), however, is another useful statistical development, namely the ability to adjust for biases introduced by the selective nature of most samples. Regardless of whether effects are to be estimated for maternal employment and child-care arrangement (as in the Baydar & Brooks-Gunn, 1991, article), divorce, maternal depression, or any number of other possible treatments, nonexperimental studies must address the issue of the selection process by which the treatment group constituted itself. That process is rarely random and can easily bias estimated treatment effects.

Suppose, for example, that mothers best able to stimulate the development of their children are also most likely to return to paid employment or to choose institutional child care. If the abilities of these mothers are not captured by measured characteristics such as formal schooling, then the estimated effect of maternal employment or institutional child care will be more positive (or less negative) than if a randomly chosen mother became employed or placed her child in institutionalized care. In other words, the apparent effect of employment or child-care choice will be biased by the selective nature of the employment and child-care decisions of mothers.

Concern for these kinds of possible biases in nonexperimental data has led to the development of statistical techniques to model the selection process and adjust for the bias. When the sample itself (e.g., employed women with children) is a selective subset of the larger population of interest (e.g., all women with children), then sample selection models (e.g., Heckman, 1979; Manski, 1989) might be useful. If the selection process involves reciprocal effects (e.g., mothers whose children are developing the best are most likely to return to paid employment), then simultaneous equations modeling (e.g., Judge, Griffiths, Hill, Luetkeppl, & Lee, 1985) is appropriate. Longitudinal data are particularly useful, because panel cases can sometimes serve as their own controls (e.g., when several births are observed during the panel period, only some of which are associated with an early return to work or institutionalized child care). Although these techniques require fairly strong assumptions and do not always produce satisfying solutions, they highlight the problem of selection bias in nonexperimental data.

### Theoretical Insights

Few economists interested in developmental outcomes fail to appreciate the insights of developmental psychologists; the smartest ones recruit developmentalists as collaborators and coauthors. But it is also true that the economic approach to modeling human behavior and development, which is based on how individuals cope with limited time and money resources in production and consumption activities, is surprisingly useful for developmental problems.

To be sure, most economists prefer a newspaper's business pages to its science pages and worry about issues—relationships among national production, investment, consumption, savings, and inflation or the investment and production activities of firms and industries—far afield from concerns of developmentalists. But an economic perspective can be used to view child development as a production activity in which parental time and money resources invested in children give rise to their cognitive and emotional growth (Stafford, 1987). In other words, parents juggling work schedules and budgets to benefit

their children is analogous to a firm juggling its employment and machinery to the benefit of its balance sheet. One need not believe that parents use calculus in the same way as economic modelers use it, or in order to benefit from the analytic conceptions (e.g., human capital) used in these models or from their qualitative predictions.

This is illustrated in three brief examples: First, Desai, Chase-Lansdale, and Michael (1989) hypothesized that nonmaternal child care may substitute adequately for the time of mothers who provide a lower quality home environment but not for mothers who provide a high-quality home environment. Their empirical work appears to have confirmed this hypothesis by showing that children of highly educated mothers do worse in nonmaternal care than in maternal care.

Second, Stafford (1987) developed a model in which parental time and money inputs as well as the presence of siblings during the preschool years affected child development. His empirical work showed that time spent by parents in learning activities with preschool children was positively related to teacher evaluations of those same children 6 years later. Furthermore, he found that siblings, especially younger brothers, had a significant negative effect on those evaluations. Although neither of these findings will surprise developmentalists, the model provides interesting insights into the nature of the choices faced by parents.

Third, Gary Becker's (1981) study developed theoretical models that delved creatively into the behavior of family members. One of his most remarkable results was the "rotten kid theorem," which holds that a sufficiently benevolent parent can ensure that even perfectly selfish children will be motivated to work for the common good of the family. His broader discussion of how altruistic and selfish motives operating within a family context can explain the behavior and development of both parents and children provides fascinating, if controversial, reading for developmentalists.

In summary, secondary data analysis and interdisciplinary collaboration have a great deal to offer developmentalists. Be forewarned, however, that there are substantial costs in learning to use any of the longitudinal surveys mentioned by Brooks-Gunn, Phelps, and Elder (1991)—all the more reason to combine secondary analysis with collaboration, especially if collaborators from other disciplines are up to speed on the data and statistical methods.

### References

- Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements in infancy on preschoolers' cognitive and behavioral outcomes: Evidence from the children of NLSY. *Developmental Psychology, 27*, 932-945.
- Becker, G. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta (Ed.), *Annals of child development—Six theories of child development: Revised formulations and current issues*. Greenwich, CT: JAI Press, Inc.
- Brooks-Gunn, J., Duncan, G. J., Kato, P., & Sealand, N. (1991). *Do neighborhoods influence child and adolescent behavior?* Princeton, NJ: Mimeo, Educational Testing Service.
- Brooks-Gunn, J., Phelps, E., & Elder, G. H., Jr. (1991). Studying lives through time: Secondary data analyses in developmental psychology. *Developmental Psychology, 27*, 899-910.
- Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. A. (1991). Children of the National Longitudinal Survey of Youth: A unique research opportunity. *Developmental Psychology, 27*, 918-931.
- Desai, S., Chase-Lansdale, P. L., & Michael, R. T. (1989). Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. *Demography, 26*, 545-562.
- Duncan, G. J., Coe, R. D., Corcoran, M. E., Hill, M. S., Hoffman, S. D., & Morgan, J. N. (1984). *Years of poverty, years of plenty: The changing economic fortunes of American workers and families*. Ann Arbor, MI: Institute for Social Research.
- England, P., Farkas, G., Kilbourne, B. S., & Dou, T. (1988). Explaining occupational sex segregation and wages: Findings from a model with fixed effects. *American Sociological Review, 53*, 544-558.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*, 153-61.
- Hill, M. S. (in press). *The panel study of income dynamics*. Beverly Hills, CA: Sage.
- Infant Health and Development Program. (1990). Enhancing the outcome of low birth weight, premature infants: A multisite randomized trial. *Journal of the American Medical Association, 263*, 22:3035-3042.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Luetkepohl, H., & Lee, T. (1985). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.
- Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources, 24*, 341-360.
- Stafford, F. P. (1987). Women's work, sibling competition, and children's school performance. *American Economic Review, 77*, 972-980.
- Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Methods and models*. San Diego, CA: Academic Press.

Received July 16, 1991

Accepted July 16, 1991 ■