

# The Endogeneity Problem in Developmental Studies

Greg J. Duncan

*Northwestern University*

Katherine A. Magnuson

*Columbia University*

Jens Ludwig

*Georgetown University*

Estimates of developmental models of processes involving contextual influences (e.g., child care arrangements, divorce, parenting, neighborhood location, peers) are subject to bias if, as is often the case, the contexts are influenced by the actions of either the individuals being studied or their parents or teachers. We assessed the nature of the endogeneity biases that may result, discuss the importance of such biases in practice, and suggest possible ways of avoiding them. Our primary recommendation is that developmentalists consider reorienting their data collection strategies to take advantage of real or “natural” experiments that produce exogenous variation in family and contextual variables of interest.

Individuals’ lives are shaped by a rich set of interactive genetic, social, structural, and historical forces and processes. Consequently, developmental science places high demands on the evidence needed to separate correlation from causation. Although social science theory can commonly be invoked to limit the scope of problems and isolate key variables, a developmental perspective often does just the opposite. Because a broad theoretical perspective holds great promise for advancing researchers’ understanding of human development, developmental scientists should not be simplifying their theories for the sake of empirical tractability. Instead, they should devote themselves to ensuring that their empirical work does justice to the theory.

We focus here on the endogeneity problem, a technical name given by econometricians to a problem that is well known within developmental circles but inadequately appreciated and addressed in most empirical studies. Many important explanatory variables in developmental models of contextual influences (e.g., child care arrangements, divorce, parenting, neighborhood location, peers) are, at least in part, determined or influenced by the actions of individuals, or the parents or teachers of the individuals, whose development is being studied. Resulting correlations between developmental outcomes of interest and these determined or influenced (i.e., endogenous) contextual variables may in fact be the result of unmeasured characteristics of the individuals themselves or their parents.

Most empirical studies in development implicitly assume that the processes through which individuals select or are selected into their developmental environments are fully explained by observable characteristics, but taking the idea of agency seriously raises the obvious question: Why would two similar children experience quite different developmental environments? What would cause these children, or their parents, teachers, or other adults, to select different environments? Standard regression methods applied to nonexperimental data will yield unbiased estimates of environment impacts only if these decisions are arbitrary or completely independent of other developmental processes and influences, or at least independent of them after conditioning on (i.e., adjusting statistically for) observed characteristics. If these selection decisions are related to developmentally meaningful characteristics or events that are not controlled for in the analysis, then bias may result.

The direction of the bias is difficult to predict *ex ante*. Consider, for example, the question of how classroom resources affect child development. If the most motivated children or parents select into the highest quality classrooms, then unmeasured aspects of motivation may lead to upward bias in the estimates of the beneficial effects of classroom resources on development. If, on the other hand, school administrators assign the most at-risk children to enriched classrooms, then unmeasured child characteristics may lead researchers to understate the benefits of such classrooms. Thus, we have the dilemma addressed in this article: Most nonexperimental developmental studies of context are subject to a host of possible endogeneity biases, and although the potential for bias is often pointed out as a limitation, researchers and consumers of research have little idea of the magnitude or even direction of the bias.

We first describe the nature of the endogeneity problem, showing that in most cases it amounts to the problem of omitted (unmeasured) variables. Omitted-variable bias is not easily resolved by including additional covariates in estimation models, because theory often suggests the need to adjust for measures not included in even the most comprehensive data collection efforts. Addressing omitted-variables bias is also complicated by the possibility that covariates are sometimes determined by, rather than determinants of, the given context of interest.

Although many developmental researchers recognize this concern, much empirical work appears to implicitly assume that endogeneity biases are a second-or-

der concern. However, a growing body of empirical research we review later calls this assumption into question. A variety of studies demonstrate that nonexperimental regression and related methods are unable to come consistently “close” to replicating the results of randomized experiments for outcomes such as juvenile delinquency, school dropout, or achievement test scores, even in cases where quite rich covariate information is available. The resulting biases are sufficiently large as to be of both scientific and policy concern.

Our central recommendation is that developmental researchers take advantage of random-assignment or “natural” experiments that produce exogenous variation in family and contextual variables of interest. For example, to assess the effects of child care on children’s development, we suggest abandoning comparisons of children whose parents have self-selected into various settings in favor of comparisons between children on a waiting list who win a lottery for entrance into a child care program and children on that same list who lose the lottery. Under certain circumstances, which we describe later, comparisons of developmental outcomes of siblings, one of whom participates in a program and the other does not, also may provide a less biased method for assessing child care effects.

To assess the effects of neighborhoods on youth development, we suggest abandoning comparisons of youth in low- and high-risk neighborhoods in favor of comparisons of youth whose families have been assigned to different neighborhoods on a random or quasi-random basis. For assessing the effects of an adult mentor on youth development, we suggest abandoning comparisons of youth with and without self-selected mentors in favor of assessing differences in youth outcomes in situations in which mentors are available for some youth but not others. Our suggested methods expand opportunities for empirical work that can better meet the demands of developmental theory.

## ENDOGENEITY IN THEORY

Classical sociological and psychological formulations of the impacts of family environments and broader social structures on development and well-being viewed individuals as passive recipients of environmental influences. Theory and research focused on identifying crucial elements of social class (e.g., Davis & Havinghurst, 1946), parenting values (e.g., Kohn, 1959, 1963), and parenting styles (e.g., Baumrind, 1968), and the processes by which these factors influenced development and well-being. In essence, this approach presumed that individuals were a *tabula rasa*, and that their development was the sum total of their experiences.

Subsequent developmental theory has both challenged the view of individuals as passive recipients of environmental influences and highlighted the role of a broad set of contextual forces. Elder (1974, 1999), Magnusson (1995), and others assign a

central role to human agency in all developmental processes, not just parent-child relations. Drawing on examples of families' coping with Depression-related hardships through work and residential strategies, Elder (1999) summarized his "principle of human agency" as follows: "Individuals construct their own life courses through the choices and actions they take within the opportunities and constraints of history and social circumstances" (1999, p. 15). Magnusson noted that researchers have referred to the importance of person-environment interactions with several different terms: *transactionism*, *reciprocal determinism*, *dialectic-contextualistic*, *process-person-context* model, and *developmental contextualism*. All of these terms describe processes in which "the individual is an active, intentional part of the environment with which he or she acts" (Magnusson, 1995, p. 34).

The role of human agency may be obvious when conducting research on adults, yet it is equally important for understanding the development of children. An agentic view of children's development arose from the recognition that parental behavior was at least in part determined by characteristics of children such as age, gender, and temperament (Bell, 1968; Kagan, 1989; Lewis, 1981). Further work, particularly that of Sameroff and Chandler (1975), suggested that parental behavior is shaped as much by changes in children's behavior as it is by stable characteristics of children. Sameroff and Chandler argued that parent-child relations should be viewed as an interactive process; both a parent's and a child's behavior may change in response to the other's actions. This *transactional model* of development assigns a central role to human agency—the scope for individuals' actions (in this case, children) to influence their developmental trajectories.

Scarr and McCartney (1983) categorized a child's influence on his or her environment as evocative and active. *Evocative effects* result when the characteristics and behavior of children elicit different reactions from family members and others in their contexts. *Active effects* describe the process by which children actively select contexts or niches. These child-driven influences may reflect genetic factors (O'Connor, Deater-Deckard, Fulker, Rutter, & Plomin, 1998; Scarr, 1992; Scarr & McCartney, 1983;) but, as will be clear in our discussion, may also be influenced by a broad set of environmental factors.

The contributions of these theoretical approaches create an appropriately sophisticated model of human development. Bidirectional and interactive processes occurring within and between several nested environments have replaced unidirectional processes between individuals and their immediate environments. Accumulating evidence supports these theoretical developments. For example, several studies find that children's temperaments and anti-social behavior influence parenting (Anderson, Lytton, & Romney, 1986; Lee & Bates, 1985; O'Connor, et al., 1998; Patterson, 1982). Parents of difficult children are more likely to be negative and use punitive techniques than parents of more easygoing children, and this behavior is at least in part a reaction to the child's behavior. Maturing children are

active agents in fashioning their unique combination of risk (e.g., involvement with problematic peers) and protective (e.g., involvement with adult mentors) factors (Boyce et al., 1998; Rutter et al., 1997).

If children themselves are not choosing contexts, then parents often play a formative role. Parents' preferences and constraints determine child care arrangements. Throughout childhood, parents' decisions about where to live and whether to send their children to public or private schools determine school quality and other neighborhood amenities available to their children.

Assigning such a central role to agency's effect on environment has profound implications for empirical studies aspiring to assess the role of family and extrafamilial contexts in development. To frame the methodological issues, consider an analysis of the determinants of an individual's achievement or problem behavior. Suppose that we wish to determine the role of family and extrafamilial risk and protective factors. A simple model in which individual  $i$ 's achievement or problem behavior ( $y_i$ ) is an additive function of  $i$ 's family ( $FAM_i$ ), extrafamilial contextual ( $CON_i$ ) influences plus a residual error term ( $e_i$ ) is:

$$y_i = A' FAM_i + B' CON_i + e_i \tag{1}$$

For the moment, we assume one child per family. Our general interest is in obtaining unbiased estimates of  $A'$  and  $B'$ , the respective effects of important elements of family and extra-familial context on our outcome of interest.

Equation 1 ignores a number of potentially important problems.  $FAM$  and  $CON$  may interact with one another so that the causal effect of  $CON$  on  $y$  may depend on  $FAM$  conditions. There may be important nonlinear relations between right- and left-hand side variables in Equation 1, and there may be unmeasured aspects of social context that are correlated with  $CON$ , which leads analysts to mistakenly conclude that, for example, neighborhood poverty influences child development when in fact the absence of affluent neighbors or high-quality public services may instead be the crucial contextual factors. None of these considerations invalidates the general lessons regarding endogeneity bias that we draw from our discussion based on Equation 1. A less tractable complication is when  $y$  and either  $FAM$  or  $CON$  simultaneously determine one another. We discuss this possibility later but do not attempt to provide guidance in how to deal with it.

We focus this article on a different kind of problem with Equation 1: that  $FAM$  and  $CON$  conditions are not allocated randomly across children. As a result, estimates of  $A'$  and  $B'$  may be biased, perhaps seriously, by not understanding the process by which  $FAM$  and  $CON$  conditions arise. In the case of  $CON$ , our concern can be expressed in the form of a supplemental equation:

$$CON_i = C' par_i + D' child_i + f_i \tag{2}$$

where  $par_i$  and  $child_i$  represent respective sets of parental and the individual child's characteristics that determine *CON*. Equation 2 reflects the process by which families and children choose or in some other way end up in the extrafamilial contexts that they do. There are, of course, literatures focused on the processes represented by Equation 2, for example, how children select their peers or how elements of choice, local supply conditions, laws and regulations, and economic constraints explain families' uses of child care and their selection of where to live. Our point is that the implications of these context-determining processes should be central to our efforts to model child outcomes as represented in Equation 1.

Only in rare cases are the contexts in which children develop completely beyond the influence of children and their parents. More common is the case where family preferences and behaviors affect developmental contexts. An example is neighborhood context, where the  $par_i$  measures that may influence neighborhood location decisions include parental beliefs about the importance of neighborhood conditions for their children's development, parental preferences regarding the desirability of living in a city versus the suburbs, parental ability to afford expensive locations, and so forth. A child's own characteristics ( $child_i$ ) may affect neighborhood context (e.g., if early signs of delinquent behavior prompt a family to move to a better neighborhood), but they are much more likely to affect a child's choice of extrafamilial contexts such as peer groups.

In the case of *FAM*, the analogous concern can be expressed as follows:

$$FAM_i = E' par_i + F' child_i + g_i \quad (3)$$

In this case, a different set of  $par_i$  and  $child_i$  factors likely come into play. For example, a child's temperament may influence the affection received from a parent, or parental ability, education and income may influence the quality of the early home learning environment.

Developmental theories inform the linkages in Equation 1 between developmental outcomes and their family and contextual determinants. In contrast, theory often does not explain processes by which family and contextual conditions arise. As a result,  $par_i$  and  $child_i$  constitute omitted variables that may bias estimates of  $A'$  and  $B'$  in Equation 1.

We can account for the effects of these omitted variables by adding them to Equation 1:

$$y_i = A' FAM_i + B' CON_i + par_i + child_i + h_i \quad (4)$$

Here the omission of explicit measures of elements of  $par_i$  will bias  $A'$  and  $B'$  to the extent that: (a)  $par_i$  is an important determinant of  $y_i$  and (b)  $par_i$  is correlated

with  $FAM_i$  and  $CON_i$ . A parallel argument holds for the omitted child-level influences  $child_i$  on  $FAM_i$  and  $CON_i$ .<sup>1</sup>

The direction of bias in estimates of  $A'$  and  $B'$  in models that omit important elements of  $par_i$  and  $child_i$  is often uncertain. In the case of estimating the effects of neighborhood contexts, suppose parents who are well equipped to resist the effects of bad neighborhoods—for example, because of good coping or problem-solving skills—choose to live in them to take advantage of cheaper housing or perhaps shorter commuting times. Unless all relevant measures of parental competence are included in the model, the estimated effects of bad neighborhoods or schools on youth outcomes will have a downward bias. However, if parents were somehow randomly allocated across neighborhoods, assuring that coping and problem-solving skills were not related to neighborhood of residence, then one would likely see a larger, less biased effect of neighborhood on youth outcomes.

It is also possible, however, that parents especially ill equipped to handle bad neighborhoods or schools are most likely to live in them, because these parents lack the (partly unmeasured) problem-solving or other skills that would enable them to move to better neighborhoods. In this case, the relation found by an equation such as Equation 1 between poor neighborhoods or schools and children's poor developmental outcomes result in part from the unmeasured inability of parents to avoid either. Controlling for all of the characteristics of parents and families that determine neighborhood residence is the only way to ensure that the effects of

---

<sup>1</sup>We are glossing over two distinct problems: (a) simultaneous causation and (b) endogeneity. The bulk of this article is devoted to the latter, endogeneity, which is fundamentally a problem of omitted variables. In the case of peer and parent-child transactional models, the problem is better conceived as a simultaneous causation between  $y_i$  and his or her parental or extrafamilial environment (Manski, 1993; Moffitt, 2001). In this case, we have a two-equation system:

$$(A) y_i = A' FAM_i + B' CON_i + e_i$$

and either:

$$(B) FAM_i = C' y_i + D' Z + w_i$$

or

$$(C) CON_i = C' y_i + D' Z + w_i,$$

where  $Z$  is a vector of other determinants of family or contextual conditions that might include the behavior of other individuals who are part of the context as well as structural and political factors. The idea that children are not only shaped by, but also shape, their family environment is a familiar one to developmentalists and a key element of transactional models of development (Sameroff and Chandler, 1975). That two-way "transactions" may play a role in extrafamilial contexts is best seen in the case of best friends or peer groups. In the case of best friends,  $CON_i$  might be the behavior of  $i$ 's best friend. Equation (a) then reflects the assumption that  $i$ 's behavior is causally linked to the behavior of his or her best friend. However, Equation (c) then reflects the assumption that  $i$ 's best friend's behavior is also causally dependent on  $i$ 's own behavior. As explained by Moffitt (2001) and Manski (1993), identification of the  $B$ s and  $C$ s in a two-equation system is a difficult task.

current neighborhood conditions on youth outcomes are not overestimated. It is an easy matter to generate similar arguments for upward and downward biases from omitted variables in studies of any number of risk, protective, or other elements of family and extrafamily contexts.<sup>2</sup>

## ENDOGENEITY IN PRACTICE

In the absence of experimental studies, “best practice” in developmental research typically consists of addressing the endogeneity problem by using multivariate regression procedures to control for child, parent, school, and neighborhood covariates. If the problem of endogeneity can be thought of as one of unmeasured variables, the standard approach amounts to an attempt to measure-the-unmeasured. We argue that alternative strategies to overcome the endogeneity problem should be considered owing to the growing body of evidence suggesting that the measure–the-unmeasured approach may still leave serious biases.

One obvious problem of a measure-the-unmeasured approach is the question of what to measure. Our reading of much of the developmental literature indicates that, with the exception of behavioral geneticists, little theoretical and empirical energy has been devoted to understanding the process by which parents and children are selected into particular contexts. Such theory may suggest the importance of gathering measures of parenting characteristics, preferences and abilities, children’s temperaments, and so forth. Even with such theory as a guide, measuring all relevant concepts will be a daunting—perhaps impossible—task for even the most elaborate (and expensive) social science survey.

The task of measuring the unmeasured becomes especially problematic with cross-sectional data, because some measures that may be central to the selection processes are themselves endogenous. This leads to the problem of overcontrolling in regression models. For example, a study of the effect of classroom quality on children’s achievement may introduce controls for student variables, such as the student’s engagement in school. The problem here is that both student engagement and achievement may be partially determined by a still unmeasured selection process. Including an endogenous variable such as engagement on the right-hand side

---

<sup>2</sup> For example, despite theoretical arguments to the contrary, most empirical studies of the effects of divorce on children have assumed that divorce is randomly assigned to children. They do this by failing to control for the fact that divorce is the product of the parents’ temperaments, resources, and other stressors that face parents, most of which will influence children’s outcomes in their own right. As a result, studies comparing developmental outcomes of children with and without past parental divorces after controlling for a handful of family background characteristics are likely to confound the effects of divorce with the effects of unmeasured parent and child variables. Indeed, studies that control for children’s behavior problems prior to a possible divorce find much smaller apparent effects of the divorce itself (Cherlin, Chase-Lansdale, & McRae, 1998).



of a regression such as Equation 1 or Equation 4 will bias the other coefficients in ways that depend on the intercorrelation among all right-hand side variables and their separate correlations with  $y$  (Hausman, 1983). Introducing a lagged measure of engagement reduces the endogeneity problem but does not eliminate it if unmeasured determinants of past engagement and current test scores remain correlated.

Note that as long as key determinants of *FAM* and *CON* are omitted from the empirical analysis, construct validity for key *FAM* and *CON* measures is not sufficient to ensure unbiased estimation of their effects on outcomes of interest (Cook & Campbell, 1979). Suppose that we seek to estimate the effects of parent–child attachment on child outcomes but believe that attachment depends at least in part on child temperament (Kagan, 1989). A regression of a perfectly measured child outcome on perfectly measured attachment will yield a biased estimate of the effect of attachment on child outcomes as long as key child temperament measures are not controlled as well.

If concerns about endogeneity are not new to the field of developmental science, why does “best practice” fail to adequately address the problem? One answer is that researchers may be concerned about the potential for substantial bias but unclear about other approaches to the problem beyond regression adjustment. In the next section, we describe a number of ways of addressing endogeneity problems beyond regression adjustment for observed covariates. An alternative answer is that researchers may believe that the magnitude of any resulting bias is likely to be modest and of little practical importance. Yet evidence to the contrary—that selection bias may be quite substantial in practice—comes from a growing body of research that begins with random-assignment experimental evidence of the effects of some environmental factor of interest on developmental outcomes and then examines the ability of nonexperimental approaches to reproduce the unbiased experimental estimate.

Experiments overcome the endogeneity problem by randomly assigning an important element of a child’s developmental environment, thereby breaking the link between the measured element of *FAM*<sub>*i*</sub> or *CON*<sub>*i*</sub> of interest and the unmeasured attributes  $\text{par}_i$  and  $\text{child}_i$  in Equation 4.<sup>3</sup> Although the experimental study of human development may strike readers as infeasible on practical or ethical grounds, the fact is that a number of experiments outside of the laboratory setting introduce random variation in the family and extrafamilial risk, protective and other contextual factors of interest to developmental scientists. For example, although random assignment of children to parents is possible in animal (e.g., Suomi, 1997) but not human studies, studies of children of parents who have been randomized to, say,

---

<sup>3</sup> It should be noted that even with randomized experiments there are possible sources of bias. Problems such as demoralization or diffusion effects should be considered in the design and analysis of experiments (Cook & Campbell, 1979).

drug treatments for depression or education treatments to raise earnings, are examples of promising ways of using random assignment in assessing effects of parental mental health or skills on children's outcomes. Also, random assignment of children from waiting lists to child care settings of varying quality, and even of families to neighborhoods of varying quality, is both feasible and, in some cases, ethical.

These experiments can provide researchers with some sense for the bias that results from nonexperimental estimates as well as providing direct evidence for the causal effects of some developmental influence of interest. For example, Wilde and Hollister (2002) compare nonexperimental and experimental results for the widely cited Tennessee Student-Teacher Achievement Ratio (STAR) class-size experiment. The STAR experiment provides an unbiased estimate of the impact of class size on student achievement by comparing the average achievement levels of students assigned to small (experimental) and regular (control) classrooms. However, Wilde and Hollister also estimated a series of more conventional nonexperimental regressions that related naturally occurring class size variation within the set of regular classrooms to student achievement, controlling for an extensive set of student demographic characteristics and socioeconomic status.

Table 1 compares the experimental and nonexperimental estimates of class size impacts by school. The table shows substantial variability across schools in the ef-

TABLE 1  
Comparison of Experimental and Nonexperimental Estimates  
for Effects of Class Size on Student Test Scores

| <i>School</i> | <i>Nonexperimental Regression</i> | <i>Experimental Estimate</i> |
|---------------|-----------------------------------|------------------------------|
| A             | 9.6                               | -5.2                         |
| B             | 15.3*                             | 13.0*                        |
| C             | 1.9                               | 24.1*                        |
| D             | 35.2*                             | 33.1*                        |
| E             | 20.4*                             | -10.5                        |
| F             | 0.2                               | 1.3                          |
| G             | -8.6                              | 10.6*                        |
| H             | -5.6                              | 9.6*                         |
| I             | 16.5*                             | 14.7*                        |
| J             | 24.3*                             | 16.2*                        |
| K             | 27.8*                             | 19.3*                        |

*Note.* Adapted from Table 3 from Wilde, E. T., & Hollister, R. (2002). *How close is enough? Testing nonexperimental estimates of impact against experimental estimates of impact with education test scores as outcomes.* Madison, WI: Institute for Research on Poverty Discussion Paper number 1242-02. Reprinted with permission of the author.

The table reports estimated effects of small classes (average 15 students per classroom) with those of larger classrooms (average 22 students) from the Tennessee STAR class-size experiment.

\* = estimate statistically significant at the 5% cutoff.

fects of smaller classes on student standardized test scores. In some cases (e.g., Schools B, D, and I), the two sets of estimates are quite close, but in some (e.g., Schools C, E, G, and H) they are quite different. A comparison of the nonexperimental and experimental results as a whole reveals that the average bias (i.e., the absolute difference between the experimental and nonexperimental impact estimates) is on the order of 10 percentile points—about the same as the average experimental estimate for the effects of smaller classes!

A second example of the bias that may result with nonexperimental estimates comes from the U.S. Department of Housing and Urban Development’s Moving to Opportunity (MTO) housing-voucher experiment, which randomly assigned housing-project residents in high-poverty neighborhoods of five of the nation’s largest cities to either a group that was offered a housing voucher to relocate to a lower poverty area or to a control group that received no mobility assistance under the program (Ludwig, Duncan, & Hirschfield, 2001). Because of well-implemented random assignment, each of the groups on average should be equivalent (subject to sampling variability) with respect to all observable and unobservable preprogram characteristics.

Table 2 presents the results of using the randomized design of MTO to generate unbiased estimates of the effects of moving from high- to low-poverty census

TABLE 2  
 Estimated Impacts of Moving From a  
 High- to a Low-Poverty Neighborhood  
 on Arrests Per 100 Juveniles

|                 |        |        |
|-----------------|--------|--------|
| Violent Crime   |        |        |
| Nonexperimental | -4.9   | (12.5) |
| Experimental    | -47.4* | (24.3) |
| Sample Size     | 259    |        |
| Property Crime  |        |        |
| Nonexperimental | -10.8  | (14.1) |
| Experimental    | 29.7   | (28.9) |
| Sample size     | 259    |        |
| Other Crimes    |        |        |
| Nonexperimental | -36.9* | (14.3) |
| Experimental    | -.6    | (37.4) |
| Sample size     | 259    |        |

*Note.* From Ludwig (1999), based on data from the Baltimore Moving to Opportunity experiment. Regression models also control for baseline measurement of gender, age at random assignment, and preprogram criminal involvement, family’s preprogram victimization, mother’s schooling, welfare receipt and marital status.

\* = estimated effect of dropout program on dropout rates statistically significant at the 5% cutoff level.

tracts on teen crime. The experimental estimates are the difference between average outcomes of all families offered vouchers and those assigned to the control group, divided by the difference across the two groups in the proportion of families who moved to a low-poverty area. (Note the implication that these kinds of experimental data can be used to produce unbiased estimates of the effects of neighborhood characteristics on developmental outcomes, even if the take-up rate is less than 100% in the treatment group and greater than 0% among the control group.)<sup>4</sup> The nonexperimental estimates simply compare families who moved to low-poverty neighborhoods with those who did not, ignoring information about each family's random assignment and relying on the set of prerandom assignment measures of MTO family characteristics to adjust for differences between families who chose to move and those who do not.<sup>5</sup>

As seen in Table 2, even after statistically adjusting for a rich set of background characteristics the nonexperimental measure-the-unmeasured approach leads to starkly different inferences about the effects of residential mobility compared with the unbiased experimental estimates. For example, the experimental estimates suggest that moving from a high- to a low-poverty census tract significantly reduces the number of violent crimes. In contrast, the nonexperimental estimates find that

---

<sup>4</sup>This is best accomplished through an instrumental variables procedure, which can be easily implemented using a two-stage procedure (Foster & McLanahan, 1996; Greene, 1993). In the first stage, the contextual measure (an example in the MTO case is a family's census tract poverty rate) is the dependent variable and is predicted by other measured variables in Equation 6 plus exogenous variables, or instruments ( $Z_i$ ), that are not themselves determinants of the outcome of interest,  $y$ . In the second stage (Equation 6),  $B'$  is estimated by replacing  $CON_i$  with the predicted value of  $CON_i$  obtained in the first stage.

$$CON_i = C' FAM_i + D' Z_i + j_i \quad (5)$$

$$y_i = A' FAM_i + B' \text{Pred}(CON_i) + k_i \quad (6)$$

For this procedure to yield unbiased estimates, the instruments must have first-stage explanatory power—that is, successfully predict variation across individuals or families in  $CON_i$ —but be uncorrelated with the unobserved determinants of the outcome of interest,  $k_i$ . Indicators for the MTO treatment group into which a family is assigned are ideal instruments because they substantially affect postprogram neighborhood poverty rates (by affecting the “price” of moving to a lower poverty area) and are otherwise uncorrelated with observed or unobserved determinants of developmental outcomes by virtue of the random assignment of families to treatment groups. As we discuss further later in the article, sometimes naturally occurring shifts in public policies or family conditions or other “natural experiments” can also generate useful instrumental variables.

<sup>5</sup>Data available for MTO families include the results of a lengthy baseline questionnaire that all families were required to complete in order to be enrolled in the program, so the response rate is 100% for program participations by construction. This baseline survey included basic sociodemographic information for all members of the household, detailed information about the family's sources of income, their residential history, and reasons why they enrolled in MTO and wished to move. In addition are the results of complete arrest histories for each program participant, which enable the analysts to control for each individual's preprogram offending history.

such moves have essentially no effect on violent arrests. In the case of “other” crimes, the nonexperimental estimates suggest that such moves reduce crime, but the experimentally based estimates do not.

A final example comes from the National Evaluation of Welfare-to-Work Strategies, randomized experiment designed to evaluate welfare-to-work programs in seven sites across the United States. One of the treatment streams encouraged welfare-recipient mothers to participate in education activities. In addition to measuring outcomes such as clients’ welfare receipt, employment, and earnings, the evaluation study also tested young children’s school readiness using the Bracken Basic Concepts Scale School Readiness Subscale. Using a method for generating experimental estimates similar to that used in the MTO analyses, Magnuson and McGroder (2002) examined the effects of the experimentally induced increases in maternal schooling on children’s school readiness. Again, the results suggest that nonexperimental estimates did not closely reproduce experimentally based estimates.

A much larger literature within economics, statistics, and program evaluation has focused on the ability of nonexperimental regression-adjustment methods to replicate experimental estimates for the effects of job training or welfare-to-work programs. Although the “contexts” represented by these programs may be less interesting to developmentalists, the results of this literature nevertheless bear directly on the question considered in this article: Can regression methods with often quite detailed background covariates reproduce experimental impact estimates for such programs? As one recent review concluded, “Occasionally, but not in a way that can be easily predicted” (Glazerman, Levy, & Myers, 2002, p. 46; see also Bloom, Michalopoulos, Hill, & Lei, 2002). And in this literature, as with the education and neighborhood programs discussed earlier, the magnitudes of these biases are often quite large judged from either a scientific or a policy perspective.

## SOLUTIONS TO THE ENDOGENEITY PROBLEM

Evidence from randomized experiments suggests that the standard practice in most nonexperimental developmental studies—multivariate regression controls for observable covariates in an attempt to measure the unmeasured—may yield biased estimates for the effects of developmental contexts on individual outcomes, even when quite detailed background data are available for participants. These findings should give researchers pause before drawing causal inferences from standard regression methods applied to nonexperimental data.

One obvious solution to the endogeneity problem is to conduct more randomized experiments. A related option is to take better advantage of experiments that are already underway by adding measures of family and child content to existing random-assignment experiments. A variety of welfare policy experiments have introduced random variation in job training or job search (Gueron & Pauly, 1991),

formal education (Bos & Fellerath, 1997), and parenting skills (Quint, Bos, & Polit, 1997), which in turn generate variation in contexts of interest to developmentalists. Thus, experimental treatments provide a way of securing unbiased estimates of the contexts affected by the experimental manipulation.

Although randomized experiments in the area of human development are perhaps not as difficult to implement or exploit as many social scientists believe, the fact remains that such experiments remain all too rare in practice. What can be done in cases where experimentation is not an option?

### Within-Child, Across-Time Variation

One useful approach sometimes taken in some developmental studies is to use longitudinal data to estimate change models. For reasons that will become clear shortly, we refer to this approach as *individual fixed-effects models*. The equation-based intuition behind them can be gleaned from Equation 7, where child  $i$ 's outcome in period  $t$ ,  $y_{it}$ , is a function of family and contextual effects in that period ( $FAM_{it}$  and  $CON_{it}$ ), unmeasured parent and child variables that are constant over time ( $par_i$  and  $child_i$ ) and unmeasured variables that vary over time ( $par_{it}$  and  $child_{it}$ ), and an error term ( $m_{it}$ ):

$$y_{it} = A' FAM_{it} + B' CON_{it} + par_i + child_i + par_{it} + child_{it} + m_{it} \quad (7)$$

One way to estimate fixed-effects models is to first-difference the data, a procedure in which each child's observation in period  $t - 1$  is subtracted from his or her observation in period  $t$ . Subtracting Equation 7 at Time  $t - 1$  from Equation 7 at Time  $t$  eliminates the time-invariant parent- and individual-level unmeasured variables from the right-hand side of the regression equation. Fixed-effects models are still subject to bias from time-varying parent- or child-level unmeasured variables, because first-differencing does not eliminate these variables from the estimating equation. However, with sufficiently long panels, more elaborate methods may be used to control for unmeasured variables whose values change over time in specific ways. For example, looking at how changes in the rate at which family or contextual variables change affect the rate at which children's outcomes change (obtained by twice-differencing the data) can help control for unmeasured parent- or child-level variables that change over time at a constant rate.

### Within-Family Variation

Another set of methods for reducing bias exploit within-family variation. An example is behavioral geneticists' use of the varying degrees of genetic relatedness to estimate the role of genes and shared and unshared environmental influences (Falconer, 1981, Plomin et al., 1990). Identical twins are 100% genetically related; fra-

ternal twins, nontwin siblings, and other first-degree relatives have an expected degree of genetic relatedness half that of monozygotic twins. Simple behavioral genetics models capitalize on the natural experiment implicit in these different kinds of sibling births to infer heritability.<sup>6</sup>

Family fixed-effects models constitute another example of using within-family variability to eliminate bias, in this case from omitted parental ( $par_i$ ) factors. In these approaches each sibling's score on the dependent and independent variables is subtracted from the average scores of all siblings in his or her family. In the case of two siblings per family, the deviation-from-family-means model becomes a sibling-difference model. If we replace the subscript  $i$  in Equation 4 with 1 (for Sibling 1) and 2 (for Sibling 2), and assume that there is sufficient cross-sibling variability in family and contextual conditions to reference  $FAM$  and  $CON$  with the sibling subscripts, the sibling difference model takes the following form:

$$y_2 - y_1 = A'(FAM_2 - FAM_1) + B'(CON_2 - CON_1) + (par_2 - par_1) + (child_2 - child_1) + (h_2 - h_1) \tag{8}$$

In estimating this regression model, sibling differences in the outcome of interest are regressed onto sibling differences in observed family and contextual characteristics. Note that coefficients  $A'$  and  $B'$  have identical interpretations in Equations 4 and 8; both reflect changes in outcomes associated with key changes in context.

Observed parental factors, such as parental educational attainment, which are the same for all siblings in a family, are differenced out of a sibling difference regression. A key advantage of sibling models is that persistent unobserved elements of  $par_i$  are differenced out as well, thus eliminating the omitted-variable bias caused by the unmeasured persistent family factors shared by siblings. The sibling-difference model thus “automatically” eliminates bias from all permanent family factors, observed or not, if the effects of these factors do not differ between siblings.<sup>7</sup> Time-varying family factors, especially those that might be producing the sibling differences in the context (e.g., divorce, income changes), are a potential source of bias in Equation 8 and should be controlled explicitly in the regression if possible. Note, however, that they will bias estimates only to the extent that

---

<sup>6</sup>It hardly needs to be said that conclusions from such models are controversial. Factors that render problematic conclusions of simple behavioral genetics models include: (a) *in utero* environmental shocks and twin competition for resources (Devlin, Daniels, & Roeder, 1997); (b) more similar treatments accorded to twins by parents, teachers, and classmates; (c) more similar environments sought by monozygotic twins relative to other siblings; (d) measurement error; and (e) more similar sibling mutual influence for monozygotic twins relative to other siblings.

<sup>7</sup>Even if unobserved family factors differ across siblings, it is often reasonable to presume a low correlation between sibling differences in those family factors and sibling differences in context, in which case even unmeasured sibling-specific family factors will not impart much bias to estimates of contextual effects ( $B'$ ).

they are correlated with the context differences. If uncorrelated with them, the unmeasured family differences between siblings will contribute to the lack of explanatory power of a sibling-difference model but will not bias the parameter estimates. A disadvantage of sibling models is that change measures are more often error ridden than level measures, which can bias parameter estimates toward zero.

Currie and Thomas (1995) and Garces, Thomas, and Currie (2002) have compared the achievement and behavior problem scores of siblings in two national longitudinal studies, one of whom attended Head Start while the other did not. They argued that such sibling differences provide a less biased estimate of the impacts of Head Start than do nonexperimental studies that compare outcomes of Head Start attendees and a matched set of children from different families who did not attend Head Start. The case for Currie and Thomas's approach is strengthened to the extent that persistent and difficult-to-measure family factors (e.g., unusual concern for their children's development, maternal depression) influence enrollment decisions and child outcomes common to all children in a family. Failure to adjust for such family factors will likely lead case control studies of children from different families to attribute to Head Start effects that ought to be attributed to these family factors.

The sibling approach is biased to the extent that decisions about Head Start enrollment reflect unmeasured differences in the mothers' perceptions of the differential needs of her children for Head Start. Key to the success of sibling models is an understanding of and statistical adjustments for the process by which children from the same family end up in different contexts of interest. Much like the measure-the-unmeasured strategy, a sibling approach requires researchers to devote considerable effort to understanding the determinants of key contexts and why these determinants might differ between siblings (Griliches, 1979).

### Other Natural Experiments

Other developmental research has been able to take advantage of novel natural experiments involving family and extrafamilial contexts. Rutter (1998) described a study of children raised in Romanian orphanages, whose lack of contact with affectionate parents or caretakers was completely beyond the children's control and whose deprivation provides valuable data on the nature of critical periods for attachment. Goldin-Meadow (1997) studied deaf children whose parents did or did not expose them to conventional sign language early in life to understand the nature of early language formation.

Among the better known developmental studies that rely on quasi-experimental variability in an extrafamilial context is Glen Elder's (1974) work on the effects of the Great Depression on children and parents. He focused on exogenous variability in *CON* introduced by the timing of the macroeconomic shock with respect to children's age. Elder (1974) found that the effect of the macroeconomic adversity on children's development differed depending on their age. Boys who experienced the



Great Depression when they were young enough to be completely dependent on their parents suffered much more than boys for whom the Depression occurred when they were already adolescents and thus able to take jobs and cope more actively with the adversity. The situation was reversed for girls. Girls who were younger children during the Depression did relatively well—perhaps, Elder (1974) speculated, because they formed tighter bonds with their mothers. Adolescent girls did not have the job and other coping strategies that boys did, thus they and fared worse.

Elder (1974) also contrasted families for whom macroeconomic conditions produced large income drops with families whose incomes did not fall as much. Here, however, the case for the exogeneity of the income changes is much weaker than the case for the exogeneity of children's ages at the time of the Great Depression. Children's birth timing is beyond their control (although not completely beyond the control of their parents), whereas the magnitude of the impact of the Depression on family income may well depend on difficult-to-measure parental characteristics such as ability, motivation, and risk aversion.

It is important to contrast the nature of environmental variation exploited in Elder's (1974) work on age-related child impacts with that found in the more typical developmental study. In the latter case, naturally occurring differences in more local contexts (e.g., neighborhood conditions, schools in a single urban school district, mentors, family structure and networks, social support) are related to differential developmental outcomes. All of these local contexts are subject to influence by the individuals or parents of individuals whose development is being studied. Accordingly, all such studies of the effects of these contexts are subject to the endogeneity biases we describe.

We recommend that researchers' first instincts for gathering data should not be to exploit variability in risk, protective, and other contextual measures of interest gathered from convenient homogeneous or even population samples. No amount of sophisticated structural equation modeling and intensive measurement is likely to solve the endogeneity problem inherent in these kinds of data. Rather, we argue for a more opportunistic strategy, exploiting variability in contextual measures of interest stemming from real or natural experiments. Analyses built around such experiments may provide powerful leverage over the endogeneity problem.

A common reaction to such advice may be to admire the ingenuity of the researchers who have conducted such studies but to consider the application of these to one's own area of interest inconceivable. In an attempt to ward off this notion, we provide a number of examples of what we have in mind in hope of stimulating thoughts of analogous designs:

- Rather than rely on self-selected differences in television viewing to study the impacts of television on aggression, Joy, Kimball, and Zabrack (1986) took advantage of the fact that a town in northwestern Canada had a delayed introduction to television because its valley location blinded it to channel transmitters for several years.

They found that two years after the introduction of just one television channel, the Canadian-owned government channel, both boys and girls of all ages and pre-television aggression levels, showed increased verbal and physical aggression.

- Rather than rely on self-selected differences in residential location to study the impacts of neighborhood on maternal employment and children's development, Rosenbaum (1991) studied families enrolled in the Gautreaux program, a court-ordered opportunity for Chicago public housing residents to relocate to city and suburban addresses on a quasi-random basis. He finds much more favorable school and career outcomes for suburban as compared with city-based movers.

- Rather than rely on teen-driven differences in fertility to study the effects of fertility timing on maternal career success, Hotz, McElroy, and Sanders (1997) used miscarriages as an exogenous source of fertility delay. Surprisingly, they found that mothers "forced" by miscarriage to have later births did no better in the labor market than mothers with earlier fertility. Similar methods could be used to assess the effects of fertility timing on child development.

- Rather than rely on parent-driven differences in family size to study the effects of family size on adolescent development, Bronars and Grogger (1994) used twin versus singleton births as an exogenous source of increased family size. They found large, but only in the short run, effects of such births on labor force participation, poverty and welfare receipt on unmarried women, and no corresponding effects on married women.

- Rather than rely on student- and parent-driven differences in high school coursework to study the effects of academic courses on student achievement, Girotto and Peterson (1999) took advantage of the fact that high school students with summer birthdays took drivers' education during the summer (in addition to academic school-year coursework), whereas all other students took drivers' education during the school year (often replacing academic courses). They estimated that one additional yearlong academic course is associated with a 0.25 standard deviation increase in standardized test scores.

- To untangle the effects of schooling on IQ, Green, Hoffman, Morse, Hayes, and Morgan (1964) took advantage of the fact that, rather than racially integrate schools, Prince Edward County, Virginia, closed down their public schools. Most Black children in the county did not attend school, and Green and colleagues compared IQ scores of children in Prince Edward County to other Black children from similar backgrounds in nearby counties. They found that a missed year of school resulted in a 6-point lower IQ score.

## DISCUSSION

Our reading of much of the developmental science literature is that empirical studies of contextual influences risk possibly serious biases from the actions of the in-

dividuals whose development is being studied. Developmental theory assigns a prominent role to human agency, but if, as seems likely, actions by individuals or their parents shape the constellation of risk, protective, and other vital contextual factors they experience, then failure to adjust in empirical work for the resulting endogeneity bias may produce a highly misleading picture of the role of context.

Because we see little value in restricting the theoretical scope of developmental models, we advocate that developmentalists consider reorienting their data collection strategies toward situations that produce random or quasi-random variability in the contexts of interest. In terms of random-assignment experiments, the policy world is producing an increasing number of interesting social experiments that either manipulate contexts directly or manipulate factors that in turn influence choice of context.

### Caveats

Natural experiments are at once plentiful and treacherous. Our many examples illustrate the wide range of natural experiments that have been used in behavioral studies. Problems arise when seemingly exogenous changes are not completely independent of the individuals and families whose behavior one is modeling. For example, family-driven geographic mobility may influence labor market conditions (Katz & Blanchard, 1993), welfare-program benefits (Meyer, 1999), and other local-area conditions of interest. Also, in the case of sibling models, parent-driven actions may determine why their children are exposed to different contexts (e.g., Head Start) of interest.

Sometimes the hoped-for variability in a context of interest is insufficient to inform an analysis of interest. Although Elder (1986) and, using the Glueck Crime Causation Study data, Sampson and Laub (1996), were able to show conditions under which military service and the GI Bill transformed developmental trajectories for the better, few historical periods are as turbulent as those that preceded and followed the second world war.

All in all, taking the endogeneity problem seriously will complicate further the already-complicated nature of empirical work in developmental research. On the other hand, opportunities along these lines abound. The number of large-scale randomized experiments conducted in the social sciences has increased substantially over the past several years. Developmentalists need to join forces with the experimenters to ensure that future randomized social science trials incorporate key process and developmental outcome measures at both baseline and follow-up.

Addressing the endogeneity problem in the absence of experimental data requires attention to natural experiments that produce exogenous variations in processes and contexts of interest. Dreaming up useful natural experiments requires a different, opportunistic kind of thinking and data collection strategy. Historians

may be helpful collaborators in this work, as may the sociologists, economists, and urban planners who study the institutional details of contexts of interest.

Only if researchers take the endogeneity problem seriously in empirical work can developmental theories be tested convincingly and policy-relevant conclusions reliably drawn.

## ACKNOWLEDGMENTS

We are grateful to the Family and Child Well-being Research Network of the National Institute of Child Health and Human Development (U01 HD30947-06) for supporting this research and Dan Levy for providing research assistance. Portions of the article parallel an assessment of methodological problems in neighborhood research presented in Duncan and Raudenbush (1999). We are grateful for helpful comments from Nancy Cohen, Rachel Dunifon, Michael Foster, Jennifer Greene, John Modell, Pamela Morris, Hiro Yoshikawa, and participants in the “Lives In Context” conference at the Murray Center for the Study of Lives, November 12–13, 1999.

## REFERENCES

- Anderson, K. E., Lytton, H., & Romney, D. M. (1986). Mother's interactions with normal and conduct disordered boys: Who effects whom? *Developmental Psychology*, *22*, 604–609.
- Baumrind, D. (1968). Effects of authoritative parental control on child behavior. *Child Development*, *37*, 887–907.
- Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review*, *75*, 81–95.
- Blanchard, O., & Katz, L. (1993). Regional Evolutions. *Brookings Papers on Economic Activity*, *1992*:1, 1–61.
- Bloom, H. S., Michalopoulos, C., Hill, C., & Lei, Y. (2002). *Can non-experimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* New York: Manpower Demonstration Research Corporation.
- Bos, J., & Fellerath, V. (1997). *LEAP: Final report on Ohio's welfare initiative to improve school attendance among teenage parents*. New York: Manpower Demonstration Research Corporation.
- Boyce, T. W., Frank, E., Jensen, P. S., Kessler, R. C., Nelson, C., Steinberg, L., & The MacArthur Foundation Research Network on Psychopathology and Development. (1998). Social context in developmental psychopathology: Recommendations for future research from the MacArthur Network on Psychopathology and Development. *Development and Psychopathology*, *10*, 143–164.
- Bronars, S., & Groger, J. (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review*, *84*, 1141–1156.
- Cherlin, A., Chase-Lansdale, L. P., & McRae, C. (1998). Effects of parental divorce on mental health throughout the life course. *American Sociological Review*, *63*, 239–249.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis for field settings*. Chicago: Rand McNally.

- Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, 85, 341–364.
- Davis, A., & Havinghurst, R. J. (1946). Social class and color differences in child-rearing. *American Sociological Review*, 11, 698–710.
- Dehejia, R., & Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053–1062.
- Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, 338, 468–471.
- Duncan, G., & Raudenbush, S. (1999). Assessing the effects of context in studies of child and youth development. *Educational Psychology*, 34, 29–41.
- Elder, G. (1974). *Children of the Great Depression: Social change in life experience*. Chicago: University of Chicago Press.
- Elder, G. (1986). Military times and turning points in men's lives. *Developmental Psychology*, 22, 233–245.
- Elder, G. (1999, April). *Historical times and lives: A journey through time and space*. Paper presented at "Landmark Studies of the 20th Century." Henry A. Murray Center for the Study of Lives, Cambridge, MA.
- Falconer, D. (1981). *Introduction to quantitative genetics*. New York: Longman.
- Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school? *Psychological Methods*, 1, 249–260.
- Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *American Economic Review*, 92, 999–1012.
- Glazerman, S., Levy, D., & Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review*. Washington, DC: Mathematica Policy Research.
- Goldin-Meadow, S. (1997). When gestures and word speak differently. *Current Directions in Psychological Science*, 6, 138–143.
- Green, R. L., Hoffman, L. T., Morse, R., Hayes, M. E., & Morgan, R. F. (1964). *The educational status of children in a district without public schools*. Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.
- Greene, W. (1993). *Econometric analysis*. (2nd ed.). New York: Macmillan.
- Giroto, J., & Peterson, P. (1999). Do hard courses and good grades enhance cognitive skills? In S. Mayer and P. Peterson (Eds.), *Earning and learning: How schools matter* (pp. 205–230). New York: Russell Sage Foundation.
- Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87, S37–S64.
- Gueron, J. M., & Pauly, E. (1991). *From welfare to work*. New York: Russell Sage Foundation.
- Hausman, J. (1983). Specification and estimation of simultaneous equation models. In Z. Griliches and M. Intriligator (Eds.), *Handbook of econometrics* (Vol. 1, pp. 391–448). Amsterdam: North-Holland.
- Hotz, V. J., McElroy, S. W., & Sanders, S. G. (1997). The costs and consequences of teenage childbearing for mothers. In R. A. Maynard (Ed.), *Kids having kids: The economic and social consequences of teen pregnancy* (pp. 55–94). Washington, DC: Urban Institute.
- Joy, L. A., Kimball, M. M., & Zabrack, M. L. (1986). Television and children's aggressive behavior. In T. M. Williams (Ed.), *The impact of television: A natural experiment in three communities* (pp. 303–360). New York: Academic Press.
- Kagan, J. (1989). *Three unstable ideas: Temperament, cognition, and self*. Cambridge, MA: Harvard University Press.
- Kohn, M. L. (1959). Social class and the exercise of parental authority. *American Sociological Review*, 24, 352–366.
- Kohn, M. L. (1963). Social class and the parent-child relationship: An interpretation. *American Journal of Sociology*, 68, 471–480.

- Lee, C. L., & Bates, J. E. (1985). Mother-child interaction at age two years and perceived difficult temperament. *Child Development, 56*, 1314–1325.
- Lewis, C. C. (1981). The effects of parental firm control: A reinterpretation of findings. *Psychological Bulletin, 90*, 547–563.
- Ludwig, J. (1999). Experimental and non-experimental estimates of neighborhood effects on low-income families. Unpublished document, Georgetown University.
- Ludwig, J., Duncan, G., & Hirschfield, P. (2001). Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. *Quarterly Journal of Economics, 116*, 655–680.
- Magnuson, K. A., & McGroder, S. (2002). *The effect of increasing welfare mothers' education on their young children's academic problems and school readiness*. Joint Center for Poverty Research, Working Paper No. 280, Northwestern University.
- Magnusson, D. (1995). Individual development: A holistic, integrated model. In P. Moen, G. H. Elder, and K. Luscher (Eds.), *Examining lives in context* (pp. 19–60). Washington, DC: American Psychological Association.
- Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies, 60*, 531–542.
- Meyer, B. (2000). *Do the poor move to receive higher welfare benefits?* Working Paper, Institute For Research on Policy, Northwestern University.
- Moffitt, R. (2001). Policy interventions, low-level equilibria, and social interactions. In S. Durlauf and P. Young (Eds.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.
- O'Connor, T. G., Deater-Deckard, K., Fulker, D., Rutter, M., & Plomin, R. (1998). Geno-type environment correlations in late childhood and early adolescence: Anti-social behavioral problems and coercive parenting. *Developmental Psychology, 34*, 970–981.
- Patterson, G. (1982). *Coercive family processes, social learning approach series* (Vol. 3). Eugene, OR: Castalia.
- Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer* (2nd Ed.). New York: Freeman.
- Quint, J., Bos, J., & Polit, D. (1997). *New Chance: Final report on a comprehensive program for young mothers in poverty and their children*. New York: Manpower Demonstration Research Corporation.
- Rosenbaum, J. (1991). Black pioneers—do their moves to the suburbs increase economic opportunity for mothers and children? *Housing Policy Debate, 2*, 1179–1213.
- Rutter, M. (1998). Developmental catch-up, and deficit, following adoption after severe global early privation. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 39*, 465–476.
- Rutter, M., Dunn, J., Plomin, R., Simonoff, E., Pickles, A., Maughan, B., et al., (1997). Integrating nature and nurture: Implications of the person-environment correlations and interactions for developmental psychopathology. *Development and Psychopathology, 9*, 335–364.
- Sameroff, A., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaking causality. In F. D. Horowitz, M. Hetherington, S. Scarr-Salapatek, and G. Siegal (Eds.), *Review of child development research* (Vol.4, pp. 187–244). Chicago: University of Chicago Press.
- Sampson, R., & Laub, J. (1996). Socioeconomic achievement in the life course of disadvantaged men: Military service as the turning point. *American Sociological Review, 61*, 347–367.
- Scarr, S. (1992). Developmental theories for the 1990s: Developmental and individual differences. *Child Development, 63*, 1–19.
- Scarr, S., & McCartney, K. (1983). How people make their own environment: A theory of genotype ? environment effects. *Child Development, 54*, 424–435.
- Suomi, S. (1997). Early determinants of behavior: Evidence from primate studies. *British Medical Bulletin, 53*, 170–184.
- Wilde, E. T., & Hollister, R. (2002). *How close is close enough? Testing nonexperimental estimates of impact against experimental estimates of impact with education test scores as outcomes*. Institute for Research on Poverty Discussion Paper No. 1242–02, Madison, WI.