

**John Iceland**

**The Dynamics of Poverty Spells and  
Issues of Left-Censoring**

**No. 97-378**

PSC Research Report Series  
January 1997

**The Population Studies Center** at the University of Michigan is one of the oldest population centers in the United States. Established in 1961 with a grant from the Ford Foundation, the Center has a rich history as the main workplace for an interdisciplinary community of scholars in the field of population studies. Today the Center is supported by a Population Research Center Core Grant from the National Institute of Child Health and Human Development (NICHD) as well as by the University of Michigan, the National Institute on Aging, the Hewlett Foundation, and the Mellon Foundation.

**PSC Research Reports** are prepublication working papers that report on current demographic research conducted by PSC associates and affiliates. The papers are written by the researcher(s) for timely dissemination of their findings and are often later submitted for publication in scholarly journals. The PSC Research Report Series was begun in 1981 and is organized chronologically. Copyrights are held by the authors. Readers may freely quote from, copy, and distribute this work as long as the copyright holder and PSC are properly acknowledged and the original work is not altered.



**PSC Publications**

<http://www.psc.lsa.umich.edu/pubs/>

**Population Studies Center, University of Michigan**

1225 S. University, Ann Arbor, MI 48104-2590 USA

## The Dynamics of Poverty Spells and Issues of Left-Censoring

**Abstract:** Although the availability of panel data has augmented our knowledge of poverty dynamics, the challenges posed by the presence of *left-censored* data remain. This analysis, using 1970 to 1985 data from the PSID, demonstrates that omitting left-censored observations, which is still the norm, could lead to serious selection bias in an analysis by systematically discarding individuals in the midst of long-term poverty spells and from certain demographic groups. Although there is still no single accepted way of dealing with left-censoring, several methods that may mitigate the problem are discussed.

**Dataset used:** Panel Study of Income Dynamics (PSID): U.S., 1970-1985

---

### Author's address

John Iceland, Postdoctoral Research Fellow, Population Studies Center, University of Michigan,  
1225 S. University Ave., Ann Arbor, MI 48104. Phone: 313-668-6019. Fax: 313-998-7415.  
Email: [iceland@umich.edu](mailto:iceland@umich.edu)

## The Dynamics of Poverty Spells and Issues of Left-Censoring

Our knowledge about the dynamics of poverty has greatly increased over the past two decades. It was only in 1964, on the eve of the War on Poverty, that an official poverty threshold was defined. Each year, the Current Population Survey (CPS), carried out by the U.S. Bureau of the Census, continues to give us successive snapshots of poverty-- showing declines in the 1960s, little change in the 1970s, and a small overall increase in the 1980s that has persisted into the 1990s. In 1994, the official poverty rate stood at 14.5 percent-- indicating that 38.1 million Americans lived in poverty (Mills, 1995).

The demographic characteristics of the poor, according to the CPS, seem to change little over time, appearing to confirm the stereotype that the same poor families are mired in poverty year after year. However, relatively recent studies using longitudinal data from such sources as the Panel Study of Income Dynamics (PSID) have revealed extensive turnover in the poverty population. Bane and Ellwood (1986), in a thorough examination of the dynamics of poverty, showed that most spells of poverty are in fact quite short: nearly 45 percent end within one year and 70 percent are over within three years. Furthermore, Duncan and Morgan (1982) and Devine and Wright (1993) show that occasional poverty is widespread. The similar successive snapshots provided by the CPS are merely a result of many people with similar demographic characteristics falling into poverty replacing those moving out.

The availability of panel data has clearly augmented our knowledge of poverty dynamics. However, persistent methodological challenges remain. One of the most serious of these is the presence of *left-censored* data-- where a subject has already been exposed to the risk of an event for an unknown amount of time before he or she comes under observation. Most longitudinal datasets, such as the PSID, do not strive to attain complete life-histories of individuals not born into the sample. At the time of the first observation of an adult individual, we may know little of his or her previous employment history or how long he or she has been in poverty. Most analysts simply ignore the problem of left-censoring by discarding left-censored observations. In fact, some of the most well-known and quoted studies on the dynamics of poverty and welfare spells were carried out with data in which left-censored spells were discarded (e.g.; Bane and Ellwood, 1986, 1994; Stevens, 1994; Pavetti, 1993). However, discarding these left-censored spells may introduce selection bias because these individuals may share common characteristics (Guo, 1993; Moffitt and Rendall, 1993).

This analysis, using PSID data from the years 1970-1985, explores the problem of left-censored poverty spells. The impact of omitting versus maintaining these observations in an analysis of poverty dynamics is examined. A logistic model identifying the traits of left-censored cases is estimated. Finally, ways of handling left-censored data in poverty studies are discussed.

### Data

This study employs data on 7,818 poor individuals aged 18 to 64 living in metropolitan areas collected by the Panel Study of Income Dynamics (PSID) from 1970 to 1985. The PSID is a longitudinal survey of individuals and their families living in the United States conducted by the Institute for Social Research (ISR) at the University of Michigan. Starting with a nationally representative sample of approximately 4,800 U.S. households in 1968, the PSID has traced individuals from those households to the present. Adults have been followed as they have grown older, and children have been observed as they advance through childhood and into adulthood, forming families of their own-- which are then added to the sample. The initial panel had a response rate of about 75 percent, but since then response rates have been above 96 percent.

Researchers who have conducted comparisons of the PSID with other data have concluded that sample continues to be representative of the nonimmigrant U.S. population (Hill, 1992).

Because the original focus of the PSID was the dynamics of poverty, the 1968 sample included a disproportionately large number of low-income households. The oversampling of poor families in the late 1960s resulted in a sizable subsample of African-Americans. Included in the PSID files are weights which allow analysts to make estimates from the sample that are representative of the U.S. population. Much of the information collected by the PSID is economic and demographic, with substantial detail on income, employment, family composition changes, and residential location. Thus, because of its longitudinal nature, its oversampling of poverty individuals, and the variety of other detailed information it contains, the PSID is an ideal data set for a study of poverty dynamics.

### **Poverty and Poverty Transitions**

Persons are classified as poor in this study in any year in which their family's money income is less than 1.25 times the official poverty line. This more generous threshold is used to account for the consistently lower rates of poverty calculated from the PSID, as income appears to be more thoroughly reported than in the Current Population Survey (Stevens, 1994). The use of the more generous threshold follows the strategy adopted by Bane and Ellwood (1986) and Stevens (1994).

Measuring the timing of poverty transitions can be tricky. Determining the length of poverty spells that precede a transition depends on the method of measurement, which are not all equal in precision. The principle problem revolves around the issue of *censoring*. That is, many spells begin before the observation window (left-censoring) while others end after the observation window (right-censoring). The calculation of the length of spells, and thus the timing of transitions, depends on how one handles censored spells. Right-censoring poses a far less serious problem than left-censoring. Statistical techniques and software have been developed to easily handle the problem of right-censoring, once one incorporates a few simple assumptions (Allison, 1982; 1984; Yamaguchi, 1991). With left-censoring, a researcher simply can not be sure how long an individual has been in the midst of spell when he or she enters the survey window.

With some longitudinal datasets, left-censoring is not a problem—such as when a cohort is observed at the beginning of its exposure to the event of interest. Consider, for example, that we wished to analyze the length of poverty spells of all persons. We would then select all persons in the dataset who we are able to observe as soon as they were born. If the longitudinal survey proceeds for enough years, we can indeed obtain a clear picture of the dynamics of poverty spells without having to deal with problems of left-censored data. Unfortunately, it is often difficult to obtain data that is this 'clean'. In particular, limiting an analysis to a single cohort often produces an insufficient number of observations, especially if the researcher desires to use the data for any sort of multivariate hazards analysis. Thus, in practice, most analysts ignore the problem of left-censoring by discarding left-censored observations. Yet this practice is problematic, as we will soon see.

In the following analysis, a spell is left-censored when: a) a person is first observed in the sample. At this point, we know little of the person's previous history or how long he or she has been in poverty. Left-censored spells here includes all spells beginning in 1970, the first year of the analysis, and all spells commencing when an individual enters the sample frame at age 18; and b) an individual reappears in the sample after one or more years of missing data-- either because of attrition or because the person lived outside of a metropolitan area for a year or more.

Likewise, spells are right censored when: a) a person is last observed in the sample. This includes all spells in 1985, the last year of the analysis, and all spells ending when an individual leaves the sample frame at age 65; and b) an individual leaves the sample due to attrition or moves out of a metropolitan area.

Generally, the calculation of spell durations involves using survival analysis, which can easily account for the fact that some observations are right-censored and estimate eventual spell length. Survival analyses, based on lifetable methods (see Duncan and Rodgers, 1988), are founded on a few related functions well-known in the event history literature (see Blossfeld, Hamerle, and Mayer, 1989). Density and distribution functions of the poverty duration  $T$  ( $T^30$ ) can be denoted by  $f(t)$  and  $F(t)$ , respectively:

$$F(t) = P(T \leq t) = \int_0^t f(u) du$$

and for all points for which  $F(t)$  may be differentiated,

$$f(t) = F'(t).$$

The *survivor* function

$$S(T) = P(T > t)$$

expresses the probability that an individual remains in the state ("survives") until time  $t$ ; that is, that an event has not yet occurred and the episode is still continuing. Note the relationship between  $S(t)$  and  $F(t)$ :

$$S(t) = 1 - F(t)$$

The survivor function is a non-increasing function of time approaching zero as time elapses.

The *hazard rate* function,  $h(t)$ , can be defined in terms of the previous functions:

$$h(t) = f(t)/S(t).$$

The hazard rate refers to the instantaneous probability that an event occurs (e.g. transition out of poverty) in a time interval given that the event has not occurred before the beginning of the interval.

In table one, we see the importance of the method of measurement on the estimation of poverty spell durations. In this example, we compute what percent of poverty spells are completed after 3 years.<sup>1</sup> Rows one and two do *not* employ the survival method described above. Rather, because they omit right-censored cases, they are based only on *observed* spell completions. Estimates in rows three and four do use the survival methods described above; these estimations account for the fact that some observations are right-censored and estimate eventual spell length. However, when left-censored observations are included in the calculations (rows two and four), no assumptions are made about the spell duration at the beginning of the

---

<sup>1</sup> The three year period chosen serves as an arbitrary example. A similar pattern holds whatever time period is chosen. This is illustrated in more detail shortly in figure one, which displays survival curves (which capture all spell lengths) for two of the rows in table 1.

observation window; rather, it is merely assumed that the first observed year in poverty is the first actual year of poverty.

**Table 1. Proportion of Poverty Spells Over After 3 Years, By Method of Measurement: 1970-1985**

| Type of spells omitted in the measurement                   | Number of spells | Proportion of spells completed after 3 years |
|---|------------------|--|
| 1) Spells where all right- and left-censored spells omitted | 3,009            | 91.5   |
| 2) Spells where all right-censored spells omitted           | 5,316            | 88.0   |
| 3) Spells where all left-censored spells omitted            | 6,504            | 56.3   |
| 4) No spells omitted  | 11,542           | 49.5   |

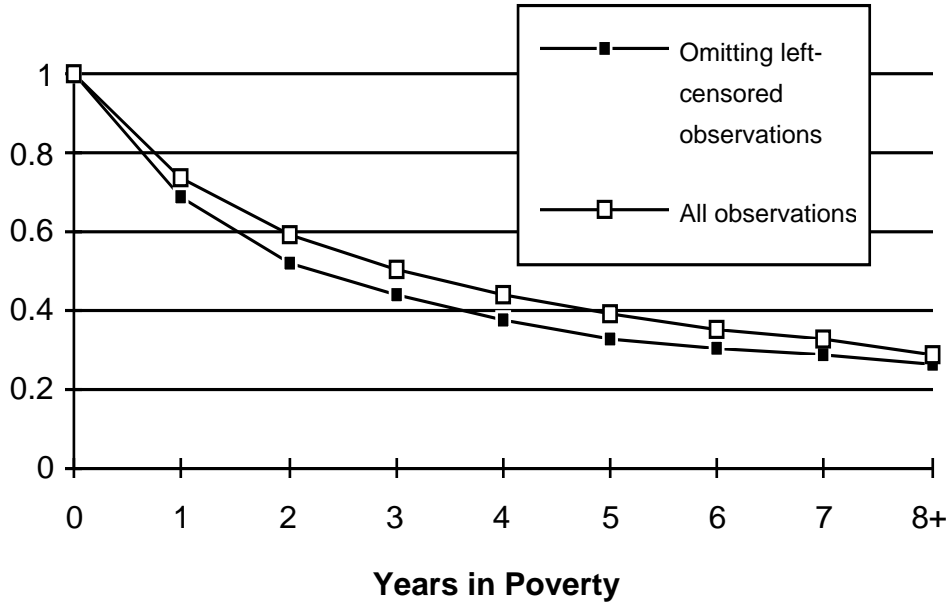
The principle lesson learned from the table is that omitting certain kinds of observations leads to an underestimation of actual spell durations. All differences in means are statistically significant.<sup>2</sup> When we omit both right- and left-censored observations, for example, we tend to omit individuals experiencing longer poverty durations-- people whose poverty spells extend either before or after the observation window. Naturally, short spells are more easily captured in the 16 year observation window and are less likely to be either right- or left-censored. When omitting right- and left-censored observations, over 90 percent of poverty spells are over after three years. On the other hand, observations included in row 4, where no spells are omitted, are most likely to include both short-term and long-term poor individuals-- the latter whose poverty spells extend beyond the observation window. Only about *half* of poverty spells are over after three years when no observations are omitted. Furthermore, we can expect that the 'true' proportion of spells that last three years or less to be even a little lower-- as that figure inevitably includes some left-censored spells that are in fact longer than the 3 years observed.

Figure 1 contains the more complete poverty spell survival curves that are associated with rows 3 and 4 of the previous table. Figure 2 contains the corresponding survival curve for nonpoverty spells for illustrative purposes. The survival curve in figure 1 basically indicates the proportion of persons still in poverty after *x* number of years after first being observed in poverty. Thus, after one year of poverty, 74.1 percent of the people are still in poverty if all observations are considered, while 69.1 percent of persons are still in poverty if left-censored observations are omitted. This difference is statistically significant even at 0.0001. This once again illustrates the fact that omitting observations understates the actual length of spells. This is because left-censored observations are more likely to include individuals in the midst of long-term poverty spells-- who have begun their spell of poverty prior to the start of the observation window. And as just mentioned, even including all observations tends to underestimate the true length of poverty spells, as the complete duration of left-censored spells can not be captured.

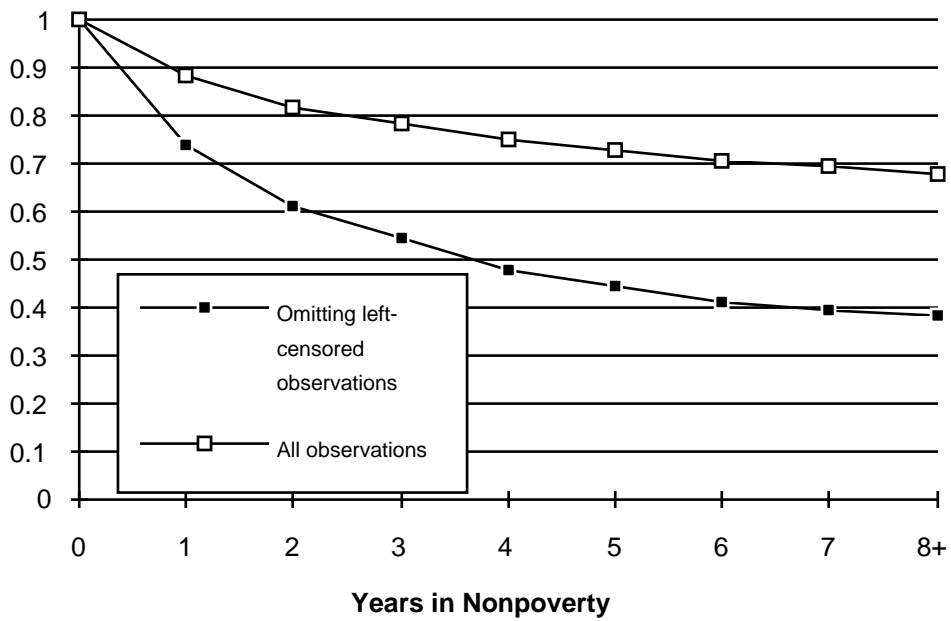
Figure 2 shows that there is an even greater disparity among nonpoverty spells by method of measurement. Spells of nonpoverty tend to be considerably longer if left-censored

<sup>2</sup> According to a two-tailed t-test, the difference between the numbers in rows 1 and 2 are statistically significant at 0.05 level. All other differences (between 2 and 3, between 3 and 4) are statistically significant at even 0.001.

**Figure 1. Poverty Duration Survival Function, By Method of Measurement**



**Figure 2. Nonpoverty Survival Function, By Method of Measurement**



observations are included (for any given year, a greater proportion of individuals are 'surviving' in nonpoverty spells). For example, after 7 years, 69.3 percent of people are still in the midst of a nonpoverty spell when all observations are included, versus 39.5 percent of people when left-censored observations are omitted. This again makes intuitive sense. By omitting left-censored observations, one tends to omit individuals who are experiencing long-term spells of nonpoverty and whose spell beginning occurred before the observation window. These individuals are more likely to never have been in poverty, and are probably less likely to ever fall into poverty.

The omission of left-censored cases poses a potentially bigger problem in the analysis of nonpoverty than poverty because long-term spells comprise a larger proportion of nonpoverty spells than of poverty spells. That is, a smaller fraction of people in poverty are in the midst of long-term spells than the corresponding fraction of people in nonpoverty.

### Individual- and Family-Level Characteristics of Left-Censored Spells

Nearly a third of all poverty spells in the sample come from left-censored spells. For all of these observations, we do not have a precise measurement for spell duration at the time of first observation. Again, the principle problem of merely discarding left-censored cases is that it will introduce selection bias into any subsequent model. This section examines the nature of this bias. More specifically, a logistic regression of left-censored spells on individual- and family-level characteristics is estimated in order to explore the characteristics of the individuals from which left-censored observations come. An outcome is coded at 1 if the spell was left-censored, 0 if it was not. Table 2 displays the results.

**Table 2. Logistic Regression of Left-Censored Poverty Spells on Individual and Family-Level Variables, 1970-1985**

| Covariates                   | Parameter Estimate | S.E.  |
|------------------------------|--------------------|-------|
| Intercept                    | 2.319 ***          | 0.189 |
| African-American             | 0.212 ***          | 0.062 |
| Female                       | 0.314 ***          | 0.062 |
| Age                          | -0.040 ***         | 0.002 |
| Number of children in family | 0.230 ***          | 0.017 |
| Female-headed family         | -0.006             | 0.088 |
| Dual-headed family           | -0.686 ***         | 0.079 |
| Years of education           | -0.199 ***         | 0.012 |
| N                            | 7,308              |       |
| -2 Log L Chi-Square          | 1065.657           |       |
| Degrees of Freedom           | 7                  |       |

\*\*\*P<.01

According to the results, left-censored observations tend to be associated with individuals who are economically disadvantaged and have longer poverty spells, as expected. Holding other factors constant, the odds are 24 percent greater that a poverty spell from an African-American individual is left-censored versus those from others, according to the odds ratio associated with that coefficient. Younger individuals are more likely to have left-censored spells; this may

partially result from the fact that all young adults who entered the sample at age 18 had left-censored observations (as defined in this study). We also see that poverty spells observed among women and less-educated individuals are also significantly more likely to be left-censored. Each additional year of education lowers the odds of observing a left-censored observation by 18 percent, according to the odds ratio associated with that variable. Finally, we also see that spells originating from individuals living in families which are dual-headed and have fewer children are *less* likely to be left-censored (the omitted category are individuals in families not dual-headed or female-headed: this includes male-headed families and individuals living alone or with unrelated people). In sum, this regression shows that omitting left-censored observations in a study of poverty will result in serious selection bias.

### **Discussion: Handling Left-Censored Data**

This analysis has shown that omitting censored cases could lead to serious selection bias by systematically discarding individuals in the midst of long-term poverty spells. These individuals also tend to be younger, less-educated, female, African-American, and live in families with more children. They are less likely to live in dual-headed families. Unfortunately, there is still no single accepted way of dealing with left-censoring. There are several methods, however, that may mitigate the problem. The 'best' one depends on the research question and patterns within the data itself. I now review a few of these methods.

One method-- which is still the most common-- is to simply ignore and discard left-censored data. In a study of poverty, this strategy is seriously flawed-- for reasons already discussed at length. It results in the omission of longer-spells, as well as spells coming from distinct demographic groups.

A second way of handling left-censored spells is to attempt to estimate their actual duration with statistical modeling. This often involves incorporating difficult to justify assumptions about the hazard rate of transitions (Guo, 1993). More specifically, an exponential hazards model, which assumes a constant hazard rate, can actually easily estimate left-censored poverty spell durations. However, the hazard for transitions out of poverty are certainly not constant (Bane and Ellwood, 1986; Stevens, 1994; Duncan and Rodgers, 1988; Iceland, 1996). The hazard rate declines the longer one is in poverty, though even the rate of decline is not necessarily constant (Duncan and Rodgers, 1988). Guo (1993) describes how left-truncated data can still be incorporated into statistical models-- if one knows at least when the start time for exposure began (perhaps with some minimal retrospective data). Again, with poverty data, exposure time often remains unknown. In sum, approaches that attempt to statistically model left-censored poverty durations are problematic.

A third way of dealing with left-censoring is to run a model both with and without left-censored data and discuss the substantive differences in the results (Moffitt and Rendall, 1993). Although this does not 'solve' the problem of censoring, it may still shed light on the issue of interest. Much of the analysis in the previous sections of this study, for example, hopefully shed light on the nature of poverty durations by comparing spell length calculated both with and without left-censored data. This type of analysis may in fact draw attention to some of the interesting complexities in poverty-related issues.

Another way to handle left-censoring is to refine the research question so that left-truncated data is no longer an issue. Duncan and Rodgers (1988) and Ashworth, Hill, and Walker (1994), for example, look at childhood poverty, where subjects are followed from birth. In such an analysis, the problem of having observations which have been exposed to the risk of an event for some time before observation becomes moot.

An analogous example of refining the data set, depending on the research question, is a study of poverty spell durations among all those who are (potentially) in the labor force-- individuals 18-64, for example. One can define the risk period as beginning at age 18. The researcher then chooses only individual observations which begin at that age. The first year observed can be considered non-left-censored-- as these people are being 'born' into the sample, in a sense. One potential problem with this strategy is that the precise research question one is interested in does not always lend itself to a refinement of the dataset which resolves left-censoring. Furthermore, the paring of the dataset may also result in a small sample size which may inhibit certain types of detailed or stratified-sample analyses. These may be issues of grave concern to the researcher.

A fifth way of handling left-censoring is applicable when one is trying to estimate the factors that impact on poverty spell durations or poverty transitions. In this type of analysis, the researcher can estimate simultaneous equations, where one derives the probability of observing a left-censored spell as a function of some parameters. The resultant parameter, an inverse mills ratio, is then entered into the main equation in order to control for sample selectivity (see Heckman, 1979). This strategy also has some disadvantages. In particular, it may be difficult to find a set a variables that affect poverty transitions only through the left-censoring variable, as such simultaneous equations require. For example, earlier we saw that left-censoring is affected by factors such as race, family structure, and education, to name a few. These factors also have a profound impact on poverty spell durations as well, and so one might not want to use them in the selection equation (as opposed to the main equation).

A final related method is most applicable when studying poverty *transitions*, using discrete-time logistic regression (see Allison, 1982). All observations from left-censored spells should be included in such a model to avoid selection bias. Dummy variables representing duration dependence are also needed-- though in such models precise data on durations may not crucial. Nevertheless, in order to control for the fact that some observations come from left-censored spells, a dummy variable indicating whether the observation is left-censored or not should also be included in the model. In this way individual- and family-level covariates can remain in the main model (versus using them in a selection equation). In other words, this method controls for duration dependence and unobservable differences of left-censored observations while still keeping important individual- and family-level covariates in the main model. It allows one to efficiently estimate the impact of the parameters of interest on poverty transitions with less bias and fewer selectivity problems than if left-censoring observations were simply ignored. The method one ultimately chooses must depend on the research questions posed and observed patterns in the data.

## References

- Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." In *Sociological Methodology*, ed. Samuel Leinhardt. San Francisco: Jossey-Bass Publishers.
- . 1984. "Event History Analysis: Regression for Longitudinal Event Data." *Sage Series: Quantitative Applications in the Social Sciences* 7, 46.
- Ashworth, Karl, Hill, Martha, and Robert Walker. 1994. "Patterns of Childhood Poverty: New Challenges for Policy." *Journal of Policy Analysis and Management*, 13, 4: 658-80.
- Bane, Mary Jo and David Ellwood. 1986. "Slipping Into and Out of Poverty: The Dynamics of Spells." *Journal of Human Resources* 21 (Winter): 1-23.
- ..... 1994. *Welfare Realities: From Rhetoric to Reform*. Cambridge, MA: Harvard University Press.
- Blossfeld, Hans-Peter, Hamerle, Alfred, and Karl Ulrich Mayer. 1989. *Event History Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Devine, Joel A. and James D. Wright. 1993. *The Greatest of Evils: Urban Poverty and the American Underclass*. New York: Aldine De Gruyter.
- Duncan, Greg and James Morgan. 1982. "Longitudinal Lessons from the Panel Study of Income Dynamics." *Review of Public Data Use*, 10: 179-84.
- Duncan, Greg J., and Willard L. Rodgers. 1988. "Longitudinal Aspects of Childhood Poverty." *Journal of Marriage and the Family*, 50 (November): 1007-21.
- Guo, Guang. 1993. "Event History Analysis for Left-Truncated Data." In *Sociological Methodology* ed. Peter V. Marsden. Washington, DC: American Sociological Association.
- Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47, 1 (January): 153-61.
- Hill, Martha S. 1992. *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park: Sage Publications.
- Iceland, John. 1996. "Urban Labor Markets and Transitions Out of Poverty." Paper presented at the 1996 PAA meetings in New Orleans.
- Mills, Kim I. 1995. "Census Report Shows Receding Poverty Levels." *The Boston Globe* (October 6): 3.

- Moffitt, Robert A., Michael S. Rendall. 1993. "Cohort Trends in the Lifetime Distribution of Female Headship in the U.S., 1968-1985." Working Paper 93-09, Population Studies and Training Center, Brown University.
- Pavetti, LaDonna. 1993. "The Dynamics of Welfare and Work: Exploring the Process by Which Women Work Their Way Off Welfare." Ph.D. dissertation, Harvard University.
- Sandefur, Gary D. and Nancy B. Tuma. 1987. "How Data type Affects Conclusions about Individual Mobility." *Social Science Research*, 16: 301-28.
- Stevens, Ann Huff. 1994. "The Dynamics of Poverty Spells: Updating Bane and Ellwood." *AEA Papers and Proceedings* 84,2 (May): 34-37.
- Trussell, James, and Charles Hammerslough. 1983. "A Hazards-Model Analysis of the Covariates of Infant and Child Mortality in Sri Lanka." *Demography* 20,1 (February): 1-25.
- Tuma, Nancy Brandon and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. Orlando: Academic Press, Inc.
- Yamaguchi, Kazuo. 1991. *Event History Analysis*. Newbury Park: Sage Publications.