

The problem of respondent attrition: survey methodology is key

Longitudinal surveys will suffer from attrition and nothing will change that; however, years of lessons learned in the field show that straightforward survey methodology can minimize the impact of losing respondents

Randall J. Olsen

The central problem of longitudinal surveys is attrition. The National Longitudinal Survey of Youth in 1979 (NLSY79), which this issue of the *Monthly Labor Review* features, is the gold standard for sample retention against which longitudinal surveys are usually measured. However, we cannot understand how the NLSY79 has done so well without considering what was done differently in the other cohorts of the NLS and what we have learned by formal evaluations of attrition aversion measures that evolved over a quarter century of field work. The lessons here are hard-won and, to some, unconventional.

Background

The NLS began in 1965 at the urging of an Assistant Secretary of Labor, Daniel Patrick Moynihan. He believed that although the Current Population Survey provided crucial snapshots of the Nation's labor force and labor market, the Nation needed a data source that was more dynamic and capable of tracking the long-run evolution of careers. The task of starting the study went to Howard Rosen at the Department of Labor, who enlisted Herb Parnes from Ohio State University, to assemble a team, design the surveys, and analyze the data. This team comprised representatives from the Census Bureau, Ohio State University, and the Department of Labor. The original plan was to follow the cohorts for 5 years to study some of the pressing questions of the time—the shrinking labor force participation rate of older men, the problem of youth unemployment and the transition from school to work, and the growing labor force participation of women whose children were entering school, leading to steady growth in the number of working mothers. Childcare was an important issue along with the problem of how the family would pay for a college education for the children of the baby boom.

Over time, the project has expanded. (Table 1 shows the various cohorts of the NLS, their start and stop dates, sizes, and age ranges covered.) Because the project began with a 5-year horizon, neither the Census Bureau, Ohio State, nor the Department of Labor had a plan for sample retention over the long run; after all, longitudinal surveys were still quite rare. The studies shortly proved their worth and the project became openended in terms of duration. However, the original limitations on intended duration led to some problems with attrition that conflicted with a revised plan to follow the respondents over the balance of their lives. In particular, the "following-rule"¹ that the Census Bureau used specified that when a respondent missed two consecutive interviews, the Census Bureau would drop that respondent from the study.

The following-rule and the original 5-year ho-

 Table 1.
 Survey groups, sample sizes, interview years, and survey status, National Longitudinal Survey, 1966–2004

Survey group	Age cohort at first interview	Birth year cohort	Original sample	Initial year/ latest year	Number of surveys	Number at last interview	Status
Older Men	45–59	4/1/06-3/31/21	5,020	1966/1990	13	2,092	Ended
Mature Women	30–44	4/1/22-3/31/37	5,083	1967/2003	21	2,236	Ended
Young Men	14–24	4/1/41-3/31/52	5,225	1966/1981	12	3,398	Ended
Young Women	14–24	1943–53	5,159	1968/2003	22	2,857	Ended
NLSY79	14–21	1957–64	² 12,686	1979/2004	21	³ 7,724	Continuing
NLSY79 Children	birth-14	(4)	(5)	1986/2004	10	³ 3,229	Continuing
NLSY79 Young Adults	615 and older	(4)	(5)	1994/2004	6	34,238	Continuing
NLSY97	12–16	1980–84	8,984	1997/2004	7	7,756	Continuing

¹ Interviews in 1990 also were conducted with 2,206 widows or other family members of deceased respondents.

² After dropping the military (in 1985) and economically disadvantaged non-black/non-Hispanic oversamples (in 1991), the sample contains 9,964 respondents eligible for interview.

³ The latest sample size available is from the 2002 survey. The 2004 survey is currently being fielded.

⁴ NLSY79 Children and young adults included by relation to NLSY79

rizon struck with the greatest force on the Young Men's cohort. In 1981, about two-thirds of the cohort responded to the survey. Some analysts believed that the rate of attrition reflected veterans' refusing to participate in a Government survey. Although the rate of attrition among black veterans was a few percentage points higher than that for nonveterans, for whites, the differential for veterans as a whole was essentially zero. Within the Young Men's cohort, blacks had the highest attrition rates. For whites, attrition in the Young Men's cohort was a bit higher than that for two women's cohorts, but male respondents have always had higher attrition than females.

The two-and-out-following-rule that the Census Bureau employed had serious ramifications, given the attrition pattern for the young men, and high attrition among blacks. By 1981, the Census Bureau had stopped tracking 11 percent of the young men because they had, at some point, missed two consecutive interviews. Blacks make up 28 percent of the young men's sample, but 57 percent of the cases dropped because of the following-rule were black. Our current ruleof-thumb is that in the next round, one can obtain an interview on about 25 percent of the respondents who have missed two interviews in a row. When interviewing began for the NLSY79, performance specifications did not allow respondents to be dropped simply based on consecutive missed interviews.

The original design for the surveys alternated in-person interviews with telephone and mail-out surveys, with the inperson version conducted every 5 years.² As a result, the content of the interview was more comprehensive every 5 females, not birth year. Children are still being born.

⁵ The size of the NLSY79 child sample depends on the number of children born to female NLSY79 respondents, attrition over time, and the gradual aging of the children into the young adult sample. The size of the young adult sample depends on the number of children who reach age 15 in each survey year.

 $^{\rm 6}$ In 1998 only, the young adults eligible for interview were limited to those ages 15 to 20.

years, with smaller updates in between. The NLS approach to the Mature Women's cohort is emblematic of the general approach of the survey. The 1967 interview of the women ages 30–44 focused on the longest held jobs between schooling and marriage, between marriage and the birth of her first child, and after the birth of the first child. The survey sought the most important (that is, longest held) job holdings, probed for significant periods not working, and ascertained why the woman did not work. The respondent answered CPS questions about the previous week; these questions accounted for a significant part of the interview.

This approach to collecting labor force behavior data left unanswered questions about work history, especially for women with frequent job transitions and women who missed the in-person interview. There were modules that collected retrospective data about fertility and marriage, but in the 1960s, marriages ending in divorce were less frequent, compared with current divorce rates. The NLS did not attempt to collect an event history on marriage, but nonetheless, the survey probably collected most of the transitions in marital status and cohabitation for the Mature Women and Older Men's cohorts.

The original cohort data collection effort frequently captured data on respondents' behavior by asking retrospective questions, sometimes at wide intervals, to capture particular data domains. For example, rather than collecting pregnancy roster data on the Young Women's cohort as those events occurred, the NLS would ask about many years' experience all at once. As Frank Mott documents, this strategy for data collection opens the way for more measurement error.³ With the strategy used then, missing one interview can leave an important part of the data record distressingly incomplete.

It is in this context that we start this article by focusing on the historical record of the completion rates for the various cohorts of the NLS and how the strategy for both data collection and the rules for continuing to follow nonrespondents generate startling impacts on the completeness of the data coming out of a longitudinal study. This article continues by describing some of the fielding techniques the NLS program has employed to offset the secular trend toward lower completion rates.

The historical record

The remainder of this article describes the two original women's cohorts: the NLSY79 and the NLSY97. The two original men's cohorts were cancelled in the early 1980s.⁴ In 1981, the Census Bureau completed interviews of 65 percent of the original respondents for the Young Men and 52.5 percent of the Mature Men. However, corrected for mortality, the numbers are higher, with 66.8 percent completed of the respondents still alive for the Young Men and 74.8 percent of the Mature Men. After 15 years, the completion rate for the Mature Women was 69.7 percent (73.5 percent of those still alive), and for the Young Women it was 68.8 percent (69.4 percent of those still alive). As mentioned earlier, the lower completion rate for the Young Men reflects a following-rule that dropped blacks at an unusually high rate.

The Census Bureau experience with the original cohorts was more favorable than its recent experience with respondents from the Survey of Program Dynamics (SPD). That survey continues the 1992-93 Survey of Income and Program Participation (SIPP) panel. By the end of the SIPP phase for the SPD respondents (the SIPP phase contained nine waves each 4 months apart), the completion rate was about 73.4 percent. The 1997 wave of the SPD completed 58.7 percent of the SIPP respondents, and by 2002, the completion rate stood at 53 percent of the 1992–93 SIPP respondents. The 10-year retention rate for the SIPP/SPD panel is below the 15-year retention rate for the original cohorts, whether or not one corrects the latter retention rates for mortality. A more striking contrast is the experience with the NLSY79, for which the National Opinion Research Center (NORC) does the data collection. The 15-year retention rate from 1979 to 1994 was 89.2 percent-almost 20 percentage points higher than the rate for the original cohorts. The disparity of experience with retention in longitudinal studies grows when we compare the NORC experience with the NLSY97 to their experience with the NLSY79. After seven rounds, NLSY97 retains 86.4 percent of the respondents, which is below the NORC experience in the NLSY79 after more than 19 years. Even if we examine the

records of the Census Bureau and NORC separately, we see marked variation in outcomes. The reason for this disparity is complex.

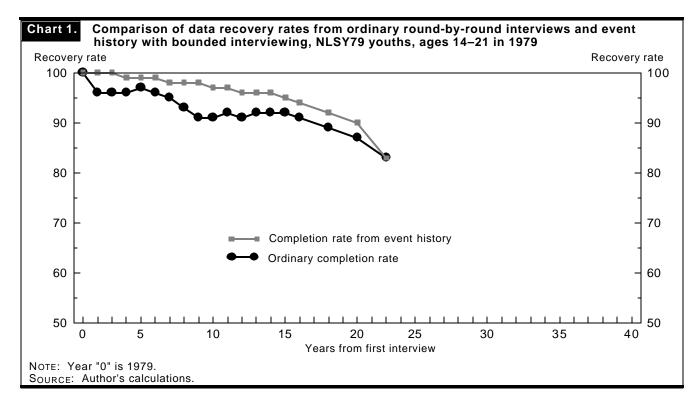
Bounded interview event histories

The most important feature that distinguishes the NLSY79 (and NLSY97) from the initial design of the original cohorts of the NLS is the use of event histories in the NLSY79 and NLSY97. The article by James Walker in this issue addresses this feature in more detail, but when the case for event histories in the NLSY79 was made by Burton Singer, he had in mind their analytic usefulness and not their effect on attrition. (See accompanying articles in this issue.) As it turns out, event histories in longitudinal surveys force us to rethink our views on attrition.

As implemented in the NLSY79, event histories carried forward respondents' answers to questions from the previous interviews in some domain. For example, if the respondent was interviewed on July 21, 1987, and he was married, we would ask, "According to our records, when we last interviewed you on July 21, 1987, you were married. Is that correct?" If the respondent agrees, he would be asked whether his marital status has changed since that date, and if so, when that happened, what the change was, and the characteristics of that transition, such as the demographics of the new spouse, and so forth.⁵ If the respondent disagrees with the data carried forward in our records, the interviewer obtains the corrected data and then carries the event history collection forward from the point of correction.⁶ This approach allows one to deal with the "seam problem," namely how to deal with recollections of the same event whose timing differs across survey waves. Indeed, when the NLSY79 switched to computer-assisted interviewing and, at the same time, to collecting a true event history for education, the incidence of seam problems declined dramatically.

Perhaps more importantly, the use of bounded interview event histories makes the data collection protocol less dependent on the interview date. If a respondent misses an interview, at the next round he is asked to pick up the collection of the event history on, say, marital status at the point he left off on the most recent completed interview. We recover event history data from a respondent whenever he or she returns to the survey. This approach generates a substantially more complete data history for respondents than suggested by a simple examination of completion rates.

Chart 1 illustrates the effect of returning to respondents. The lower line is the round-by-round completion rate for the NLSY79 from the 1st through 20th rounds. The higher line shows the fraction of the data for the year preceding each of the 20 rounds that we recovered either from an interview in that or a subsequent round. Because we fill in missing data



for the event histories whenever we interview the respondent, analysts have well in excess of 90 percent of the data for 20 years after the initial interview. Bounded interviewing requires a sophisticated instrument, adds to interview length, and works best with salient events. With long retrospectives, accuracy may suffer, but missing data are the worst data. In the NLSY79 and NLSY97, we employ event histories for employment, marriage and cohabitation, fertility, training, education, and program recipiency. These histories are the core of the NLS.

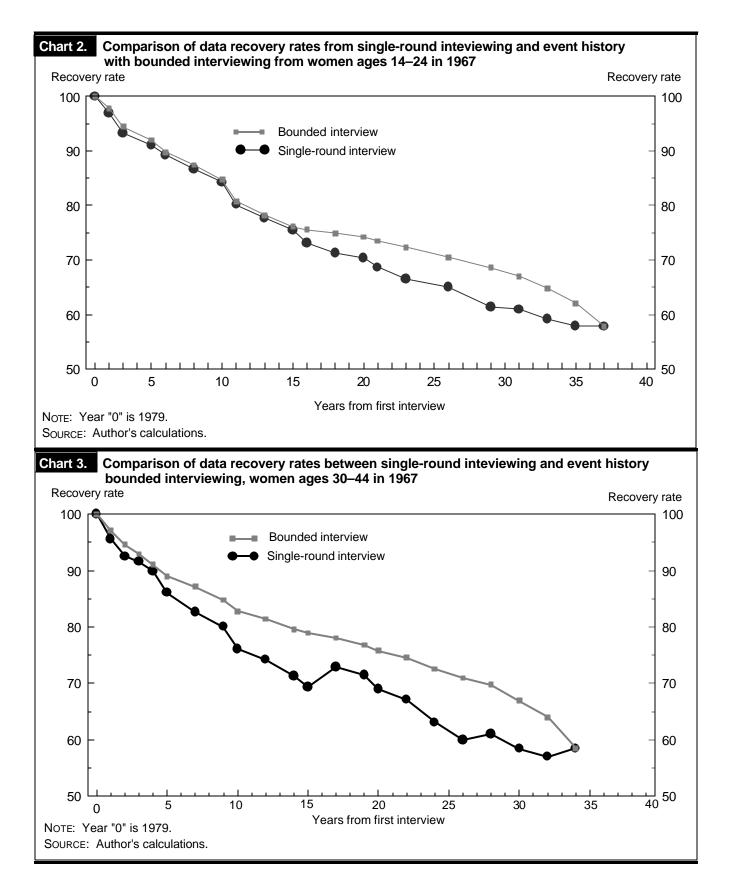
Using bounded interviewing event histories generates a more complete data record, especially when combined with a field strategy that emphasizes returning to past nonrespondents. Charts 2 and 3 illustrate how the decision on which respondents to contact in a longitudinal survey is crucial to generating a complete data record.

In the Young Women's and Mature Women's cohorts, the Census Bureau originally adopted the strategy that it would not return to any respondent who refused two straight interviews. In the mid-1980's, the Census Bureau changed that strategy. Starting in 1985, the Census Bureau returned to past nonrespondents to the Young Women's survey, and in 1986, the Census Bureau no longer dropped respondents missing two straight interviews, but did not return to past nonrespondents.

Chart 2 shows that the rate of attrition slowed after the Census Bureau no longer dropped nonrespondents after they missed two straight interviews, effective in 1985. The event history completion rate naturally shows less data loss once the Census Bureau changed their following-rules. Chart 3 illustrates the impact of the Census Bureau's decision in 1986 to return to some (but not all) of the Young Women who had been dropped from the survey roster after two straight noninterviews. The completion rate jumped up when the Census Bureau started the new following-rules, and in subsequent rounds, the completion rate dropped more slowly. The impact on the event history completion rate of former nonrespondents, (especially in contrast to chart 2 for the Mature Women), provides stark evidence that an aggressive following-rule plus bounded interviewing event histories are valuable data-collection strategies for the NLS. Next, we turn to encouraging the respondents to give us more interviews.

Encouraging respondent cooperation

The NLS has a life-cycle perspective; it tries to follow respondents from their first interview to the end of their life. Sometimes funding constraints limit how long it can follow a cohort, but the project tries to keep its focus on how lives of individuals evolve in their entirety. This scientific agenda requires patience and a focus on long-term cooperation. The other long-term survey best known among social scientists is the Panel Study of Income Dynamics (PSID). That study focuses on the household rather than on the individual, so if we lose the cooperation of one reference person for the household, we seek another informant for that household. In



the case of the NLS, we only use proxy respondents for ill individuals or in other unusual cases. This puts a premium on retaining the good will of respondents. We seek the goodwill of the respondents in three major ways: informing the respondent about survey objectives and findings, using incentive payments, and employing the mode of interview that the respondent prefers.

Fundamentally, to gain respondent cooperation, field staff must stress that the study is important, if not for the respondent, then for society at large. The questions have to be valid and must relate to the objectives of the study that form the basis for its societal usefulness. For the doubting respondent, the study must provide evidence that it is substantiated by findings, recommendations, or a body of research consistent with the message that it is important. Interviewers play a pivotal role in the "selling of the survey." Thus, a less appreciated corollary of convincing respondents that the study is important is that the *interviewers* must be convinced the study is important. What good is a salesperson who does not believe in the product?

The NLS program mails out informative brochures to respondents and provides the interviewers with packets of materials that provide support for the central thesis that the study has societal importance. It also provides items to rebut common objections to cooperating with the survey. However, the most important tool we have is the enthusiasm and commitment of the interviewers. We communicate this in two forms—first by engaging and "selling" the interviewers when they attend training, and second through their interaction with the field managers who have day-to-day responsibility for overseeing the field work. Tepid and boring training sessions for interviewers are not only ineffective venues for learning, but they communicate a subliminal message to interviewers that the larger survey is equally boring and, hence, unimportant.

Although we attempt to communicate the scientific importance and social utility of the survey, some respondents remain unconvinced. Respondents are not monolithic. About half the respondents are extremely cooperative and easy to work with. As of 2002, almost exactly 60 percent of the respondents in continuing sample types⁷ had completed every interview. There is also a core group of very disaffected respondents for whom we have little hope they will complete the survey, although from time to time a few rejoin. For these hard-core refusers there is not much that will make a difference in their cooperativeness, at least nothing we have been able to identify and use. However, for a sizeable minority of the sample, respondent incentives, either money or in-kind, can play a useful role in securing cooperation with the survey. Small gifts that are tailored explicitly to the interests and situation of the respondent can be very effective-they say in a tangible way that we care about the respondent and pay

attention. Although incentive payments likely influenced some of the respondents who have remained with the survey throughout, incentives appear to work on a minority of respondents—a fact that needs to be integrated with any plan to use incentives in longitudinal surveys.

In 2000, we conducted an experiment, with randomization of subject and treatment, investigating the effects of a \$40 versus \$80 incentive. With the two incentive amounts, we computed how much additional money we spent to obtain an additional interview. On the one hand, when additional incentives were offered to respondents not interviewed in 1998 (the previous round-the NLSY79 is now administered every 2 years), we made \$167 in additional incentive payments, on average, for every additional interview when increasing the payment from \$40 to \$80. On the other hand, for respondents who had cooperated in 1998 we paid an additional \$264 for each additional interview. This experiment was conducted when we had encountered strong resistance and sharply higher field costs when using an across-the-board \$20 incentive. Had the experiment been mounted at the beginning of the field period, the higher payments would have been even more cost-ineffective.

Across-the-board incentive payments increase response rates at a substantial monetary cost. However, one unexpected result of the experiment in 2000 was a reduction in field costs that coincided with the higher incentive payments—field costs fell in a manner we had not seen in previous rounds. The savings in field costs very nearly offset the higher respondent fee costs. Survey organizations should consider the effects of incentive fees, not only on respondent cooperation, but also on the overall cost-effectiveness of a survey campaign.

The most frequent objection to differential fees cites equity issues. This objection, together with the surprising finding in the 2000 round that higher fees reduced field costs, led us to experiment in 2002 with a strategy aimed at using higher incentives with cooperative respondents. For this experiment, called the "Early Bird," we mailed a flyer to respondents offering a higher fee if they called in to make an appointment for an interview. Because interviewers expend substantial time locating, contacting, and making an appointment with respondents, the "Early Bird" offer saves money despite the higher respondent fee involved. Once the conventional field period begins, the offered incentive payment reverts to the normal level.8 Respondents requesting a higher fee are encouraged to participate in the Early Bird programinterviewers redirect requests for a higher payment toward enrolling the respondent in a mutually advantageous compact exchanging more cooperation for a higher fee.

Although incentive payments encourage some respondents to participate, these payments cannot, within a reasonable range, convert all or even nearly all nonrespondents. Fees work for some and fail for others. In addition, for some groups of respondents, such fees can be very expensive in terms of how many cases the extra money spent produces. In short, respondent reactions to incentive fees are remarkably heterogeneous.⁹

Heterogeneity also rules when it comes to the interview mode that respondents prefer. The survey choices available to respondents are via telephone, in-person interview, or, to a lesser degree, Webpage. However, the NLS surveys are too complex to encourage most respondents to take a mail survey or a Web survey-most interviewers require about 2 days of training on the survey before they are ready to administer it. Some respondents, having done the survey 20 times before, are more "experienced" with the survey than new interviewers, but there are sections of the interview where interviewer training plays an important role. For this reason, the NLS program does not routinely offer respondents the choice to do the interview over the Web, although we could offer this option.¹⁰ When it comes to telephone versus in-person interviews, some respondents insist on a telephone interview and others insist on in-person interviews. The initial approach to each respondent is usually over the telephone, except in unusual circumstances. At that point, respondents for the NLSY79 can choose how they want to do the interview. With the NLSY97, the survey contains a substantial self-administered section containing sensitive questions, and interviewers emphasize in-person interviews. The latter constraint generates comments from the field staff indicating the emphasis on in-person interviewing tends to increase costs; however, if the choice is between a telephone interview and no interview, the field staff always goes for the interview. The essential point is that respondents vary in their preferences, and insisting on either telephone or in-person interviews carries substantial risk of alienating respondents.

In this same vein, organizations conducting longitudinal

surveys often question whether having the same interviewer do a case year after year encourages response.¹¹ Knowing whether continuity in the assigned interviewer encourages response would allow a field organization to assign cases more strategically and to make a more intelligent trade-off between reducing travel costs and reducing assignment turnover. Our ability to infer the relationship between attrition and interviewer turnover is reduced by the fact that field operations rarely employ random assignment of interviewer and respondent to a dyad. If one interviewer has trouble with a case, we assign a new interviewer and hope for better results. In short, random influences that make a respondent less likely to do an interview can also lead to a change in the assigned interviewer. Indeed, a case may have several interviewers assigned to it over the course of the field period. Because of these problems, we look at respondent attrition using variables that summarize his or her past tendency to do the survey, his or her attitude at the most recently completed interview, and respondent age. To capture the effect of interviewer continuity, we use two additional variables, whether the interviewer doing the respondent's most recent interview is still on staff and, if so, how many times that interviewer had interviewed the respondent. These two variables only measure the ability of the field staff to exploit interviewer continuity in the current round, not whether interviewer continuity held for the current round.

The results, in table 2, using 202,245 observations from rounds 2 through 20 of the NLSY79, suggest that interviewer continuity is not a major factor. There is a net advantage to interviewer continuity after the respondent has been interviewed twice by the same interviewer, and after that, having the same interviewer decreases attrition by about 0.7 percent for each additional round. If an interviewer had interviewed the respondent for the previous round, having that interviewer on staff again generates no positive effect (for the

Variable	Coefficient	Effect ¹ (dP/dX)	Standard error	t-ratio
Sample size, N=202,245				
Intercept	-2.4002		0.2219	10.82
Respondent did round before?	2.8604	0.212	.0296	96.64
Previous interviewer on staff?	0914	007	.0289	3.16
Previous interview by phone?	5067	037	.0256	19.79
Number of interviews done by previous				
field interviewer, previous interviewer on staff	.0881	.007	.0089	9.90
Percent of previous interviews done	3.8157	.282	.0721	52.92
Respondent hostile at previous interview?	-1.5691	116	.0917	17.11
Respondent impatient, restless?	9364	069	.0421	22.24
Respondent cooperative, not interested	5107	038	.0249	2.51
Age R	.0142	.000	.0141	1.01
Age R squared	0011	.000	.0002	5.50

¹ The effect on the probability of an interview on a unit of change in the explanatory variable.

NLSY79, if the same interviewer is available, she will most likely have any of her previous cases that are in the same geographic area).

Also in table 2, the estimated model is a logit. The third column gives the effect on the probability of an interview of a unit change in the explanatory variable. The standard error is for the coefficient. Results show that it is, by far, more important to keep the respondent interested in the project and happy with how they are treated than to keep the same interviewer. Of course, interviewer continuity may make it easier for the field staff to remember how best to deal with a particular respondent.

LONGITUDINAL SURVEYS WILL SUFFER from attrition and nothing will change that. However, certain survey methodologies can minimize the impact of attrition. First, if consistent with its objectives, the survey should utilize event histories to recover data not collected when a respondent misses a round.

Notes

¹ This is the rule determining to which respondents interviewers would return in the event the respondent did not complete one or more interviews.

² Census interviewed the young men and young women in person for several consecutive years and then reverted to the interview pattern for the two older cohorts—personal, skip a year, telephone, skip a year, telephone, personal. In the late 1980s the women's interviews were done in person, although many cases were done over the phone when circumstances dictated.

³ See Frank Mott, "Looking Backward: Post Hoc Reflections on Longitudinal Surveys," in Erin Phelps, Frank Furstenberg, and Anne Colby, eds., Looking at Lives: American Longitudinal Studies of the Twentieth Century (New York, Russel Sage, 2002).

⁴ The two original men's cohorts—NLSY79 and NLSY97 were cancelled in the early 1980s.

⁵ With the rise of nonmarital unions, event histories on marriage and cohabitation have become more complex. The approach of asking the retrospective question with explicit reference to the respondent's previous answers is referred to as "bounded interviewing." Second, the following-rules must emphasize persistence. If a respondent refuses a round, return in the next round. When respondents miss a round, in about half the cases they will grant an interview for the next round. If they miss two straight interviews, the probability of success drops to about 25 percent, but certainly not to zero. Third, targeted incentive payments should be used because they are cost effective ways of holding attrition in check. Fourth, allow respondents to choose the interview mode. Just as differences in respondents make incentives effective only for some respondents, differences among respondents make it important to acquiesce to their preferences over interview mode (phone versus personal or even Web). Finally, longitudinal surveys must be "sold." They must be sold to the interviewers who face the job of convincing the respondent that the survey is important, and they must be sold to the respondent who, in the majority of cases, will offer their cooperation so long as the study engages their attention and they are confident that they are providing their time for a worthy endeavor.

⁶ The discrepancy is handled at Ohio State. This style of data collection generates fewer "seam problems" than histories that are not collected using bounded interviewing. Our current practice is to accept the date given initially and place the "seam" on the day after the date of interview.

⁷ The low-income, non-black, non-Hispanic oversample was nearly eliminated in 1993 and most of the military oversamples were dropped earlier.

⁸ Currently the normal incentive is \$40.

⁹ In the past 15 years, we have also experimented with monetary incentives to interviewers. Unfortunately, those experiments failed. At best, they simply shifted the timing of when cases were completed.

¹⁰ Our CAPI system is based entirely on Web browsers. Some interviews done over the phone utilize the Web to present the interview to the interviewer, so offering a Web interview to respondents is a minor step.

¹¹ The other side of repeated contact is that the respondent may be more likely to give normative responses to an interviewer once they have established rapport.