Center for Demography and Ecology

University of Wisconsin-Madison

Comparing Data Quality of Fertility and First Sexual Intercourse Histories

Lawrence L. Wu

Steven P. Martin

Daniel A. Long

CDE Working Paper No. 99-08

COMPARING DATA QUALITY OF FERTILITY AND FIRST SEXUAL INTERCOURSE HISTORIES

Lawrence L. Wu, Steven P. Martin, and Daniel A. Long
Department of Sociology and
Center for Demography and Ecology
University of Wisconsin-Madison

October 1998

[revised April 1999]

Paper presented at the conference, “Data Quality in Longitudinal Surveys,” sponsored by the Panel Study of Income Dynamics, the National Longitudinal Surveys, and the Health and Retirement Study. We gratefully acknowledge research funding from NICHD (HD 29550), the Institute for Research on Poverty, and the University of Wisconsin Graduate School, with additional support from a core grant to the Center for Demography and Ecology (NICHD HD 05876). We thank Tom MaCurdy for helpful comments and Nora Schaeffer for numerous discussions in which she generously shared her expertise. Direct all correspondence to Lawrence L. Wu, Department of Sociology, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706, lwu@ssc.wisc.edu.

ABSTRACT

This paper evaluates the data quality of two demographic variables in light of hypotheses on respondent recall from the literature on survey methodology. An emerging consensus in this literature is that recall of the timing of an event declines with duration of recall unless the dating of an event is frequently "rehearsed." We provide empirical evidence consistent with this hypothesis by assessing the quality of demographic data supplied by female respondents on selected event history outcomes in multiple nationally representative surveys. A first demographic variable concerns the interval between a first and second birth. We compare second birth intervals using birth registration data from the Vital Statistics on Natality (VSN) and individual-level survey data from the June Current Population Survey (CPS), the 1979−94 waves of the National Longitudinal Survey of Youth (NLSY), and the 1988 National Survey of Family Growth (NSFG). Despite marked differences in survey design, we find relatively few differences in the quality of birth interval data across these four surveys. A second demographic variable is age at first sexual intercourse. We engage in two sets of analyses of this variable. First, we use data from the NLSY to analyze discrepancies between successive reports on age (to the nearest year) at first intercourse. Second, we analyze a form of partially missing data (respondent inability to recall the calendar month of intercourse) that occurs in both the NLSY and NSFG. In both sets of analyses, we find that data quality varies significantly with duration of recall and measures of respondent ability related to arithmetic facility and memory. Observed differences by race and ethnicity narrow substantially when controlling for these and other background factors. We find evidence for a nonlinear association between duration of recall and data quality, with similar nonlinear patterns observed for both NLSY and NSFG respondents.

Sample surveys have increasingly included event history data on a variety of demographic outcomes. Surveys containing demographic questions often routinely ask respondents to report dates of birth for all biological children. Increasing numbers of surveys have obtained data on the timing of entry and exit from marital and cohabiting unions, including the timing for the dissolution of formal and informal unions (Sweet, Bumpass, and Call 1988). Surveys such as the 1995 National Survey of Family Growth have significantly expanded the types of event history data collected (Abma et al. 1997), including data providing rough schooling attendance and attainment histories, including spells not in secondary school, the timing of entry and exit from post-secondary institutions (e.g., college, vocational training), and dates attached to important degrees (high school diploma, GED, post-secondary degrees); data on the onset of sexual activity, as proxied by age at first sexual intercourse; spells of sexual activity, contraceptive use and nonuse, and, for contracepting individuals, the type of contraceptive method used; pregnancy and pregnancy resolution histories, including data on the timing of abortions; and data on fertility cessation, including the timing of contraceptive sterilization for men and women and the timing of onset of menstruation for women.

A growing consensus among survey methodologists is that event history data impose unusual demands on sample respondents by requiring responses on both the occurrence and timing of events (Schwarz and Sudman 1994). Recent changes in two longitudinal surveys, the National Longitudinal Survey of Youth 1979 and the Panel Study of Income Dynamics, from an annual to biannual interviewing schedule, have added to concerns about the quality of event history data obtained using traditional survey instruments. Many studies in the survey methodology literature directly address these issues, but findings in this literature are typically based on small-scale surveys or experiments.

This paper complements these studies by examining fertility-related outcomes in several nationally representative surveys. We examine two fertility-related outcomes—age at first sexual intercourse and the interval between a first and second birth. We begin by comparing birth intervals using data from the Vital Statistics on Natality (VSN), the June Current Population Survey (CPS), the 1979−94 waves of the National Longitudinal Survey of Youth (NLSY), and the 1988 National Survey of Family Growth (NSFG). We find little evidence of duration heaping in any of the surveys we examine.

We then turn to data on age at first sexual intercourse from the NLSY and NSFG. We provide two analyses of age at first sexual intercourse. First, we use data from the NLSY to analyze discrepancies between succesive reports on age (to the nearest year) at first intercourse. Second, we analyze a form of partially missing data (respondent inability to recall the calendar month of intercourse) that occurs in the NLSY and NSFG. In both cases, the data we analyze are unusual in that they provide both rough measures of data quality for an event *and* a rough measure of the duration of time that has elapsed between the event and interview. We exploit these properties to examine how respondent ability to recall detailed information on the timing of an event varies with the duration of recall and other observed attributes. Our results suggest nonlinearities in respondent recall with duration, with similar nonlinearities observed in both NLSY and NSFG. Our findings also suggest a significant association between data quality and measures of respondent ability related to general memory retrieval and arithmetic ability.

RESPONDENT RECALL OF THE TIMING OF EVENTS[1]

Event history data impose unusual demands on sample respondents. Even simple event history data (for example, an item on first marriage) requires responses on both whether an event occurred and the timing (often to the nearest month) of an event. In more complicated settings, obtaining event histories requires respondents to report on the timing and sequencing of multiple events, for example, the timing and sequencing of changes in contraceptive methods (Martin and Wu 1998), or the timing and seqeuencing of changes in the numbers and types of parental figures present while the respondent was growing up (Wu and Martinson 1993; Wu 1996). The complexity of these data,

[1]This discussion is adapted from a review in Schaeffer, Thomson, Dykema, and Wu (1998).

and the corresponding ability of respondents to comprehend and respond to such complexities, give rise to issues concerning the quality of event history data obtained in standardized survey settings.

In general, demographic researchers have held two presumptions with respect to the likely quality of event history data obtained under alternative sample designs. A first is that, all else being equal, the quality of event history is lower in an omnibus survey that obtains data across a variety of substantive domains than in a single-purpose survey that obtains data concentrated in a few substantive domains that are related in content. Examples of single-purpose surveys include the National Survey of Family Growth for demographic variables and the Consumer Expenditure Survey for economic variables related to the consumption of individuals. Examples of omnibus surveys, at least from the standpoint of the collection of data on demographic variables, include the National Longitudinal Survey of Youth (NLSY), the Panel Study of Income Dynamics (PSID), the various June Current Population Surveys (CPS), and the Health and Retirement Survey.

A second presumption is that, all else being equal, the quality of event history data is higher in a prospective survey than in a retrospective survey. Examples of prospective designs include the NLSY and PSID, while examples of retrospective designs include the CPS and NSFG. It should be emphasized that the terms “prospective” and “retrospective” are somewhat misleading in that both types of surveys require retrospection by respondents. The key distinction is that prospective surveys employ shorter periods of recall, with the recall period typically tied to the duration (often one or two years) between successive survey interviews, while retrospective surveys may employ long periods of recall.

These presumptions are mirrored in arguments found in the survey methodology literature. Empirical findings in this literature concerning the accuracy of respondent self-reports on the occurrence, frequency, and dating of events are primarily based on studies using experiment, diary, and “record-check” designs. Studies using diaries and record checks are of particular interest for our purposes. Subjects in diary studies are typically asked to keep a diary; at some later point, respondent recall on specific items is compared to diary entries (Smith and Jobe 1994;

Thompson, Skowronski, Larsen, and Betz 1996). Subjects in record-check studies are interviewed, with data provided in the survey interview compared to records in independent, external data sources. Examples include comparisons of self-report data on smoking cessation with records obtained from smoking-cessation programs (Means et al. 1991). Findings from such studies have led survey methodologists to argue that respondent responses in the standardized survey setting are best conceptualized as a *process* in which the survey respondent attempts to retrieve selected memories in response to interviewer questions (Sudman, Bradburn, and Schwarz 1996). Because respondent recall requires motivation, effort, and time on the part of both the respondent and interviewer, memory retrieval may be aided when questions are related in substantive content or when a survey concentrates on a few substantive domains.

Another consistent finding is that the accuracy of respondent reports for events varies inversely with the frequency of similar events. Large-scale record-check studies have compared self-reports about spells of unemployment with company records (Mathiowetz and Duncan 1988) or about amounts of child support owed and paid with court records (Schaeffer 1994; Dykema 1996). Findings from these studies are consistent with smaller scale laboratory work (Menon 1993, 1994), which show that when events are simple (constant, episodic, or regularly patterned), reports in interviews are fairly accurate. Thus, respondents who were employed continuously during a reference year period or who received no child support are likely to be fairly accurate in their answers concerning income or child support receipt. But when events are complex, irregular, or variable in nature, accuracy of self-report data declines (Schaeffer 1994; Dykema 1996).

A second finding is that data quality suffers with the passage of time, with some authors arguing that accuracy declines exponentially with time (Neter and Waskberg 1964; Mathiowetz and Duncan 1998). So-called “flashbulb” memories concerning prominent or “landmark” events were once thought to be exceptions to general tendency for recall accuracy to decline with time. However, more recent studies suggest that even these memories can be prone to error. For example, Neisser and Harsch (1992) obtained two reports by subjects about where they were when they heard about the Challenger space shuttle explosion, with a first report obtained within 24 hours of the disaster and a second report obtained two and a half to three years later. In comparing contemporaneous and later reports, Neisser and Harsch (1992) found that only 7% remembered the circumstances exactly, with 25% supplying completely inaccurate accounts. Similarly, in one diary study, subjects could date about half of the events that occurred within the previous 100 days accurately, but accuracy of events declined to under 20% for events that were about a year old (Thompson et al. 1996). It is notable that this degradation occurred even though the experimental design (i.e., the process of keeping the diary) might have, on average, improved respondent recollection of the events.

These results have led survey methodologists to conclude that the dates of events are not routinely stored in the memories of individuals; instead, it is argued that there are several distinct types of memory, which are stored and accessed in different ways that have implications for the accuracy of respondent reports (Brewer 1994). A particularly useful distinction is between event dates that are frequently "rehearsed" and those for which event dates are reconstructed by the respondent during the interviewing process. Examples of dates that might be frequently rehearsed include landmark events such as the birth dates of one's children, marriage anniversaries, when a couple first met, and so forth. However, when the date of an event is not frequently rehearsed, it is argued that respondents reconstruct data by employing various heuristics or simplifying strategies to construct answers during the interview (Bradburn, Rips, and Shevell 1987; Blair and Burton 1987). For example, Schaeffer (1994) and Dykema (1996) interpret evidence showing that when child support payment patterns are relatively constant or regularly patterned, respondent reports are fairly accurate as consistent with the view that the heuristics used by respondents work well in such situations. Conversely, they find that when events are complex (e.g., when child support payments are irregular or vary in important ways), accuracy of respondent reports declines, a result again consistent with the view that respondents employ heuristics when responding to survey items. Similarly, Brewer (1994) and Thompson et al. (1996) argue that the dates for many events are reconstructed from several kinds of information, for example, information (sometimes including visual images) of the people, locations, and actions that occurred contemporaneously with the event in question. Unfortunately, the peripheral details used to reconstruct when something "must have happened" appear to be forgotten relatively quickly (Thompson et al. 1996). Moreover, the cognitive processes involved in retrieving such personal memories do not appear to be mirrored by cognitive processes that store the date or timing of the event in question.

DATA

In this section, we describe our data sources used to assess data quality for the interval between a first and second birth, and age at first sexual intercourse. Throughout, we limit our investigation to data provided by U.S. women. It is important to note that our assessments of data "quality" rely heavily on measures of presumed data consistency. For event history data on first sexual intercourse, it is difficult to assess data quality against some "external" standard; hence, one is typically forced to assess data quality by examining the internal consistency of data. For fertility histories, comparison against external standards (e.g., registers of natality statistics) are feasible. Because individual-level natality data are readily available, we are able to compare birth interval data from the CPS, NLSY, and NSFG against official birth registers. By contrast, our assessments of data quality for data on age at first sexual intercourse are more indirect and rely on (a) analyses of factors associated with discrepancies between respondent reports in successive interview waves of the NLSY, and (b) analyses of the factors associated with the prevalence of partially missing data (missing calendar month of first intercourse) that occurs in both the NLSY and NSFG.

Our analyses of fertility histories examine duration heaping on the interval between a first and second birth obtained from female respondents in four nationally representative data sources. One data source, the 1990 Vital Statistics on Natality (VSN), provides individual-level birth registers. Two data sources employ retrospective designs: the June 1990 Current Population Survey (CPS) and the 1988 National Survey of Family Growth (NSFG). A final data source, the 1979−93 National Longitudinal Survey of Youth (NLSY), employs a prospective design, following a nationally-representative sample of individuals aged 14−21 at initial interview in 1979. In our analyses of birth intervals (but not in our analyses of missing data of age at first intercourse), we have restricted the samples in the CPS, NSFG, and VSN to match the ages of NLSY women in 1993 to aid comparability of results across surveys.[2]

Data on age at first sexual intercourse are obtained from the NLSY and NSFG. Both surveys thus provide retrospective data on age at first sexual intercourse. Data on the calendar year and month of first sexual intercourse were gathered by NSFG in 1988 and by NLSY in 1983−85. In the NLSY, the 1983 wave obtained age at first intercourse to the nearest year; this item was repeated to all female respondents in the 1984 and 1985 waves, but obtained age at first intercourse to the nearest month. In both surveys, a screener asked if the respondent had ever engaged in sexual activity; if so, the respondent was asked to supply the calendar year and month of first intercourse. While 98% of NLSY and 95% of NSFG women were able to provide the calendar year of first sexual intercourse, substantial numbers (16% and 20% in the NLSY and NSFG, respectively) were unable to recall the calendar month of first intercourse. Thus, both the NLSY and NSFG contain a form of partially missing data—missing calendar month of intercourse. In addition, the NLSY lets one identify discrepancies in respondent reports on age at first sexual intercourse (to the nearest year) using multiple respondent responses in the 1983−85 waves.

Vital Statistics on Natality. The 1990 Vital Statistics on Natality provide individual-level birth register data for all U.S. births that occurred during calendar year 1990. An advantage of these data are that they provide national birth registers and hence very large sample sizes; their disadvantage is that the covariate information contained in these data are restricted to items such as respondent's age, race and ethnicity, whether the respondent is married, live birth order, months

[2]NLSY respondents were approximately 29−37 at the 1993 wave, and we imposed this age restriction (at interview) on our CPS, NSFG, and VSN samples. We imposed these restrictions to make years of exposure roughly equivalent across surveys; note, however, that they yield CPS, NSFG, and VSN cohorts that were born a few years earlier than the NLSY birth cohorts.

since last birth, gestational period, and birthweight. These data allow us to identify U.S. women with a second birth during 1990. We further restricted the universe of all second births recorded in the 1990 VSN by retaining second births only for those women aged 32 or younger in 1990. This sample restriction yielded a sample of 1,128,726 second births. Of these second births, 13,342 were multiple births (e.g., twin births to a nulliparous woman); hence, these were coded by DHHS as births with a zero months duration between births. For analytical convenience, we ignore such births in our analyses. Another 62,240 cases (or approximately 6% of second birth records) were missing data on duration since a first birth. The resulting sample consists of 1,053,144 second births for which we have valid data on the interval between a first and second birth.

June 1990 Current Population Survey. The 1990 CPS sampling frame was derived from the 1980 decennial census and involves 729 sample areas comprising 1,973 counties and independent cities covering all 50 states and the District of Columbia. The sample is self-weighting and employs no oversampling of racial or ethnic minorities; the sample universe consists of individuals in the U.S. civilian noninstitutional population aged 15 and older. The June 1990 CPS supplement includes items that provide retrospective marital and fertility histories. The sample universe for fertility questions consists of married women aged 15 or older and never-married women aged 18 or older. As a result, the CPS lacks fertility data for never-married women who were younger than 18 at survey. The resulting truncation bias should not seriously affect our analyses because births at very young ages will be available for women who began childbearing prior to age 18 but who were older than 18 at the June 1990 survey date. A screener question obtains the number of children born (excluding stillbirths), followed by items on the date of birth (in calendar month and year) for the first four and last child. This allows one to construct a retrospective birth history through the fifth parity for women with five or fewer births at interview, which we have used to construct the interval between a first a second birth. As in the VSN data, the June 1990 CPS provide large samples of women, but little covariate information.

One caution in using fertility data from the CPS involves the high incidence of missing data.

In the June 1990 CPS, more than 20 percent of birth dates of children to CPS female respondents were "allocated" by census personnel. Published census documentation contains little detail about such allocations, but O'Connell (1998, personal communication) reports that allocations were made under a variety of circumstances, including refusals and when the designated household respondent was unavailable for interview. In some cases, interviewers obtained proxy reports from other household members, while in other cases, dates were imputed by a variety of census procedures. Unfortunately, the allocation flags in public use data do not allow us to distinguish if children's birth dates were supplied by proxy reports from other household members or whether children's birth dates were imputed by procedures such as hot decking by census personnel. We return to this issue below.

1988 National Survey of Family Growth. The National Survey of Family Growth is a retrospective survey of the fertility behavior of U.S. women conducted by the National Center for Health Statistics. The sample consists of 8,450 women aged 15−44 at 1988 interview. The sampling frame was drawn from a stratified sample of households surveyed by the National Health Interview Survey; black and teenage women were oversampled. These data contain a complete retrospective fertility history, including the date of birth (calendar month and year) for all children born to sample respondents. A corresponding retrospective marital history obtains dates (to the nearest month) of marriage, separation, and widowhood/divorce for the respondent's first, second, and most recent marriage.

These data also contain retrospective reports by respondents on age (to the nearest month) at first sexual intercourse. A screener item was used to determine if the respondent had ever engaged in sexual activity. If so, the respondent was asked to supply data on the date (year and month) of first intercourse. Of respondents who have initiated sexual activity, 5% have missing data on calendar year of first intercourse, but 20% have missing data on calendar month of first intercourse. Unfortunately, the public use files mask the raw data on calendar year and month of first intercourse and instead provide an imputed value of century month of first intercourse for respondents with missing calendar month, together with a dummy variable indicating if data have been imputed by NSFG staff. Hence, we possess a form of partially missing data on the timing of an event in a nationally representative sample that allows us to determine (a) which respondents are unable to recall a particular detail of the event in question (calendar month of first intercourse) and (b) the duration between interview and the event of interest.

A disadvantage of the NSFG is that it provides only limited background information for respondents. Available covariates include the respondent's race and ethnicity, family structure at age 14, highest grade completed by the respondent's mother, mother's age at first birth, and religious affiliation at interview. In some analyses, we control for interview characteristics to determine if the prevalence of missing data varies with characteristics of the interview process. Interview characteristics examined include the length of the interview (in minutes), interview mode (person or telephone interview), and month of interview.[3]

1979−93 National Longitudinal Survey of Youth. The NLSY is a prospective survey consisting of a household-based national probability sample of young adults aged 14–21 in 1979. At initial interview in 1979, there were 12,686 respondents, consisting of a main sample of 6,111 individuals, an oversample of 5,295 minorities and poor whites, and a military sample of 1,280 armed forces personnel. The poor white oversample was dropped in 1991. Reinterviews have been conducted annually from 1979 to 1995, and biannually since 1996. These data contain first intercourse, fertility, and marital event histories. Attrition has been low, with 9,011 of the original sample reinterviewed in 1993, fourteen years after initial survey, corresponding to an average annual retention rate of 97.6%. Note that of the 3,675 NLSY respondents not interviewed in 1993, 1,621 respondents were from the poor white oversample suspended in 1991. Adjusting for these

[3]NSFG interviews were conducted over a 10 month period. During this period, interviewers and field operations staff presumably gain experience, which would presumably lead to increases in average data quality during the period of fielding, net of respondent characteristics. On the other hand, it is possible that other changes in the fielding operations over time (interviewer turnover, selection over time on sample respondents who are difficult to contact) could produce decreases in average data quality during the fielding period, net of respondent characteristics.

1,621 respondents increases the average retention rate to 98.5%.

In the initial 1979 interview, respondents were administered retrospective birth and marital histories that obtained dates of all births and marital transitions. In subsequent interviews, respondents were asked to report changes in marital status and periodically re-asked dates of births for all children ever born. When discrepancies arose between respondent reports of dates of birth for children at different waves, we chose the date most frequently reported. Less than 4% of the NLSY birth intervals we analyze were reconciled in this way.

Data on age at first sexual intercourse were obtained in the 1983–1985 interviews, when all respondents were at least 18 years old. In all three waves, a screener item asked the respondent if she had ever engaged in sexual activity; if so, the respondent was asked to supply data on the timing of first intercourse. In 1983, age at first intercourse was obtained to the nearest year. In 1984, these items were repeated for all female respondents, with an item on age at first intercourse followed by items on the corresponding calendar year and month of first intercourse. These questions were repeated in 1985 for female nonrespondents in the 1984 wave. While virtually all NLSY women were able to provide age at first sexual intercourse (to the nearest year), 16% were unable to recall the calendar month of first intercourse. In addition, because the NLSY obtained two respondent reports on age at first sexual intercourse, age at intercourse to the nearest year in 1983, and age at intercourse to the nearest month in 1984 or 1985, a comparison of data from the 1983–85 waves allows us to identify discrepancies in respondent reports.

The NLSY contains a richer array of covariate data than the other data sources we examine. As noted above, covariates common to both the NLSY and NSFG include race and ethnicity, family structure at age 14, mother's education, mother's age at first birth, and respondent's religious affiliation during childhood and adolescence. Covariates available in the NLSY but not in the NSFG include father's SEI, number of siblings, educational expectations, nativity, and a scale for the presence of reading materials in respondent's household at age 14,

One final NLSY variable of particular interest consists of respondent scores on the

Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB was administered to NLSY respondents in 1980, and many analyses of these data use respondents' scores on the Armed Forces Qualifying Test (AFQT), a rough measure of cognitive ability. AFQT is defined operationally by a weighted sum of the four ASVAB subsections on arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations. These sections vary considerably in content. Paragraph comprehension was designed as a test of the respondent's ability to acquire information from written passages. Arithmetic reasoning consisted of "word problems" intended not to test the respondent's ability to translate real-life problems into mathematical terms. Numeric operations assessed the respondent's speed and accuracy on basic arithmetic operations. Word knowledge tested basic vocabulary. As noted above, the wording of event history items in standard survey instruments places demands not only on the respondent's ability to recall an event, but also on the resondent's arithmetic facility to convert one form for the timing of an event (age at event) into another form (the equivalent calendar year and month at event). As a result, we report results for the four subcomponents of AFQT, as well as the usual single composite variable.

As in the NSFG, we control for interview characteristics in some analyses to determine if the prevalence of missing data varies with aspects of the interview process. NLSY interview characteristics examined include the length of the interview (in minutes), items on the respondent's attitude during interview and comprehension of the interview, whether another person was present during interview, and the number of callbacks required to obtain a completed interview. Note that with the exception of interview length, these variables differ from those in the NSFG.

RESULTS

Interval between a first and second birth. We begin by comparing the distributions for the duration between first and second births in the CPS, NLSY, NSFG, and VSN. The VSN data are based on birth registers. Panels A and B of Figure 1 compare the interval between the first and second births obtained from the VSN against those from the NLSY and NSFG, respectively. In both panels, the two curves track one another closely, suggesting that both the omnibus prospective design of the NLSY and the retrospective single-purpose using of the NSFG yield high quality data on the intervals between first and second births.

[Figure 1 about here.]

Panel C in Figure 1 compares intervals obtained from the VSN and CPS. We observe considerable duration heaping in the distribution of birth intervals in the CPS relative to the VSN, with notable spikes occuring at 24, 28, and 36 months in the CPS curve. Figure 2 decomposes the CPS curve in Figure 1 into two components reflecting whether intervals were calculated from allocated or unallocated dates of a first or second birth. The curve corresponding to intervals in which allocations have been made departs significantly from the curve based on intervals obtained from birth registers. For intervals exceeding 9 months in duration, there are noticeable spikes at yearly intervals, coupled with fewer births between these intervals. For intervals shorter than9 months in duration, we observe a larger than expected proportion of birth intervals (given typical gestational durations and postpartum infecundability) in which one or more dates have been allocated.[4] By contrast, the VSN and unallocated CPS curve track one another closely, which suggests that data quality of birth intervals can be high even in retrospective omnibus surveys such as the CPS.

[Figure 2 about here.]

Overall, Figures 1 and 2 suggest little duration heaping across the four data sources, with the exception of CPS cases in which dates have been "allocated." Thus, we find little if any increase in the prevalence of duration heaping in retrospective vs. prospective surveys or in omnibus vs. single-purpose surveys. These findings are consistent with arguments in the survey methodology literature concerning the dating of events that are frequently "rehearsed;" indeed, the classical

[4] As noted above, the public-use CPS files do not permit us to determine if these cases reflect proxy reports for the intended adult CPS respondent or or whether these cases reflect poor properties of hotdeck procedures for missing data used to impute missing dates.

example for rehearsed dates in this literature is dates of births for one's children.

In Figures 1 and 2, we have used the VSN data as a baseline for comparison, implicitly assuming that these data provide an "external" baseline by which to compare sample survey data on birth intervals. However, within the demographic community, data obtained from birth registers are often regarded with considerable suspicion, in part because of worries concerning local and regional variation in the care with which such data are recorded by local authorities. In addition, it is important to note that the birth interval data in all of our data sources may understate certain types of births—for example, births to women who relinquish the infant to adoption. While our analyses do not directly address such births, it is well-documented that survey respondents seriously underreport pregnancies, especially aborted pregnancies, relative to national estimates obtained from abortion providers (Jones and Forrest 1992; Mosher, Pratt, and Duffer 1994). Finally, it is possible that the aggregate nature of the distributions in Figures 1 and 2 could mask duration heaping, telescoping of dates, or other systematic errors in particular population subgroups. Nevertheless, both figures provide little evidence for systematic and substantial biases in respondents' reports of the dates of first and second births, or for systematic variation in duration heaping by type of survey.

Age at first sexual intercourse

Our analyses of the interval between first and second births made use of comparisons against official registers of natality statistics. By contrast, there is no such external standard by which to assess the quality of data on age at first sexual intercourse; hence, our assessments must necessarily proceed more indirectly using comparisons across surveys or by documenting data inconsistencies that appear for respondents in a single survey. Table 1 presents one such assessment by comparing the cumulative percentage of NLSY and NSFG women who experienced first intercourse by age and race and ethnicity.[5] The aggregate distributions, based on Kaplan-Meier (1958) estimates of

[5]To facilitate comparisons, we have restricted the birth cohorts in the NSFG to match those sampled in the

survivor probabilities, show close agrement between the NLSY and NSFG. The largest observed deviations between distributions occur for white women at age 19 and black women at age 15. In neither case are differences statistically significant (results not shown).

[Table 1 about here.]

In the remainder of the paper, we focus on two measures of data quality. The first relies on NLSY data and examines several measures constructed from the difference between the respondent's report of age at first intercourse in the 1983 and 1984/1985 waves. The second uses data from both the NLSY and NSFG to define a dichotomous outcome equal to one if calendar month of first sexual intercourse is missing. For both outcomes, we use standard methods (OLS, simple and multinomial logistic, and poisson regressions) to estimate the association between the outcome and observed covariates.

Discrepancies between 1983 and 1984/85 NLSY respondent reports. As noted above, female respondents in the NLSY provided up to two reports on age at first intercourse, one in the 1983 wave (age to the nearest year), and a second (age to the nearest month) in either the 1984 or 1985 wave. Table 2 presents the frequency distribution for the difference between these two reports for those women with valid responses in both waves. Fully 43.5% of women are observed with discrepancies between reported ages across waves; however, only 11.3% of women have discrepancies that exceed one year in absolute magnitude. Overall, the distribution of differences is roughly symmetric, with relative few women in the tails of the distribution.

[Table 2 about here.]

It is important to note that the figure of 43.5 percent of *observed* discrepancies between waves will overstate the probability that a respondent provides incorrect data in a given wave (O'Muircheartaigh 1991). Intuitively, this occurs because a discrepancy is observed if an error is

NLSY and imputed missing calendar month for NLSY respondents using a simple hot-deck procedure. In both surveys, we censored data on first intercourse either at age at the relevant interview if a woman reported never having experienced sexual intercourse or at her age of first marriage if she initiated sexual activity on or after marriage.

made in either or both waves; hence, the proportion of observed discrepancies will be a sum of the probability of an error from either or both waves. More formally, let y_1 and y_2 denote respondent reports in the two waves of data collection, and let y' denote the unobserved true value of y. From elementary probability theory, we have:

$$\begin{aligned} 1 \; = \; & \Pr(y_1 = y', y_2 = y') + \Pr(y_1 = y', y_2 \neq y') + \Pr(y_1 \neq y', y_2 = y') + \\ & \Pr(y_1 \neq y', y_2 \neq y', y_1 \neq y_2) + \Pr(y_1 \neq y', y_2 \neq y', y_1 = y_2) \end{aligned} \tag{1}$$

The proportion of observed discrepancies across waves corresponds to the second through fourth terms on the right-hand side of (1); hence,

$$\begin{aligned} p \; = \; & \Pr(y_1 = y', y_2 \neq y') + \Pr(y_1 \neq y', y_2 = y') + \Pr(y_1 \neq y', y_2 \neq y', y_1 \neq y_2) \\ = \; & 1 - \Pr(y_1 = y', y_2 = y') - \Pr(y_1 \neq y', y_2 \neq y', y_1 = y_2) \\ = \; & 1 - \Pr(y_1 = y', y_2 = y') - \Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y') \Pr(y_2 \neq y' | y_1 \neq y') \Pr(y_1 \neq y') \end{aligned} \tag{2}$$

Let $e_1 = \Pr(y_1 \neq y')$ and $e_2 = \Pr(y_2 \neq y')$ denote the probability of an error in waves 1 and 2, respectively. If one assumes that e_1 and e_2 are independent and equal in value, with $e_1 = e_2 = e$, then (2) further simplifies as:

$$\begin{aligned} 0 &= p - 1 + (1 - e)^2 + e^2 \Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y') \\ &= p - 2e + e^2 + e^2 \Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y') \\ &= p - 2e + e^2(1 + \lambda) \end{aligned} \tag{3}$$

where $\lambda = \Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y')$. Equation (3) is a quadratic in the probability of error e; hence,

$$e = \frac{2 \pm \sqrt{4 - 4p(1 + \lambda)}}{2(1 + \lambda)} = \frac{1 \pm \sqrt{1 - p(1 + \lambda)}}{(1 + \lambda)} \tag{4}$$

with one of the roots of substantive interest.[6]

To provide a rough sense of how the quantity in (4) differs from the raw proportion of discrepancies reported in Table 2, consider estimating $\Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y')$ from the

[6]Note that p and λ cannot vary independently, with this dependence helping to identify one of the roots.

observed proportion of discrepancies equal to zero in Table 2. (Note that this is equivalent to assuming independence of all quantities in the term $\Pr(y_1 \neq y', y_2 \neq y', y_1 = y_2)$.) Under this admittedly rough assumption, $\hat{e}$ = 27.8 percent, which contrasts with the 43.5 percent of observed discrepancies across the two waves. A similar calculation for probabilities of discrepancies greater than one in absolute magnitude yields $\hat{e}$ = 6.0 percent. Since $\lambda = \Pr(y_1 = y_2 | y_1 \neq y', y_2 \neq y')$ is also simply a probability, it is possible, following Manski (1995), to calculate bounds on e; these calculations yield values of e in the range [24.6%,31.5%], again under the assumption that errors are independent and equal in value across waves.

Overall, Table 2 and our accompanying calculations suggest that inconsistent reports are frequent, with our estimates suggesting somewhat more than 1 in 4 inaccurate reports occuring in a given wave. However, arithmetic errors by either the respondent or interviewer in converting between calendar year and age at occurrence could plausibly generate small discrepancies between waves. Table 2 shows that most discrepancies are small, with our estimates suggesting that only 6 percent of reports in a given wave contain errors of at least 2 years in magnitude.

These observations, coupled with the roughly symmetric distribution of discrepancies in Table 2, suggest two related, but conceptually distinct questions. A first question concerns the factors that might be associated with observed discrepancies between waves; a second concerns the factors associated with the absolute magnitude of observed discrepancies. Both questions are related to reliability, but the first question is motivated primarily by issues of bias—whether systematic bias in data quality appears associated with observed characteristics—while the second question addresses variation in data quality. These two questions also parallel arguments made in the survey methodology literature concerning the heuristic processes that are thought to be used by respondents to reconstruct when an event “must have occurred” for dates of events that are not frequently rehearsed. Note in particular that this hypothesis suggests that longer durations of recall will be positively correlated with the absolute magnitude of observed discrepancies, but this line of reasoning does not provide similar expectations regarding the *sign* of observed discrepancies.

Table 3 presents means and standard deviations for the covariates used in our OLS, poisson, and multinomial regression analyses presented below. Calculating discrepancies across two waves requires data on age at first intercourse in both the 1983 and the 1984/1985 waves. Note, however, that defining the sample in this way allows some women to provide second wave data in which a first intercourse is reported to occur between the first (1983) and second (1984/1985) waves. This possibility is observed only for a very small number of cases (n = 51), but these cases are of concern for two reasons. A first is that for such cases, the two waves of data differ not only in the reported ages at intercourse, but are logically inconsistent, in that first intercourse was reported to have occur prior to the 1983 interview based on data from the first wave, but after the 1983 interview based on data from the second wave. A second cause for concern is statistical, in that defining recall duration based on data from the second wave produces unusually small values of recall duration, with these cases exerting unusual influence (Cook and Weisberg 1982) on parameter estimates for the effects of recall duration.

To deal with this complication, Table 3 reports means and standard deviations for two samples, a full sample (n = 4,113) containing all NLSY female respondents meeting our sample criteria, and a restricted sample ($n = 4,062$) that deletes these 51 influential cases. The distribution of the dependent variable in our OLS analyses is given in Table 2; Table 3 shows that the mean for this variable is close to zero in both the full ($\bar{y} = -.03$) and restricted ($\bar{y} = -.05$) samples. The poisson and multinomial regression results reported below examine the absolute value of this variable; the corresponding means are .62 and .61 in the full and restricted samples, respectively.[7]

[Table 3 about here.]

Our review of the survey methodology literature suggested that the ability of respondents to accurately date events varies inversely with the length of recall, with the ability of respondents to

[7] In our analyses of missing calendar month of first intercourse using NLSY and NSFG data, we control for selected interview characteristics. We do not report such effects in our NLSY analyses of observed discrepancies across waves reported in Tables 4 and 5 below; however, controlling for interviewing characteristics does not substantially change the results reported in Tables 4 and 5 (estimates available upon request).

accurately recall and date events thought to decay exponentially with time. As noted earlier, data on age at first intercourse can be used to determine the duration between the date of interview and the respondent's report of the event. It is important to note, however, that this variable provides at best a rough proxy of recall duration, since an inaccurate report of age at intercourse will generate corresponding measurement error in the corresponding measure of recall duration. In our analyses, we examine two measures for recall duration to provide some sense of the sensitivity of results to this problem. A first measure uses the time between the date of 1984/1985 interview and the respondent's reported age (in months) at first intercourse; this yields a measure of recall duration to the nearest month. A second measure averages recall duration as provided from the 1983 and 1984/1985 waves. Note that the 1983 measure of recall duration yields a figure to the nearest year only; hence this second measure of recall duration averages one variable measured to the nearest year and another measured to the nearest month. Despite these complications, Table 3 shows close agreement in the descriptive statistics for these two measures of recall duration.

Table 4 regresses discrepancies between respondents' reports of age at first intercourse in the 1983 and 1984/1985 waves on observed covariates. To model possible nonlinearities in the effect of recall duration, we use a piecewise linear spline function with a knot specified at 48 months. In Models 1 and 2, we measure recall duration by computing the number of months between date of the 1984/1985 interview and the respondent's 1984/1985 report for age at first intercourse. In subsequent models, we measure recall duration by averaging this measure and one constructed from data obtained in the 1983 wave. We also examine alternative specifications for the effects of ability by contrasting parameter estimates for the single AFQT composite measure (Models 1, 3, and 5) with parameter estimates for scores on the four subtests used to define AFQT (Models 2, 4, and 6). Finally, we present estimates for both a full and restricted sample (Models 1–4 and 5–6, respectively), where the restricted sample drops a small fraction of outliers with unusually small values of recall duration.

[Table 4 about here.]

The results from Models 1 and 2 suggest that only recall duration and educational expectations are significantly associated with discrepancies in respondent reports across waves. Our linear spline specification for the effect of recall duration specifies one linear effect of recall duration at durations between 0 and 48 months, and another linear effect of recall duration at durations greater than 48 months. Estimated coefficients are presented in the row labelled $[0, \infty)$, corresponding to the overall effect of recall duration, and the row labelled $[0, 48)$, corresponding to a contrast for recall durations of less than 48 months. Estimated coefficients are negative and highly significant in Models 1 and 2; however, these estimates correspond to only small changes in predicted discrepancies across waves. For example, each addition year of recall associated with the overall effect yields a $12 \times -.007 = -.084$ year predicted change, or equivalently a $12 \times -.084 = 1.01$ month decrease, in discrepancies observed across waves. For recall durations less than four years, each additional year of recall is associated with a $12 \times -.016 = -.192$ year, or $12 \times .192 = 2.30$ month decrease in discrepancies observed across waves.

Results in Models 3–6 illustrate these issues. Models 3 and 4 substitute the measure of recall duration based on second wave data, as used in Models 1 and 2, by a measure averaging recall duration computed from both waves. Doing so reduces substantially both the magnitude and significance of the effect for durations exceeding four years, from $-.007$ (with a corresponding z-ratio of -10.0 in Models 1 and 2) to $-.000$ (z-ratio of -0.3 in Models 3 and 4). We observe similar, but smaller, change in the magnitude of the effect for durations less than four years, with these effects remaining highly significant in Models 3 and 4. Models 5 and 6 replicate the analyses in Models 3 and 4 on a restricted sample that deletes the 51 influential cases described above. Note that these cases comprise a very small fraction of the full sample ($51/4113$ =1.2 percent) and cases in the full sample with recall durations of less than 48 months ($51/1022$ =5.0 percent). The influence of these cases on the $[0, 48)$ recall duration parameter is readily apparent, with estimates declining sharply in magnitude and significance.

Overall, the results in Table 4 suggest little association between observed covariates and discrepancies across respondent reports of age at first intercourse. The few associations that we do observe appear to be artifacts that are highly sensitive to alternative definitions of key measures and to alternative sample criteria. Thus, we find relatively little empirical structure in discrepancies between respondent reports across waves in age at first intercourse; put another way, the distribution of discrepancies observed in Table 2 appears relatively random with respect to the observed characteristics of respondents.

The analyses in Table 4 assume that observed characteristics of respondents influence the sign, as well as magnitude, of observed discrepancies across waves. However, we argued above that the factors usually held responsible for generating respondent error make predictions about the magnitude, but not the sign, of such discrepancies. The analyses in Table 5 speak to these hypotheses by regressing the absolute value of observed discrepancies on observed covariates. In our poisson regressions, we examine the full distribution of the absolute value of discrepancies; in our multinomial logistic regressions, we pool discrepancies of 4 years or greater, thus limiting our examination to the log odds of no observed discrepancy vs. absolute discrepancies of 1, 2, 3, and 4 or more. As in Table 4, we allow recall duration to have nonlinear effects via a linear spline specification and contrast estimates of the effect of the single composite measure of AFQT with estimates of the four subscores that define AFQT. Throughout, we examine the restricted sample (n = 4,061) and measure recall duration using the average of the 1983 and 1984/1985 values of recall duration.

[Table 5 about here.]

The results generally support expectations suggested in our review of the literature on survey methodology. The first column reports estimates from a poisson regression. In this model, estimates for the two duration parameters are .000 and .021. As in Table 4, these estimates correspond to an overall effect of recall duration and a contrast term for durations lying in the range between 0 and 47 months. Under this model, the association between recall duration and the absolute value of discrepancies increases to a maximum at 48 months, and remains constant at this level for longer recall durations. We observe significantly higher levels of absolute discrepancies for black women relative white women, but no significant differences between white women and women in other racial and ethnic groups. We observe a significant negative effect of AFQT, but, somewhat surprisingly, positive and significant effects of mother's years of schooling completed and respondent's educational expectations.

The next four columns present estimates from our multinomial logistic regressions. Several results parallel those in the poisson regression. As in column 1, the parameter estimate for the overall effect of recall duration is small and not significant across all levels of discrepancies; similarly, the parameter estimate for the contrast term for durations between 0 and 47 months is positive and (with the exception of discrepancies of ± 3) significant. AFQT has a significant negative effect at all levels of discrepancies, with the effect increasing in magnitude at higher discrepancy levels. However, we observe fewer significant differences by race and ethnicity than in the poisson regression of column 1, with significant black/white differences confined to discrepancies of ± 1. Note also that the positive and significant associations for mother's education and educational expectations observed in the poisson regressions remain positive, but decline in significance in the multinomial logistic regressions.

The remaining columns of Table 5 modify these analyses by decomposing the effect of AFQT into four subcomponents. Estimated coefficients for recall duration and race and ethnicity generally parallel those in previous models; note, however, that black/white differences decline slightly in magnitude and are no longer statistically significant. Higher scores on the vocabulary and arithmetic subsections are associated with smaller absolute discrepancies in both the poisson and multinomial regressions. These effects are significant in the poisson regression but only in two out of eight coefficients in the multinomial logistic regression. Effects of respondent scores on the numerical operations and paragraph comprehension subtests are not significant in either the poisson or multinomial logistic regressions. Effects of the background factors are generally similar to those of previous models.

Overall, the results in Tables 4 and 5 are largely consistent with recent arguments in the survey methodology literature. Our analyses in Table 4 suggested that the distribution of discrepancies observed in Table 2 is relatively random with respect to observed covariates, while our analyses in Table 5 suggested that the absolute magnitude of these discrepancies is significantly associated with observed characteristics of respondents such as the duration of recall and measures of ability. We interpret the first finding as suggesting that the difference in respondents reports across waves is, to a first approximation, random with respect to the *sign* of observed discrepancies; we interpret the second finding as evidence of nonrandom variation with respect to the *magnitude* of observed differences. In additional analyses in Table 5, we decomposed the effects of respondent ability into four subcomponents corresponding to respondent scores on word knowledge, arithmetic reasoning, numerical operations, and paragraph comprehension. The first two tests measure the respondent's basic vocabulary and the respondent's ability to translate real-life problems into mathematical terms—skills that we argued are similar to those involved in dating events retrospectively—while the latter two tests assess the respondent's ability to acquire information from written text passages, and the respondent's speed and accuracy on basic arithmetic operations—skills that we argued are less immediately related to the task of dating events retrospectively. Our findings were consistent with these expectations, with significant effects of vocabulary and arithmetic reasoning, but no significant effects of numerical operations and paragraph comprehension.

Missing calendar month of first intercourse. We now turn to analyses of a form of partially missing data—respondent inability to recall the calendar month of intercourse. This form of partially missing data is present in both the NLSY and NSFG. Table 6 presents means and standard deviations for variables used in our logistic regressions. Because our analyses contrast women who can and cannot recall calendar month of first intercourse, our NLSY analyses are restricted to data from the 1984 wave, or, for nonrespondents in 1984, to data from the 1985 wave. As a result, our NLSY sample selection criteria differs from that used in Tables 3–5; these differences are reflected in the sample sizes reported in Table 6. Note that NSFG respondents were aged 15−44 at 1988 interview; hence, NSFG women will have longer durations of recall than NLSY women. Table 6 shows that recall averaged approximately 69 months for NLSY women and approximately 157 months for NSFG women. These longer durations of recall in the NSFG are accompanied by a slightly higher prevalence of missing calendar month of first intercourse (20% vs. 16% in the NSFG and NLSY, respectively).

[Table 6 about here.]

Do differences in average recall durations across surveys produce different patterns in the association between recall and respondent inability to recall the calendar month of intercourse? The two panels of Figure 3 examine this issue empirically. We grouped data into duration intervals that contain at least 10 respondents with a missing calendar month of first intercourse. Within each interval, we computed the empirical log odds ratio, plotted this ratio at the midpoint of each interval, and passed a smooth curve through the resulting scatterplot of empirical log odds ratios. The resulting curves exhibit a monotonic pattern with duration, with the log odds of missing month increasing more rapidly at short durations of recall than at longer durations of recall. Note that we observe NSFG log odds ratios over a wider range of durations than for NLSY log odds because of the wider age range of NSFG respondents. Both curves have roughly similar shapes; however, the slope of the NLSY curve appears to be slightly steeper at intermediate durations than the NSFG curve.

[Figure 3 about here.]

Overall, the curves in Figure 3 suggest that the log odds of missing month does not vary linearly with duration, suggesting instead a somewhat more complicated functional form than a simple exponential decay with time. To model these nonlinearities, we use a piecewise linear spline function with knots at 48 months and (in the NSFG) at 120 months duration; note that this specification parallels that used in Tables 4 and 5 for discrepant data in the NLSY.

Table 7 presents results from our logistic regressions. The first column (Model 1) presents of estimates for NSFG women and specifies effects of recall duration, race and ethnicity, and background characteristics that are common to the NLSY and NSFG. As in previous tables, the row labelled $[0, \infty)$ estimates the overall effect of recall duration, while the row labelled $[0, 48)$ presents the estimated contrast for recall durations of less than 48 months. A third row, labelled $[120, \infty)$, estimates a final contrast for recall durations of 120 months or greater; we specify this effect only for NSFG women to account for differences in the tail of the recall duration distribution between the NLSY and NSFG. The coefficients for recall duration in Model 1 suggest an overall nonlinear pattern for the effect of recall duration. During the first 48 months of recall, the odds of missing month for NSFG women increase by exp(.007 + .019) = 1.026 or 2.6 percent per month. Between 48 and 120 months, the odds of missing month increase at a slower rate of exp(.007) = 1.007, or 0.7 percent per month; after 120 months, the odds increase at an even slower rate of $\exp(.007 - .005) = 1.002$, or 0.2 percent per month. Note that this overall pattern—a positive association at short recall durations, followed by positive, but smaller, associations at longer recall durations—resembles the pattern of nonlinear effects for recall duration observed in Table 5.

[Table 7 about here.]

Estimates in Model 1 suggest higher log odds for missing month for all three nonwhite racial and ethnic groups; however, only the effect for black women is significant, with this estimate suggesting a 42% (1.42 = exp[.35]) higher odds of a missing month of first intercourse for black women relative to white women. The remaining coefficients for the background controls are not significant, except for the coefficient for mother's education; note, however, that this coefficient is in the opposite direction of typical expectations.

Model 2 adds controls for interview characteristics. Longer interviews are associated with lower odds of missing month, while telephone interview and calendar month of interview are associated with increased odds of missing month; however, the coefficient for telephone interview is not significant. The positive effect of calendar month of interview may reflect the changing composition of the interview pool if the proportion of sample households that are difficult to contact increases over the fielding period. Estimated coefficients for recall duration, race and ethnicity, and background characteristics are relatively stable between Models 1 and 2.

The remaining columns of Table 7 present estimates for NLSY women. Overall, results closely resemble those for NSFG women. Estimates in Model 1 for recall duration are close in magnitude to those for NSFG women, although the estimated contrast for recall durations of less than 48 months is somewhat smaller for NLSY women than for NSFG women (.019 vs. .024). We also observe a slightly smaller coefficient for black women in the NLSY than in the NSFG. Neither of these differences in coefficients is statistically significant. Note however that we observe a negative but not statistically significant coefficient for mother's education, which differs from the positive and significant coefficient for mother's education for NSFG women.

Estimates in Model 2, which add controls for NLSY interview characteristics, result in few changes in the estimated coefficients for recall duration, race and ethnicity, and background characteristics. Longer interviews decrease the odds of missing month, while telephone interview and a scale indicating poor respondent attitude are associated with increased odds of missing month; however, only the last coefficient is significant, with the other two coefficients approaching significance. The remaining interview variables are not significantly associated with the odds of missing month.

Models 3 and 4 add effects of the ability and the remaining background controls available in the NLSY but not in the NSFG. We again observe few changes in estimated coefficients for recall duration, race and ethnicity, and the background characteristics common to both surveys. The additional background variables (father's SEI, number of siblings, educational expectations, foreign born, and the scale for reading materials) are negatively associated with the log odds for missing month; however, only the coefficient for foreign born is statistically significant. In

Model 3, women with higher ability scores, as measured by AFQT, are observed to have lower log odds of missing month; however, Model 4 suggests that this effect loads primarily on on the vocabulary subtest, with no significant association between the log odds of missing month and the other three tests.

DISCUSSION

In this paper, we examined the quality of demographic data supplied by female respondents on two event history outcomes in light of hypotheses on respondent recall drawn from the survey methodology literature. A first demographic variable concerned the interval between a first and second birth. We compared birth interval data using birth registration data from the Vital Statistics on Natality (VSN), and more typical individual-level data from the June Current Population Survey (CPS), the 1979−94 waves of the National Longitudinal Survey of Youth (NLSY), and the 1988 National Survey of Family Growth (NSFG). Although these surveys vary markedly in design, we found little evidence of duration heaping for second birth intervals in the surveys examined. The single exception to this general finding concerned data from CPS respondents in which dates of births had been flagged as allocated by census personnel.

Our remaining analyses focused on a second demographic variable—age at first sexual intercourse. A first set of analyses examined discrepancies in age at first intercourse as reported in successive NLSY waves. Large numbers of NLSY women supply inconsistent reports on age at first intercourse across waves; however, the vast majority of such observed inconsistencies are small in magnitude. Based on these data, we estimate that roughly 1 in 4 NLSY women provide an inaccurate report for age at first intercourse in a given wave, but that only 6% of the reported ages at first intercourse in a given NLSY wave contain errors of at least 2 years in magnitude. Results from OLS regressions suggested that the distribution of observed discrepancies was random with respect to observed characteristics of NLSY women; results from poisson and multinomial logistic regressions suggested that the *absolute value* of such discrepancies is significantly associated with observed characteristics of respondents. Thus, our results suggested little empirical structure in the sign of observed discrepancies in successive reports of age at first intercourse, but considerable empirical structure in the absolute magnitude of such discrepancies. In particular, the duration of recall and measures of respondent ability were significantly associated with the absolute magnitude of discrepancies in reports across waves.

A second set of analyses examined a form of partially missing data—respondents' inability to recall the calendar month of first sexual intercourse—that occurs in both the NLSY and NSFG. Although most NLSY and NSFG respondents were able to provide data on the calendar year of first intercourse, substantial percentages—16 percent in the NLSY and 20 percent in the NSFG—were unable to recall the calendar month of first intercourse. As in our other analyses, we found significant associations between the duration of recall, race and ethnicity, measures of respondent ability, and some interview characteristics on the log odds of missing reports of the calendar month of first intercourse. Estimated effects in logistic regressions were generally close in value across the two surveys; however, black/white differences observed in both the NLSY and NSFG declined in both size and significance when controlling for additional covariates only available in the NLSY.

Our analyses of the effects of recall duration on the absolute magnitude of discrepant reports across waves of the NLSY and on the log odds of missing data on the calendar month of first intercourse are consistent with the hypothesis that data quality declines with the duration of recall. However, we find evidence of nonlinearities in the effect of recall duration, with data quality deteriorating more slowly at longer durations of recall than at shorter durations of recall. We observe this pattern in both our analyses of the absolute magnitude of discrepant reports in the NLSY and of the log odds of missing data on calendar month in the NLSY and NSFG. In the latter analyses, our results suggest strikingly similar parameter estimates for these nonlinearities in both the NLSY and NSFG.

These analyses also provide hints about other individual-level factors that may affect the accuracy of self-reported event history data in standardized survey settings. Our NLSY analyses suggest that discrepancies are smaller in absolute magnitude for respondents who score higher on the Armed Forces Qualifying Test (AFQT), a measure of general ability. Because the NLSY survey instrument on age at first intercourse required respondents to convert a report on age at first intercourse to the equivalent calendar year and month at first intercourse, arithmetic errors by either the respondent or interviewer could plausibly generate the observed association between ability and data quality. To examine this possibility, we examined the effects of respondent scores on four subsections used to define AFQT. Our results suggested that data quality varies significantly with respondent scores on tests of basic vocabulary and the respondent's ability to translate real-life problems into mathematical terms. These effects are plausible to the extent that tests of basic vocabulary are correlated with more generalized abilities related to recall and memory, and to the extent that aptitude in translating word problems into mathematical terms correlates with the skills required to convert age at the onset of an event into the corresponding calendar year at onset. Conversely, we found no significant effects of respondent scores on tests of paragraph comprehension or scores on tests that measure the respondent's speed and accuracy on basic arithmetic operations—skills that we argue are less immediately related to the task of dating events retrospectively.

Taken as a whole, our findings are consistent with arguments that respondent recall of the timing of an event declines with duration of recall unless the dating of the event is frequently "rehearsed." Taken at face value, these results suggest relatively few means by which to increase the quality of event history data, given the constraints encountered in most social surveys. However, our findings also suggest other sources of variation in respondent recall that may help reduce some sources of error in respondent reports. Thus, computer-aided interviewing techniques may help reduce respondent and interviewer burden in situations that require simple arithmetic operations to convert the timing of an event from one metric to another. Similarly, it is possible that extended question sequences or the use of memory prompts such as event calendars tied to "landmark" events may help respondents reconstruct data on the timing of other events with greater accuracy. Still, the presumed utility of memory prompts such as event calendars rests heavily on the accurate dating of “landmark” events. It is important in this respect to emphasize that the accuracy of data for the demographic events examined in this paper—births and the initiation of first sexual activity—are both plausibly construed as “landmark” events, but that the quality of respondent self-reported data varies substantially for these two events.

REFERENCES

Abma, Joyce C., Anjani Chandra, William D. Mosher, Linda S. Peterson, and Linda J. Piccinino. 1997. “Fertility, Family Planning, and Women’s Health: New Data from the 1995 National Survey of Family Growth.” *Vital and Health Statistics* Series 23, No. 19. U.S. Department of Health and Human Services, Publication No. PHS-97-1995.

Blair, Edward and Scot Burton. 1987. “Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions.” *Journal of Consumer Research* 14(2): 280-288.

Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. “Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys.” *Science* 236(4798): 157-61.

Brewer, William F. 1994. “Autobiographical Memory and Survey Research.” Pp. 11-20 in Norbert Schwarz and Seymour Sudman (Eds.,) *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer Verlag.

Dykema, Jennifer. 1996. “Events, Instruments, and Reporting Errors: Combining Knowledge From Multiple Perspectives.” Unpublished Master’s Thesis, Department of Sociology, University of Wisconsin-Madison.

Jones, Elise F., and Jacqueline Darroch Forrest. 1992. “Underreporting of Abortion in Surveys of Women: 1976 to 1988.” *Demography* 29(1): 113–26.

Kaplan, E. L., and Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53(282): 437–81.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences.* Cambridge, MA: Harvard University Press.

Martin, Steven P., and Lawrence L. Wu. 1998. “Teenage and Subsequent Childbearing.” CDE Working Paper No. 98-01, Center for Demography and Ecology, University of Wisconsin-Madison.

Mathiowetz, Nancy A., and Greg J. Duncan. 1988. “Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment.” *Journal of Business and Economic Statistics* 6(2): 221-29.

Means, Barbara, Swan Gary E., Jared B. Jobe, and James L. Esposito. 1991. “An Alternative Approach to Obtaining Personal History Data.” Pp. 167-84 in Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman (Eds.), *Measurement Errors in Surveys.* New York: Wiley.

Menon, Geeta. 1993. “The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies.” *Journal of Consumer Research* 20: 431-40.

Menon, Geeta. 1994. “Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies.” Pp. 161-72 in Norbert Schwarz and Seymour Sudman (Eds.,) *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer Verlag.

Mosher, William D., William F. Pratt, and Allen P. Duffer. 1994. “CAPI, Event Histories, and Incentives in the NSFG Cycle 5 Pretest.” Pp. 59-63 in *Proceedings of American Statistical Association, Section on Survey Research Methods*, Toronto, Canada.

Neisser, Ulric and N. Harsch. 1992. “Phantom Flashbulbs: False Recollections of Hearing the News About Challenger.” Pp. 9-31 in E. Winograd and Ulric Neisser (Eds.), *Affect and Accuracy in Recall: Studies of “Flashbulb” Memories.* New York: Cambridge University Press.

Neter, J., and J. Waksberg. 1964. "A Study of Response Errors in Expenditures Data from Household Interviews." *Journal of the American Statistical Association* 59: 18-55.

O'Muircheartaigh, Colm A. 1991. "Simple Response Variance: Estimation and Determinants." Pp. 551–74 in Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman ((Eds.), *Measurement Error in Surveys.* New York: Academic Press.

Schaeffer, Nora Cate. 1994. "Errors of Experience: Response Errors in Reports About Child Support and Their Implications for Questionnaire Design." Pp. 141-60 in Norbert Schwarz and Seymour Sudman (Eds.,) *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer Verlag.

Schaeffer, Nora Cate, Elizabeth Thomson, Jennifer Dykema, and Lawrence L. Wu. 1998. "Technical Review of the National Survey of Family Growth Cycle 6 Questionnaire." Report submitted to the National Center for Health Statistics, October 1998.

Schwarz, Norbert, and Seymour Sudman, (Eds.). 1994. *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer Verlag.

Smith, Albert F. and Jared B. Jobe. 1994. "Validity of Reports of Long-Term Dietary Memories: Data and a Model." Pp. 121-40 in Norbert Schwarz and Seymour Sudman (Eds.,) *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer Verlag.

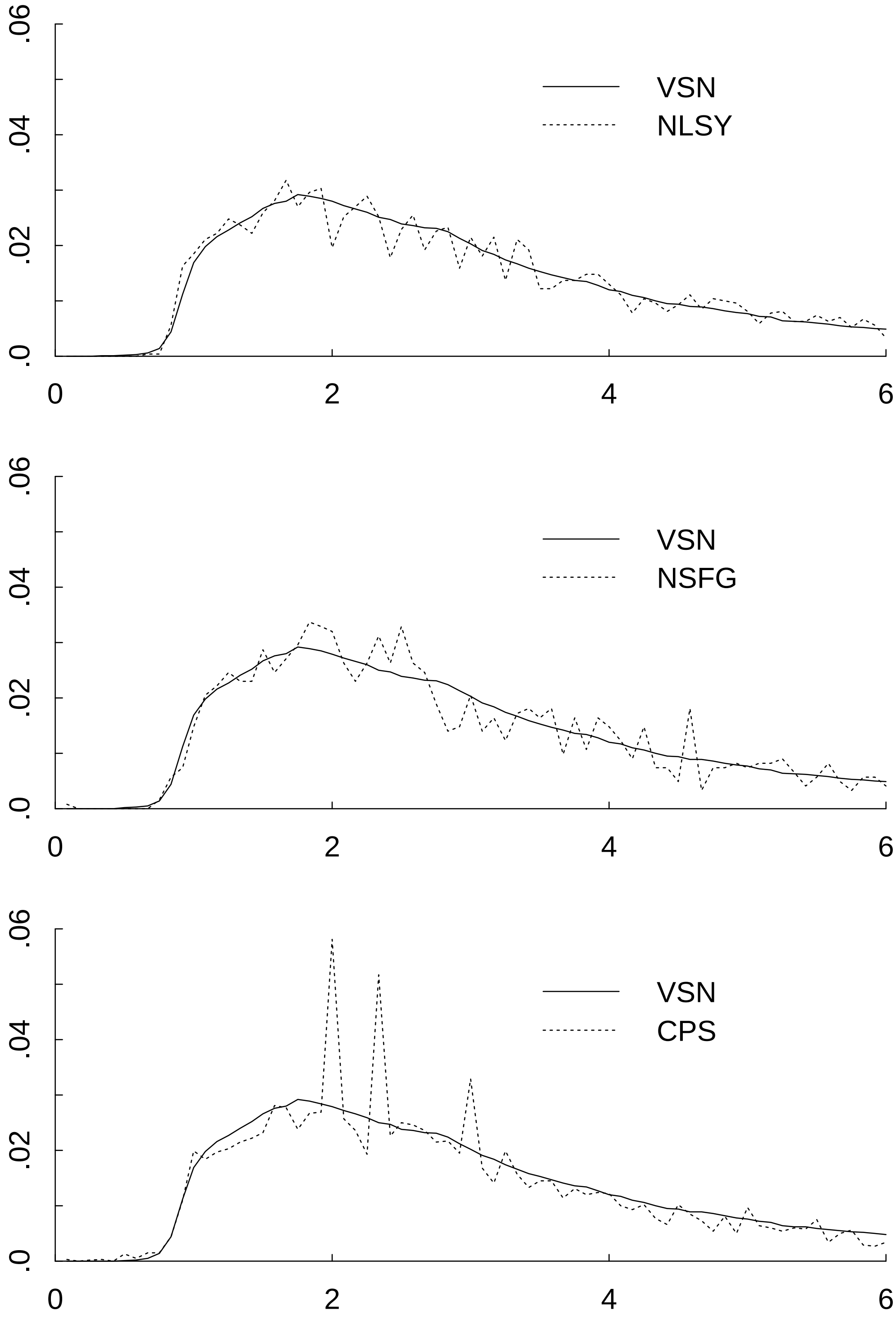
Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.

Sweet, James, Larry L. Bumpass, and Vaughn Call. 1988. "The Design and Content of the National Survey of Families and Households." NSFH Working Paper No. 1, Center for Demography and Ecology, University of Wisconsin-Madison.

Thompson, Charles P., John J. Skowronski, Steen F. Larsen, and Andrew L. Betz. 1996. *Autobiographical Memory: Remembering What and Remembering When.* Mahwah, N.J.: Lawrence Erlbaum.

Wu, Lawrence L. "Effects of Family Instability, Income, and Income Instability on the Risk of a Premarital Birth." *American Sociological Review* 61(3): 386-406.

Wu, Lawrence L., and Brian C. Martinson. 1993. "Family Structure and the Risk of a Premarital Birth." *American Sociological Review* 58(2): 210-32.


.06
.04
.02
.0
VSN
NLSY
0
2
4
6
.06
.04
.02
.0
VSN
NSFG
0
2
4
6
.06
.04
.02
.0
VSN
CPS
0
2
4
6

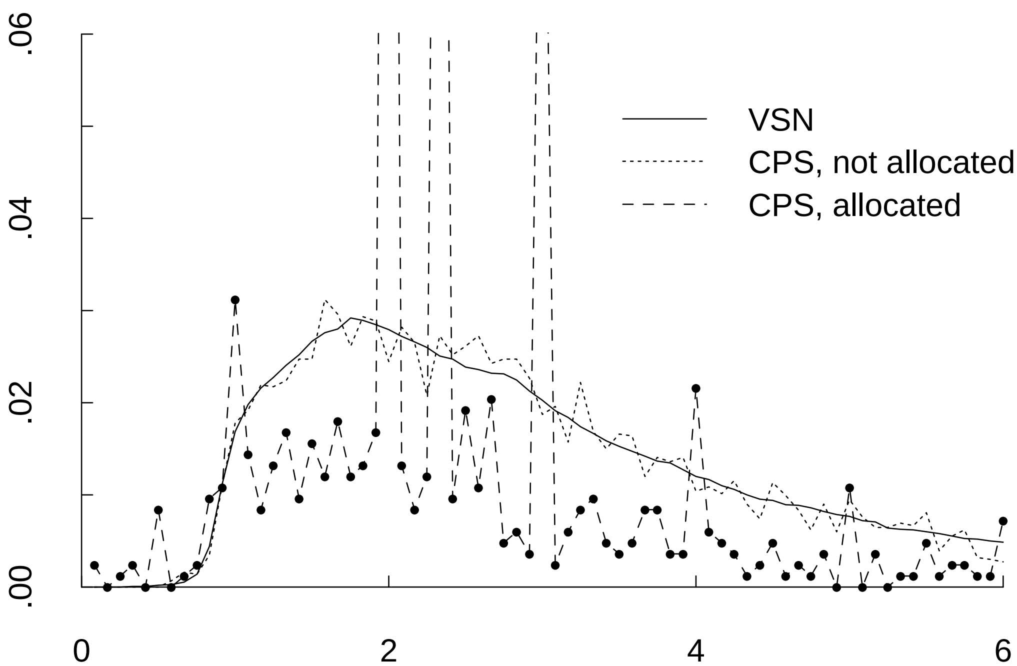
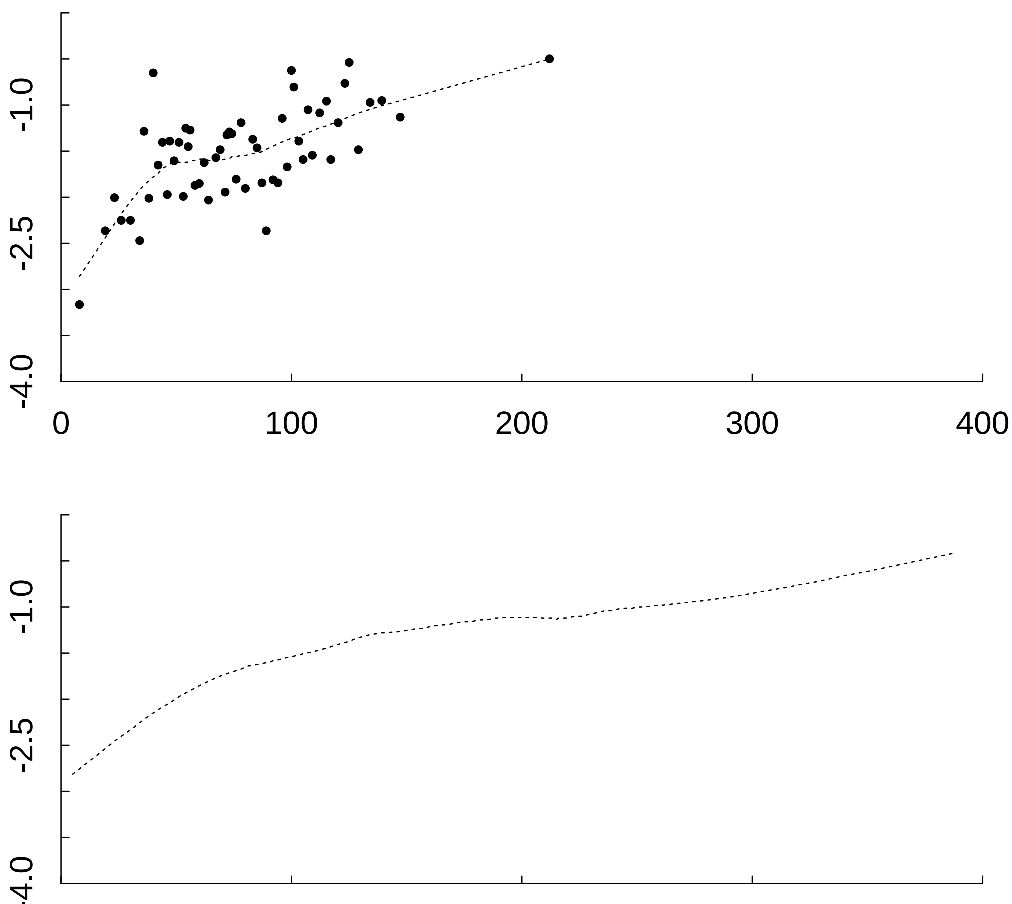


Figure 2. Duration heaping in the interval between a first and second birth with allocated and unallocated dates of CPS births. Females, 1990 Vital Statistics on Natality and June 1990 Current Population Survey.



n = 7,259) and National Survey of Family Growth (n = 4,774)

Table 1. Kaplan-Meier estimates of the cumulative percent experiencing first sexual intercourse by age. White and black women born 1958−65.

Age	13	14	15	16	17	18	19	20	21	22	n
White women, NLSY	1	3	8	17	35	52	69	77	82	86	2561
White women, NSFG	2	4	8	18	33	49	63	72	79	83	1353
Black women, NLSY	2	6	13	30	52	68	83	88	92	94	1271
Black women, NSFG	3	8	18	31	50	68	81	87	92	94	784

Sources: National Longitudinal Survey of Youth, 1979−93, and National Survey of Family Growth 1988.

Table 2. Frequency distribution for the difference between reported age at first sexual intercourse from 1983 and 1984/85 respondent reports. Females, National Longitudinal Survey of Youth, 1979-85 (n = 4,113).

Difference	Frequency	Percent	Cumulative Percent
−8	0	.00	.00
−7	6	.15	.15
−6	3	.07	.22
−5	11	.27	.49
−4	11	.27	.75
−3	55	1.34	2.09
−2	159	3.87	5.96
−1	691	16.80	22.76
0	2323	56.48	79.24
1	632	15.37	94.60
2	144	3.50	98.10
3	41	1.00	99.10
4	26	.63	99.73
5	7	.17	99.90
6	2	.05	99.95
8	2	.05	100.00
Total	4113	100.00	100.00

Table 3. Variable means and standard deviations, full (n = 4,113) and restricted (n = 4,062) samples. Females, National Longitudinal Survey of Youth, 1979–85.

	full sample		restricted sample	
	mean	s.d.	mean	s.d.
Dependent variables				
discrepancy between 83 and 84/85 waves	−.03	1.11	−.05	1.11
absolute value of discrepancy	.62	.92	.61	.91
Duration of recall				
months between first sex and 84/85 waves	72.30	32.41	73.10	31.81
average of recall from 83 and 84/85 waves	72.12	31.51	72.82	31.06
Race and ethnicity				
black	.24		.24	
hispanic	.12		.12	
other	.14		.14	
missing	.01		.01	
Ability				
AFQT	.01	.93	.01	.93
vocabulary	.01	.94	.01	.95
arithmetic	−.11	.92	−.11	.92
numerical operations	.11	.96	.10	.96
paragraph comprehension	.09	.94	.09	.94
missing AFQT	.03		.03	
Background characteristics				
mother's education	10.79	3.11	10.79	3.11
missing mother's education	.05		.05	
father's SEI, age 14	28.35	14.10	28.34	14.11
father did not work at age 14	.46		.46	
father not present at age 14	.02		.02	
missing father's SEI	.03		.03	
mother's age at first birth	20.77	5.20	20.75	5.19
missing mother's age at first birth	.19		.19	
R does not know biol. mother	.00		.00	

Table 3. (continued)

	full sample		restricted sample	
	mean	s.d.	mean	s.d.
Background characteristics				
number of siblings	3.93	2.66	3.93	2.67
intact family at age 14	.66		.66	
missing intact family at age 14	.14		.14	
educational expectations	13.65	2.40	13.64	2.41
missing educational expectations	.01		.01	
reading materials	1.74	1.17	1.74	1.17
missing reading materials	.01		.01	
catholic	.33		.33	
foreign born	.06		.06	

Table 4. Least squares estimates of the effects of recall duration, race and ethnicity, ability and background characteristics on discrepancies in reports of age at first sexual intercourse. full and restricted samples of females, National Longitudinal Survey of Youth, 1979-85.

	full sample				restricted sample	
	1	2	3	4	5	6
Recall duration						
$[0, \infty)$						
84/85	−.007*** (.001)	−.007*** (.001)				
83 and 84/85			−.000 (.001)	−.000 (.001)	−.000 (.001)	−.000 (.001)
$[0, 48)$						
84/85	−.009*** (.003)	−.009*** (.003)				
83 and 84/85			−.006* (.003)	−.006* (.003)	−.001 (.003)	−.001 (.003)
Race and ethnicity						
black	.02 (.05)	.02 (.05)	.03 (.05)	.03 (.05)	.03 (.05)	.03 (.05)
hispanic	.02 (.06)	.03 (.06)	.09 (.07)	.09 (.07)	.10 (.07)	.10 (.07)
other	.01 (.07)	.00 (.07)	−.01 (.08)	−.02 (.08)	−.02 (.08)	−.02 (.08)
Ability						
AFQT	−.02 (.02)		.02 (.03)		.02 (.03)	
vocabulary		.02 (.03)		.02 (.04)		.03 (.04)
arithmetic		−.02 (.03)		−.02 (.03)		−.02 (.03)
numerical operations		−.02 (.02)		.01 (.03)		.01 (.02)
para. comprehension		.01 (.03)		.00 (.03)		.01 (.03)

Table 4. (continued)

	full sample				restricted sample	
	1	2	3	4	5	6
Background characteristics						
mother's education	.00	.00	.00	.00	.00	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
father's SEI $\times$10	.00	.00	.01	.01	.01	.01
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
mother's age at first birth	−.00	−.00	−.00	−.00	−.00	−.00
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
number of siblings	.01†	.01†	.01	.01	.01	.01
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
intact family at age 14	.05	.05	.06	.06	.05	.05
	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
educ. expectations	−.02**	−.02**	−.02†	−.02†	−.02†	−.02†
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
catholic	−.05	−.05	−.04	−.04	−.04	−.05
	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
foreign born	.01	.02	.04	.04	.04	.04
	(.07)	(.07)	(.08)	(.08)	(.08)	(.08)
reading materials	.02	.02	.01	.01	.01	.01
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Constant	1.10***	1.11***	.41*	.41*	.21	.22
	(.18)	(.18)	(.19)	(.19)	(.19)	(.19)
sample size	4113	4113	4113	4113	4062	4062
R^2	.07	.07	.01	.01	.01	.01

Note: Standard errors reported in parentheses. All models control for dummy variables for missing values of: race and ethnicity, ability, family structure at age 14, father's SEI, educational expectations, and reading materials.

† $p < .10$ * $p < .05$ ** $p < .005$ *** $p < .0005$ (two-tailed test)

Table 5. Poisson and multinomial logistic estimates of the effects of recall duration, race and ethnicity, ability, and background characteristics on the absolute value of discrepancies in reports of age at first sexual intercourse. Restricted sample of females, National Longitudinal Survey of Youth, 1979-85 (n = 4,062).

	Poisson	Multinomial logistic				Poisson	Multinomial logistic			
	all	1 vs. 0	2 vs. 0	3 vs. 0	4+ vs. 0	all	1 vs. 0	2 vs. 0	3 vs. 0	4+ vs. 0
Recall duration										
$[0, \infty)$	.000	−.001	−.005†	.006	−.001	.000	−.001	−.005†	.006	−.001
	(.001)	(.002)	(.003)	(.004)	(.005)	(.001)	(.002)	(.003)	(.004)	(.005)
$[0, 48)$	.021***	.013*	.049***	.050†	.278*	.020***	.013*	.048***	.049†	.281*
	(.004)	(.006)	(.013)	(.028)	(.126)	(.004)	(.006)	(.013)	(.028)	(.127)
Race and ethnicity										
black	.12*	.22*	.27	.30	.08	.10†	.20†	.25	.25	.01
	(.06)	(.11)	(.18)	(.31)	(.37)	(.06)	(.11)	(.19)	(.32)	(.37)
hispanic	.03	−.06	.10	.13	.28	.02	−.07	.08	.08	.26
	(.08)	(.14)	(.25)	(.44)	(.50)	(.08)	(.14)	(.25)	(.44)	(.50)
other	−.05	.17	.13	−.24	−1.05	−.05	.18	.13	−.25	−1.07†
	(.09)	(.15)	(.29)	(.50)	(.65)	(.09)	(.15)	(.29)	(.50)	(.65)
Ability										
AFQT	−.15***	−.14*	−.19*	−.35*	−.66***					
	(.03)	(.05)	(.09)	(.15)	(.18)					
vocabulary						−.09*	−.16*	−.15	−.10	−.27
						(.04)	(.07)	(.13)	(.22)	(.25)
arithmetic						−.09*	−.02	−.09	−.28	−.63*
						(.03)	(.06)	(.11)	(.19)	(.25)
numerical operations						.02	.03	.08	.19	−.03
						(.03)	(.05)	(.09)	(.15)	(.18)
paragraph comprehension						−.02	.00	−.05	−.21	.08
						(.04)	(.07)	(.12)	(.20)	(.23)

Table 5. (continued)

	Poisson	Multinomial logistic				Poisson	Multinomial logistic			
	all	1 vs. 0	2 vs. 0	3 vs. 0	4+ vs. 0	all	1 vs. 0	2 vs. 0	3 vs. 0	4+ vs. 0
Background characteristics										
mother's education	.02*	.03†	.02	.07	.06	.02*	.03†	.02	.07	.06
	(.01)	(.01)	(.03)	(.05)	(.05)	(.01)	(.01)	(.03)	(.05)	(.05)
father's SEI ×10	−.03†	−.00	−.01	−.14	−.20	−.03†	−.00	−.01	−.13	−.20
	(.02)	(.03)	(.05)	(.10)	(.14)	(.02)	(.03)	(.05)	(.10)	(.14)
mother's age at first birth	.00	−.01	.03*	−.00	−.01	.00	−.01	.03*	−.00	−.01
	(.00)	(.01)	(.01)	(.02)	(.03)	(.00)	(.01)	(.01)	(.02)	(.03)
number of siblings	−.01	−.00	.01	−.04	−.05	−.01	−.00	.01	−.04	−.05
	(.01)	(.01)	(.03)	(.04)	(.05)	(.01)	(.01)	(.03)	(.05)	(.05)
intact family at age 14	−.02	.11	.02	−.31	−.11	−.03	.11	.02	−.32	−.12
	(.05)	(.08)	(.15)	(.25)	(.30)	(.05)	(.08)	(.15)	(.25)	(.30)
educ. expectations	.04***	.03	.06†	.08	.11†	.04***	.03	.06*	.09	.11†
	(.01)	(.02)	(.03)	(.05)	(.06)	(.01)	(.02)	(.03)	(.05)	(.06)
catholic	−.07	.05	−.20	−.08	−.50	−.07	.06	−.20	−.09	−.54
	(.05)	(.09)	(.16)	(.29)	(.38)	(.05)	(.09)	(.16)	(.29)	(.39)
foreign born	.14†	−.08	.18	.11	.61	.15†	−.08	.17	.10	.66
	(.09)	(.16)	(.27)	(.47)	(.48)	(.09)	(.16)	(.27)	(.47)	(.48)
reading materials	.00	.01	.11	−.07	−.13	.00	.01	.11	−.08	−.14
	(.03)	(.04)	(.08)	(.13)	(.15)	(.03)	(.05)	(.08)	(.13)	(.15)
Constant	−2.09***	−1.70***	−5.91***	−7.14***	−17.45**	−2.12***	−1.75***	−5.97***	−7.24***	−17.72**
	(.24)	(.39)	(.77)	(1.55)	(6.05)	(.24)	(.39)	(.77)	(1.56)	(6.08)

Note: Recall duration variable defined as the average of recall duration from the 1983 and 1984/1985 waves. Standard errors reported in parentheses. All models control for dummy variables for missing values of: race and ethnicity, ability, family structure at age 14, father's SEI, educational expectations, and reading materials.

† $p < .10$ * $p < .05$ ** $p < .005$ *** $p < .0005$ (two-tailed test)

Table 6. Variable means and standard deviations. Females, National Longitudinal Survey of Youth, 1979-85 (n = 4539), and National Survey of Family Growth, 1988 (n = 7259).

	NLSY		NSFG	
	mean	s.d.	mean	s.d.
Missing calendar month	.16	.37	.20	.40
Duration of recall				
months between first sex and interview	68.85	34.55	156.63	86.17
Race and ethnicity				
black	.24	.42	.31	.46
hispanic	.13	.34	.07	.26
other	.14	.34	.02	.15
missing	.01	.09	.02	.06
Ability				
AFQT	.03	.93		
vocabulary	.03	.94		
arithmetic	−.09	.93		
numerical operations	.12	.96		
para. comprehension	.11	.94		
missing AFQT	.04	.19		
Interview characteristics				
interview length (in minutes)	75.84	24.16	79.38	26.95
attitude during interview	1.18	.44		
understanding during interview	1.10	.33		
other person present?	.15	.36		
number of callbacks	5.57	4.79		
telephone interview			.00	.05
century month at interview	1010.84	2.39	1059.44	1.55

Table 6. (continued)

	NLSY		NSFG	
	mean	s.d.	mean	s.d.
Background characteristics				
mother's education	10.82	3.13	11.33	3.00
missing mother's education	.05	.22	.10	.30
father's SEI, age 14	28.43	14.14		
father did not work at age 14	.46	.50		
father not present at age 14	.02	.13		
missing father's SEI	.03	.16		
mother's age at first birth	20.86	5.20	20.97	4.22
missing mother's age at first birth	.18	.39	.09	.29
R does not know biol. mother	.00	.04	.00	.00
number of siblings	3.89	2.65		
missing number of siblings	0.0	0.0		
intact family at age 14	.67	.44	.69	.46
missing intact family at age 14	.13	.34	.01	.30
educational expectations	13.70	2.42		
missing educational expectations	.00	.09		
catholic	.34	.47	.24	.43
foreign born	.06	.24		
reading materials	1.75	1.17		
missing reading materials	.01	.09		

Table 7. Effects of recall duration, race and ethnicity, interview characteristics, and background characteristics on the log odds of missing calendar month for date of first premarital sexual intercourse. Females, National Longitudinal Survey of Youth, 1979-85 (n = 4539), and National Survey of Family Growth, 1988 (n = 7259).

	NSFG		NLSY			
	1	2	1	2	3	4
Recall duration						
$[0, \infty)$	.007*** (.002)	.007*** (.002)	.007*** (.002)	.007*** (.002)	.006*** (.002)	.006*** (.002)
$[0, 48)$	.024** (.008)	.024** (.008)	.019** (.006)	.019** (.006)	.019** (.006)	.019** (.006)
$[120, \infty)$	−.005* (.002)	−.005** (.002)				
Race and ethnicity						
black	.48*** (.07)	.47*** (.07)	.37*** (.10)	.35** (.10)	.17 (.12)	.15 (.12)
hispanic	.02 (.13)	−.03 (.14)	.04 (.15)	.02 (.15)	−.05 (.16)	−.06 (.16)
other	.19 (.21)	.19 (.22)	−.09 (.14)	−.09 (.14)	−.30† (.18)	−.29 (.18)
Background characteristics						
mother's education	.05*** (.01)	.04*** (.01)	−.02 (.01)	−.02 (.01)	.01 (.02)	.01 (.02)
mother's age at first birth	−.01 (.01)	−.01 (.01)	−.00 (.01)	−.00 (.01)	−.00 (.01)	−.00 (.01)
intact family at age 14	−.12 (.07)	−.12 (.07)	−.14 (.09)	−.14 (.09)	−.07 (.10)	−.07 (.10)
catholic	.08 (.08)	.07 (.08)	.01 (.10)	.01 (.10)	.08 (.11)	.08 (.11)
father's SEI $\times 10$					−.03 (.04)	−.03 (.04)
number of siblings					−.02 (.02)	−.02 (.02)
educational expectations					−.02 (.02)	−.02 (.02)
foreign born					−.56* (.21)	−.56* (.21)
reading materials					−.09† (.05)	−.08† (.05)

Table 7. (continued)

	NSFG		NLSY			
	1	2	1	2	3	4
Ability						
AFQT					−.22** (.06)	
vocabulary						−.17* (.08)
arithmetic						−.03 (.07)
numerical operations						.00 (.06)
para. comprehension						−.05 (.08)
Interview characteristics						
length		−.30** (.12)		−.30† (.17)	−.22 (.17)	−.21 (.18)
telephone		.08 (.61)		.46† (.28)	.48† (.28)	.49† (.28)
month of interview		.10*** (.02)				
R's attitude				.22* (.09)	.21* (.09)	.21* (.09)
R's understanding				.01 (.13)	−.13 (.13)	−.14 (.13)
other person present				−.14 (.12)	−.19 (.12)	−.19 (.12)
number of call backs				−.00 (.01)	−.00 (.01)	−.00 (.01)
Constant	−3.35*** (.47)	−3.10*** (.48)	−1.82*** (.29)	−2.33*** (.45)	−2.73*** (.55)	−2.77*** (.55)

Note: Standard errors in parentheses. All models control for dummy variables for missing values of race and ethnicity and family structure at age 14. NLSY models also control for dummy variables for missing values of: ability, father's SEI, educational expectations, and reading materials.

† $p < .10$ * $p < .05$ ** $p < .005$ *** $p < .0005$ (two-tailed test)

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: lwu@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu