

The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country

Megan Beckett
Julie DaVanzo
Narayan Sastry
Constantijn Panis
Christine Peterson

**Reprinted from
The Journal of Human Resources**

**Labor & Population
Reprint Series 01-05**

RAND

RP-956

**The Quality of Retrospective Data:
An Examination of Long-Term
Recall in a Developing Country**

Megan Beckett
Julie DaVanzo
Narayan Sastry
Constantijn Panis
Christine Peterson

**Reprinted from
The Journal of Human Resources**

REPRINTS

Labor & Population

RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. The RAND reprint series contains offprints of RAND research originally published in journals or books. The text is reproduced here with permission of the publisher. RAND® is a registered trademark. RAND's publications do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published 2001 by RAND

1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

1200 South Hayes Street, Arlington, VA 22202-5050

201 North Craig Street, Suite 102, Pittsburgh, PA 15213-1516

RAND URL: <http://www.rand.org/>

To order RAND documents or to obtain additional information, contact

Distribution Services: Telephone: (310) 451-7002; Fax: (310) 451-6915;

Email: order@rand.org

The Quality of Retrospective Data

An Examination of Long-Term Recall in a Developing Country

Megan Beckett
Julie DaVanzo
Narayan Sastry
Constantijn Panis
Christine Peterson

ABSTRACT

The literature on reporting error provides insights into the quality of retrospective reports, particularly as it pertains to short-term recall. Less is understood about the generalizability of these findings to longer-term retrospective reports. We review studies analyzing the quality of retrospective reports in the Malaysian Family Life Surveys (MFLS), fielded in Peninsular Malaysia in 1976 and 1988, and conclude that many of the data quality problems found previously are present in the MFLS. We summarize this literature, place studies based on the MFLS within the context of the reporting error literature, and discuss implications for the design of future surveys.

I. Introduction

Many social science research issues require data on events that have occurred over time. To reduce the cost of collecting such data, or to collect data on the period before a baseline survey or between waves, surveys often ask retrospective questions. Studies of the elderly, such as the Health and Retirement Study (HRS), commonly collect information about life events (such as fertility, marriages, and

The authors are researchers at the RAND Corporation. This research was supported by the National Institute on Child Health and Human Development under grant number P01 HD 28372 and the National Institute on Aging under grant number 1 T32 AG00244-03. The authors are grateful to Robert Moffitt, Robert Groves, and two anonymous reviewers for their insightful comments on earlier drafts of this paper. All figures in this article that are reproduced from other publications are done so with the permission of the publishers. The data used in this article can be obtained beginning October 2002 through September 2005 from Megan Beckett, RAND, 1700 Main Street, Santa Monica CA 90407 email to Megan_Beckett@rand.org.

THE JOURNAL OF HUMAN RESOURCES • XXXVI • 3

jobs) that occurred in the distant past. In some cases, collecting retrospective histories may be the only option for obtaining information at reasonable cost and in a timely manner (for example, women's fertility histories).

Asking retrospective questions on periods as long as several decades has several important advantages over a longitudinal study design. It can take many decades to assemble a prospective, longitudinal data set whereas the same information can be collected retrospectively much quicker. Also, single-round retrospective studies do not require tracking of respondents over long periods and distances in order to conduct periodic reinterviews, which is necessary with a longitudinal study design. Consequently, retrospective reports have a relatively low cost in addition to being quicker to collect.

The desirability of collecting retrospective data that cover a long period of an individual's life is counterbalanced by important questions about the quality of such data. For example, to what extent are events forgotten, especially those that are not highly salient to the respondent but are important to researchers? Does the precision of retrospective answers decrease as the recall period lengthens? What are the correlates of omission? The existing literature on reporting error provides some insight into these issues, particularly as they pertain to short-term retrospective reports. Less is known about the generalizability of these findings to longer-term retrospective reports, such as provided in the Malaysian Family Life Surveys (MFLS) and HRS. Analyses based on selective or inaccurate reports may yield biased and misleading results. A good understanding of the shortcomings of retrospective data can help analysts anticipate (and perhaps correct for) such problems and can provide survey designers and analysts with knowledge about when retrospective data will be adequate. Finally, very little is understood about how data quality lessons from the United States and other developed countries apply to low-income countries characterized by lower levels of education and different cultural and ethnic backgrounds.

We examine data quality issues associated with retrospective reports in a large representative survey of Peninsular Malaysia. The MFLS provides a unique opportunity to examine data quality because it contains repeated retrospective histories across a number of topics, ranging from fertility to migration, that were collected 12 years apart. This has allowed researchers to assess a broader range of data quality issues than is possible using only a single retrospective history. In contrast, most current panel studies that incorporate retrospective histories (such as the HRS) use bounded recall to collect information on events between waves. This precludes a comparison of overlapping retrospective histories and many types of data quality assessments.

The next section reviews what is known about how respondents formulate reports on events that occurred in the past and about common reporting errors and biases. Many studies in this literature employ an experimental or qualitative design or convenience samples. With a few exceptions, the studies based on survey data are based on retrospective reports covering a few weeks to, at most, several years. Section III describes the MFLS, which collects life history data covering periods of up to several decades. Section IV presents original analyses and reviews studies that have analyzed the quality of MFLS retrospective data. We identify the pattern of reporting errors

and biases for retrospective events in the MFLS, drawing parallels with findings from the broader survey research literature. The section is organized around four major issues: the relationship between time since event and reporting error; the association between the salience or social acceptability of an event and its omission; the presence of telescoping effects; and the association of respondent characteristics with reporting error. We close by addressing the implications of reporting error in the MFLS and elsewhere for survey design.

II. Literature on Recall Error

In this section we provide an overview of research that has investigated how respondents retrieve and report retrospective histories, how and why errors in these histories emerge, and the relationship between recall errors and respondent characteristics.

A. Retrieval of Autobiographical Memories

Survey researchers and cognitive psychologists study the reporting process via several methods, such as protocol analysis or the cognitive interview, in which respondents “think aloud” as they answer questions (Sudman, Bradburn, and Schwarz 1996). Researchers have identified four steps in the reporting process (Tourangeau 1984; Strack and Martin 1987). The respondent first interprets the question, and second, retrieves the answer to the question or the information needed to construct an answer. Next he or she formulates an answer based on the recalled information. Finally the respondent edits the answer and decides whether and how to respond. The retrieval and reporting processes are the same for answers to all types of questions, including past events, current status reports, and attitudinal questions.

When discussing recall error with respect to answering autobiographical questions, researchers generally focus on the second stage of the reporting process—errors in memory retrieval. Events can be “forgotten”—or, more generally, omitted—in three ways (Sudman, Bradburn, and Schwarz 1996). First, some events may never be processed, encoded, and stored in memory. Even if an event is stored in memory, finer details of an event may not be recalled. For example, research shows that while an event such as a pregnancy itself is rarely (or slowly) forgotten, the exact date when it ended is the detail most often lost (Warnecke et al. 1997). Second, the event may be missed through retrieval failure (Sudman, Bradburn, and Schwarz 1996). Retrieval failure occurs if the recalling of an event is not “rehearsed” (that is, remembered, thought about, or talked about). Events that are regularly rehearsed include, for example, birth dates and wedding anniversaries. The process of memory rehearsal is thought to strengthen the memory trace and thus increase the ease of recalling an event and decrease the time this takes (Mathiowetz 1999). Third, events can be omitted through the inaccurate reconstruction of a memory (Sudman, Bradburn, and Schwarz 1996). For example, details from similar events that happened to the respondent or that happened to another person who shared their memory with

the respondent may override the original memory or leave a stronger memory trace.

B. Recall Strategies and Recall Errors

Research reveals several types of reporting errors and biases that are characteristic of retrospective reports. The earliest identified and most consistently reported correlate of quality of retrospective reports in the empirical literature is the time since an event occurred.¹ The association between time and recall ability is complex and not yet well understood. Ebbinghaus (1964) in his seminal work identified a negatively accelerating “forgetting curve”—that is, rates of forgetting were highest immediately after the event, then fell and eventually leveled off over time. He also showed that rate of decline in recall ability over time varies by the type of event. For instance, nonsense syllables were forgotten after a few minutes, whereas certain street names and personal events could be recalled after many years (Sudman, Bradburn, and Schwarz 1996). Finally, not all memories decay with time. Warnecke and colleagues (1997) assessed the accuracy of retrospective reports from women about annual physical examinations (for example, breast examinations) they received over a six-year interval by comparing these reports with their HMO records. The authors concluded that “there really was very little decay in recall associated with the length of the recall period” (Warnecke et al. 1997). In another study, length of recall was a significant predictor in the reporting of robberies, but not in the reporting of assaults, burglaries, and larcenies (Dodge 1970). However, for most events, the passage of time results in memory decay. It is also the case that reporting errors and incompleteness increase with the length of recall. For instance, events in the distant past are more likely to suffer from “heaping,” in which respondents provide a “prototypical” response nearest the actual value when they do not recall the actual value (Sudman, Bradburn, and Schwarz 1996).

A second consistent finding in the experimental and empirical literature is that events that are highly salient to the respondent are better recalled. Salience depends on many factors, including the relationship of the event to respondent’s lifestyle or self-identity. In a longitudinal survey of older Taiwanese, the most important predictor of the likelihood that a health condition that was reported at one interview was also reported at a subsequent interview was the initial severity of the health condition, with severity defined as the extent to which the condition limited daily activities, such as walking or bathing (Beckett et al. 2000).

A third finding from the empirical literature is the occurrence of telescoping, in which more events are recalled as having occurred in the most recent period and fewer in the more distant past. Three factors are believed to contribute to telescoping (Rubin and Baddeley 1989; Huttenlocher, Hedges, and Bradburn 1990). First, normal forgetting is greater for events that took place further back in time. Second, errors in dating events are random or unbiased but increase with time since the event occurred. Third, there are intrusions of events that occurred outside the reference period. For example, if respondents are asked to report on the number of office visits in the past month, they might easily include earlier visits by mistake. The combina-

1. See Mathiowetz (1999) for a review of these studies.

tion of the second and third factors means that events that occurred further back in time will be more likely to be remembered as having occurred within the recent past. This process leads to errors in reports of the timing of events and tends to produce an overestimate of the number of events in the recent reporting interval.²

Closely related to telescoping is a strategy called the "availability heuristic" (Tversky and Kahneman 1973), also referred to as the "accessibility principle" (Brown, Rips, and Shevell 1986). These terms refer to the phenomenon that a memory that is easily recalled is believed by the respondent to be more frequent (and recent) than is a memory that is remembered with more effort. Conversely, an event that is recurring but recalled with great difficulty may be believed by the respondent to be rarer than it really is. Both qualitative and survey data lend empirical support to the availability heuristic. For example, Bowman, Sanson-Fisher, and Redman (1997) matched women's reports about pap smears in the past three years with pathology records and examined the correlates of a match between self-reports and medical records. When asked how sure they were of their responses, women who said they were "very certain" were more likely to be accurate (that is, to have had a smear according to medical records) than were "less certain" women. The empirical evidence is mixed on whether the availability heuristic contributes to forward telescoping (Thompson, Skowronski, and Lee 1988).

When a respondent cannot recall events clearly, several strategies have been identified to aid in reconstructing them (Sudman, Bradburn, and Schwarz 1996), although some of these strategies can lead to reporting bias. One common strategy, for instance, is to probe for individual instances of the event by first asking for a count of them. However, this may lead to heaping of counts. A second strategy is to use information about current-status attitudes or behaviors to infer past attitudes and behaviors. The resulting "stability bias" is evident in several empirical studies that examined retrospective reports of items such as income and drug use (Whitney 1954; Collins et al. 1985). Retrospective reports were heavily influenced by current behaviors, indicating that respondents assumed a greater degree of stability in their behavior than really existed. However there may be times when retrospective reports are more accurate than current status reports. Analysis of the National Health and Nutrition Examination Survey (NHANES-I) Epidemiologic Follow-up Study (NHEFS) indicates that respondents may more accurately report past heavy drinking than current heavy drinking (Dufour et al. 1990). In this study, respondents' current-status reports at time of the initial interview were compared with their retrospective reports for the same period during a reinterview about 10-years later. Respondents may have been less embarrassed to report their sensitive behaviors in the past than in the present.

2. A troublesome counteracting source of error that results in backward displacement occurs when there is an incentive to report events as having occurred before a reference period in order to avoid detailed follow-up questions and thereby save time. This phenomenon has been documented in the Demographic and Health Surveys (Arnold 1990). For births that occurred within a fixed interval before the survey, respondents were asked detailed questions about breastfeeding, contraceptive use, and health. Interviewers who wanted to decrease their workload could deliberately misreport the date of births of children as having occurred outside this window.

C. Respondent Characteristics Associated with Recall Error

Many empirical studies drawing on survey data have investigated respondent characteristics that may be associated with recall error. Some of these studies are based on special populations, such as HMO members, for whom it is possible to link survey reports with external data such as medical records. The implicit assumption is that the external record is the "gold standard," although one cannot necessarily assume this is true. For example, some debilitating but nonfatal conditions, such as arthritis, are highly salient to the respondent but do not require ongoing medical attention; these may be more accurately reported by the respondent than by a physician (Beckett et al. 2000). There is limited experimental evidence of systematic differences in recall error according to respondent characteristics. In one experiment, based on data from a metropolitan sample, Herzog and Rodgers (1989) showed a systematic negative relationship between age and respondents' ability to recall the topics of questions asked earlier in a personal interview. There was a somewhat weaker relationship between recall ability and age when education was controlled.

Empirical research on respondent characteristics that are associated with poor reporting has yielded inconsistent findings. Marquis, Cannell, and Laurent (1972) reviewed early studies that compare self-reports of health events, such as hospitalizations, office visits, and number of chronic conditions, with medical records. The periods for which respondents were asked to recollect health events in these studies were relatively short, ranging from the previous two weeks to the previous year. Generally, these studies confirmed that a shorter time since an event was associated with better recall, as was the greater salience of an event. However, they revealed no consistent patterns of underreporting by respondent characteristics including education, sex, family income, or race/ethnicity of the respondent. Likewise, the Bowman, Sanson-Fisher and Redman (1997) study mentioned above found that none of 25 items (covering demographic, attitudinal, and behavioral factors) predicted a match between self-reports and medical records. Unfortunately the authors do not note which, if any, of these 25 items may be influencing the match through their effect on the degree of certainty with which a respondent characterizes her response. For example, better-educated women may be more aware of their past use of health care and thus more sure of their response; since this latter factor is controlled for, the association between education and a match may be obscured.

In contrast to reporting about health status, several recent studies show differences in recall of sexual activity according to race, sex, cognitive ability, and other factors. Upchurch et al. (2000) indicated that 11 percent of respondents in the National Longitudinal Survey of Adolescent Health who reported that they were sexually active at the time of the baseline interview denied this fact in the first follow-up, one year later. Males, black females, and teens with lower tested ability were more likely to rescind their report. Upchurch and colleagues also found that males, and especially black males, had larger differences between their two reports in the reported month and year of first sexual experience as did teens with low verbal ability. Wu, Martin, and Long (1998) examined how recall of the calendar month of first intercourse varied with respondent characteristics using data from the 1979–94 waves of the National Longitudinal Survey of Youth (prevalence of missing 16 percent) and the

1988 National Survey of Family Growth (20 percent missing). They found that the ability to recall this information decreased with the duration of recall, was lower for Hispanics and especially blacks, was negatively associated with interview length, and, finally, was higher for the foreign-born, for women who initiated sexual activity after the teen years, and for those who scored higher on a test of ability. Fu et al. (1998) found that higher level of education was associated with a lower level of reporting for induced abortions in the main interview—compared with a self-report procedure in the 1995 National Survey of Family Growth.

The studies above generally examine the completeness of retrospective reports for relatively short recall periods. It may be the case, however, that respondent characteristics more strongly affect the recall of events that occurred in the more distant past or that these relationships are different in settings outside the U.S. Using the NHEFS data set, Beckett et al. (2000) examined the correlates of consistent reports of chronic conditions between the 1971–75 and 1982–84 waves of the NHEFS data set and also in a survey of older Taiwanese (between 1989 and 1993 and between 1993 and 1996). A consistent report was defined as the percentage of people who reported a chronic condition at baseline who also reported it at follow-up. The assumption is that once a person has been diagnosed with a certain chronic condition, they should always report having had this diagnosis. A nontrivial proportion of respondents with a diagnosed chronic condition in the baseline did not report this diagnosis in the follow-up. This ranged from 14 percent of people with hypertension to 53 percent of people with arthritis. In Taiwan, education was positively associated with consistent reporting, whereas in the U.S. being a female was the strongest predictor. In the survey of older Taiwanese, education and age were significantly associated with consistent reporting in some models. Neither age nor education was significantly related with consistent reporting among older Americans. In contrast, Hahn, Eaker, and Rolka (1997) used the same U.S. NHEFS sample to compare the reliability of reported age at menopause at two points in time approximately ten years apart. They found that women with higher education, controlling for age and years since event, were more reliable reporters than less educated women.

III. Data: The First and Second Malaysian Family Life Surveys

The First and Second Malaysian Family Life Surveys (MFLS-1 and MFLS-2), fielded in Peninsular Malaysia in 1976–77 and 1988–89, respectively, are comprehensive household demographic and socioeconomic surveys that represent two of the most widely used databases collected in a developing country.³ MFLS-1 was designed with a primary interest in studying influences on fertility and infant feeding and survival. For this reason the sample was restricted to ever-married women of childbearing age. In addition to collecting life histories on the variables of primary interest, retrospective data were also collected on other life areas, for example, migration, marriage, and her own and her husband's employment, that may

3. RAND and Malaysian collaborators are planning to field the Third MFLS in 2001.

have influenced decisions that would affect fertility and child health (for example, where the women was living at the time of a birth, how long she had been married, and the type of work her husband was doing). The MFLS-1 was conducted in three rounds, four months apart, in an effort to document seasonality in some key areas (for example, economic activities). MFLS-1 interviewed 1,262 ever-married women up to age 50 and their husbands (DaVanzo and Butz 1978; DaVanzo et al. 1978).

The MFLS-2 contained four samples, two of which—the Panel Sample and MFLS-2 New Sample—are used in the studies discussed in this paper. The Panel Sample consists of the ever-married female respondents from MFLS-1 who were successfully contacted and reinterviewed in MFLS-2; the re-interview rate was 72 percent. The MFLS-2 New Sample is representative of women of reproductive age (under age 50), and, like in MFLS-1, husbands were interviewed as well.⁴ The MFLS-2 questionnaires were fairly similar to those used in MFLS-1. In particular, they again collected full retrospective life histories on the life areas noted above.⁵ Hence, for MFLS-2 Panel respondents (who were aged 61 or younger in 1988) replicate histories are available for the period before the 1976 MFLS-1. The Panel sample has been used in a number of studies to compare 1988 retrospective reports about pre-1976 events to MFLS-1 reports about these same events.

Several strategies were employed in the MFLS event histories to collect complete and accurate retrospective data. If respondents were unable to give the exact month of an event, they were asked whether it was in the early, middle, or late part of the year, and this was recorded using special codes for these parts of the year. If a respondent was unable to report the year of the outcome, she was asked about her age at the time. Interviewers probed in terms of other events, such as holidays or other events in the respondents' life history. Although the survey did not formally use a calendar, interviewers were encouraged in the retrospective questions about each life area to cross-reference events in other life areas; for example, did a move occur before or after the birth of a particular child, and if afterward, how old was the child at the time?

For the Panel Sample, dates of birth for all MFLS-1 household members and any other of the women's children who were alive at the time of the MFLS-1 interview were brought forward in MFLS-2. (This was done to update the characteristics of these individuals, in part for the purpose of selecting respondents for the MFLS-2 Children Sample.) Hence, for the Panel sample, there is typically not an independent report of birth dates for children who were alive at the time of the MFLS-1 interview.

The analyses of the quality of MFLS retrospective data have considered a number of different topics, including fertility and related topics (for example, infant and child mortality, birthweight, infant feeding, and post-partum amenorrhea), migration, and employment.⁶

4. The two other samples in MFLS-2 were the Children Sample, consisting of the grown children of women in the Panel Sample, and the Senior Sample, consisting of households with a person aged 50 or older. (For more information about MFLS-2, see DaVanzo et al. 1993).

5. The full life histories were collected again because the technology at the time did not readily facilitate bringing forward the life-history data to provide benchmarks for each life area.

6. More than 250 papers have been published using the MFLS (for a list of publications, see <http://www.rand.org/FLS/MFLS/MFLSpubs.html>). Some investigated data quality as their main focus; many others include a side analysis of data quality or have findings that have implications about data quality.

IV. Quality of Retrospective Reports in MFLS

In this section, we review in detail the variation in quality of reporting in the MFLS across characteristics of events and respondents. In many cases, the findings for Malaysia, which are based on long-term retrospective reports, accord well with what is found in the survey literature. For example, the further back in time an event occurred, the poorer the quality of reporting. The papers we review utilized a variety of techniques to measure data quality of retrospective reports. Below we briefly describe some of these techniques. For more discussion of the range of methods employed to analyze data quality of MFLS retrospective reports, see Beckett et al. (1998). We begin this section with an overview of data quality in MFLS.

A. Overview of MFLS Data Quality

Overall, analyses of data quality of the MFLS retrospective histories show that the quality is quite good⁷—or at least consistent between MFLS-1 and MFLS-2. This is the case for reports on topics ranging from fertility to work status to migration events, especially at the aggregate level. In Table 1 we show the correspondence between summary statistics for the reports of Panel respondents in MFLS-1 and MFLS-2 (that is, the matched sample of respondents who were interviewed in both waves) across a variety of topics. For breastfeeding, birthweight, and child mortality, we compare the two sets of pre-1976 reports and for contraceptive use, marriage, and work we compare MFLS-1 current status reports with MFLS-2 retrospective reports about this same point in time. Current status reports (from MFLS-1) should be less prone to reporting error than retrospective reports (from MFLS-2) since in MFLS-1 the respondents are reporting on contemporaneous events.

Overall, Panel respondents provided very consistent reports across topics. Summary statistics differ significantly for just two of the six topics: breastfeeding and marital status. In MFLS-2, Panel respondents said they breastfed 83 percent of their infants born before 1976, compared with 79 percent based on reports for these same infants from MFLS-1. A small change in the qualifier for the main breastfeeding question between MFLS-1 and MFLS-2 may have contributed to at least some of this difference. In MFLS-1, the question “Did you breastfeed (NAME OF CHILD)?” was followed by the qualifier “I want to know even if it was just for a few days,” whereas in MFLS-2 the qualifier was changed to “I want to know even if you just tried once or twice.” Later, we review other factors, such as changes in social norms related to breastfeeding that may have contributed to discrepant breastfeeding reports. Not only are women more likely to say they breastfed a given child in MFLS-2 they also reported breastfeeding for a significantly longer duration. The largest difference is observed for reports of whether a woman was married in 1976. In MFLS-1, about 95 percent of women reported they were currently married, whereas in MFLS-2 only 84 percent of women reported they were married at that time. The perfect correspondence in age of marriage probably occurs because reports

7. See Beckett et al. (1998) for summary of all of the studies analyzing data quality.

Table 1
Comparison of Summary Statistics for Panel Respondents, MFLS-1 and MFLS-2

	MFLS-1	MFLS-2
Breastfeeding of children born before 1976		
Percent ever breastfed	79.1	82.6*
Mean duration (months)	13.8	15.5*
Percent breastfed < 4 months	25.0	21.5*
Birthweight of children born before 1976		
Mean weight (lbs)	6.8	6.7
Percent \leq 5.5 lbs	16.1	16.0
Mortality among children born before 1976		
Number of children died per woman	1.2	1.2
Contraceptive use at time of MFLS-1 interview		
Percent practicing any method of contraception	35.3	37.5
Marital status at time of MFLS-1 interview		
Percent married	94.7	84.3*
Age at marriage	19.1	19.1
1977 Work status at time of MFLS-1 interview		
Percent ever worked	80.8	78.3
Mean age 1976 job began	26.4	23.1

* MFLS-1 and MFLS-2 statistics differ significantly at $p < .05$.

are compared only for women who consistently reported being currently married (as of 1976) at both waves.

When individual reports about the same events are matched, the quality of reporting is lower than is reporting in the aggregate, but is still generally high. As part of our analysis of the MFLS-1 current status and MFLS-2 retrospective reports, we computed the Kappa coefficient to assess agreement between the two reports.⁸ Using guidelines for the interpretation of Kappa coefficients provided by Landis and Koch (1977), the level of agreement for these variables was substantial for current breastfeeding status (Kappa = 0.75), moderate for marital status (0.43), and fair for contraceptive use (0.38) and whether the woman ever worked (0.39). The consistency of matched reports is lower for reported details about events. For example, the level of agreement about whether a child was breastfed is substantial (Kappa = 0.67), but is poor for the number of months breastfed (0.16). Using a dichotomous distinction (for example, whether the child was breastfed for less than four months) increases the consistency of reporting compared with the continuous report in the matched sample and in a comparison of the MFLS-1 and MFLS-2 New Samples

8. Defined as $(O-E)/(N-E)$, where O is the observed number of respondents giving the same answer at two times; E is the expected number of same answers, given the marginal probabilities of responses in the two waves; and N is the number of respondents that responded both times.

Table 2
Summary of Retrospective Reports on Migration for the Pre-1976 Period

	Males		Females	
	All	All	Inter-district	Intra-district
Number of moves in MFLS-1	968	1,878	690	1,188
Number of pre-1976 moves in MFLS-2	884	1,652	700	952
Number of pre-1976 moves reported in MFLS-2 as a percentage of number of moves reported in MFLS-1	91.3	88.0	101.4	80.1
Percent of respondents reporting same number of pre-1976 moves in MFLS-1 and MFLS-2	57.0	42.8	70.2	44.8

Table adapted from Smith and Thomas (1997).

(Peterson 1993). A similar pattern of higher agreement on dichotomized birthweight (≤ 5.5 lbs. versus > 5.5 lbs.) compared to reported exact birthweight also occurs (Peterson 1993). The vast majority (91 percent) of reports on whether a birthweight was less than 5.5 pounds agreed, whereas only about one quarter of birthweights were exactly the same (when pounds and ounces were reported in MFLS-1 and MFLS-2).

B. Time Since Event and Recall

Analyses of the MFLS data reveal that some events may be forgotten over time. Smith and Thomas (1997) examined the reliability of migration histories covering all male inter-district⁹ moves and female inter- and intra-district between age 15 and the time of the MFLS-1 interview. Table 2 summarizes the comparison of moves for the pre-1976 period reported at both waves. In MFLS-1, men reported 968 moves; in MFLS-2, men reported 91 percent as many moves for the same period. Women reported 1,878 total moves in MFLS-1; in MFLS-2, they reported 88 percent as many moves for the same period. Because women's migration histories collected information on both inter- and intra-district moves, one can see that the distance of the move is strongly related with the level of apparent omission of events. Women actually reported slightly more moves that involved crossing a district boundary in MFLS-2 (700) than they did in MFLS-1 (690). Below we discuss in more detail the difference in reliability of reporting about events that have greater salience to the

9. A district in Malaysia is analogous to a county in the U.S.

Table 3

Summary of Retrospective Reports on Pre-1976 Infant Deaths in the Matched Sample

Number of infant deaths in MFLS-1	185
Number of pre-1976 infant deaths in MFLS-2	191
Difference	6
Percent of respondents reporting more infant deaths in MFLS-2	15.7
Percent of respondents reporting same number of infant deaths in MFLS-1 and MFLS-2	73.0
Percent of respondents reporting fewer infant deaths in MFLS-2	11.3

respondent. Smith and Thomas (1997) compared the number of pre-1976 moves reported in both surveys and found a higher rate of inconsistency: 57 percent of men and 70 percent of women reported the same number of inter-district moves (and 45 percent of women the same number of intra-district moves) in MFLS-1 and MFLS-2.¹⁰

In Table 3 we examine the consistency of reporting for infant deaths, presumably an event of high salience. The table summarizes the number of infant deaths occurring before 1976 that were reported by Panel respondents in MFLS-1 and MFLS-2. A total of 185 infant deaths were reported in MFLS-1, compared to 191 pre-1976 infant deaths in MFLS-2. As with inter-district moves for female Panel respondents, in the aggregate, women are not under-reporting infant deaths 12 years after the initial report. Most women (73 percent) provide consistent reports of the number of infant deaths. Of the remainder, slightly more women (16 percent) report more pre-1976 infant deaths in MFLS-2 than report fewer pre-1976 infant deaths (11 percent). As discussed below, it is likely—especially with infant deaths—that inconsistent reports are the result of blurring of details of events, such as the date of the child's death or the child's age at the time.

Panis and Lillard (1994) examined trends in rates of miscarriage, an event that is presumably less salient than pregnancies ending in a live birth, using data from the MFLS-2 New Sample. Figure 1, reproduced from their study, shows the survival curves for pregnancies before 1960, between 1960 and 1980, and after 1980. Over a period characterized by improving health conditions and declining infant, child, and maternal mortality, these data imply rising rates of miscarriage (Panis and Lillard 1994).¹¹ To explain this result, Panis and Lillard posited that women forgot about miscarriages that occurred in the more distant past—especially those from the first three months of pregnancy. They showed that pre-1960 pregnancies failed to result in a live birth in 4.4 percent of the cases, with 59 percent of these miscarriages occurring in the first three months of pregnancy. Among post-1980 pregnancies,

10. District splits between MFLS-1 and MFLS-2 were accounted for by reclassifying MFLS-2 interdistrict migrations as intra-district migrations where appropriate.

11. Note that induced abortions are not included among miscarriages.

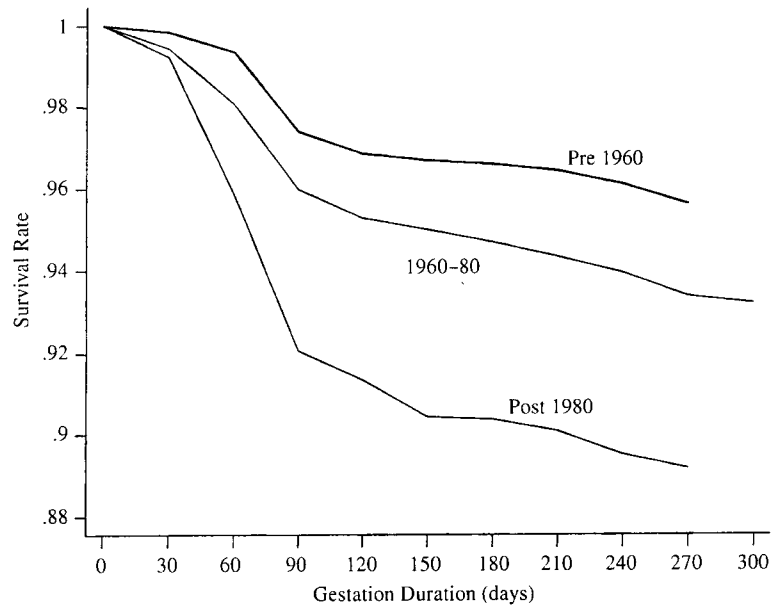


Figure 1

Empirical fetal survivor function (by pregnancy cohort) (Panis and Lillard 1994). Reproduced with permission from Elsevier Science.

however, 10.8 percent resulted in a miscarriage and 73 percent of miscarriages occurred in the first three months. In other words, the apparent increases in rates of miscarriage over time implied by the MFLS-2 New Sample data were largely due to an underreporting of early miscarriages in the distant past.

C. Time Since Event and Reporting of Event Details

Analyses of the MFLS reveal that not only are events omitted over time, but also details about past events can blur, resulting in incomplete and inexact responses. In particular, there is a tendency to provide less precise answers about dates and other numerical information as one moves backwards in time and the period of recall increases. For example, mothers are less likely to provide the exact month and year of birth for children born in the more distant past (Haaga 1986; Sine 1993; Sine and Peterson 1993) and birthweight is reported less exactly for births that occurred further back in time.¹² Peterson (1993) matched reported birthweights among births with an exact birthweight reported in MFLS-1 and MFLS-2. Two-thirds of reported birthweights in 1976 had non-zero ounces, which was the case for only half of 1988

12. In both MFLS-1 and MFLS-2, women were first asked to report the exact birthweight, in pounds and ounces, of each of their liveborn children. If a woman did not know the exact birthweight, she was asked to provide an approximate birthweight using five categories, ranging from "very low" to "heavy."

Table 4

Percentage Reporting Zero Ounces Birthweight in MFLS-1 and the MFLS-2 New Sample, by Time Period

Year of Child's Birth	MFLS-1		Year of Child's Birth	MFLS-2 New Sample	
1971-76	37.2	(1,185)	1983-88	33.4	(2,013)
1965-70	39.1	(1,157)	1977-82	37.2	(1,721)
1959-64	40.0	(881)	1971-76	39.3	(1,089)
1954-58	47.6	(466)	1965-70	39.6	(535)
Before 1954	52.6	(268)	Before 1965	38.5	(239)
Total	40.6	(3,957)	Total	39.5	(5,597)

Based on births where exact, rather than approximate, birthweight reported. Cell sizes are in parentheses.

reports (for these same pre-1976 births). If birthweights were in reality not more likely to be clustered on whole pounds we would expect 1/16 of birthweight reports (6.25 percent) to involve 0 ounces. Hence, we observe a tendency of respondents to heap on 0 ounces in MFLS-1, and an even greater such tendency in MFLS-2 (for the same pre-1976 births). When considering all MFLS-1 (or MFLS-2) births with an exact birthweight, Table 4 shows that the proportion of births with 0 ounces increases monotonically with recall length (with one minor exception). In MFLS-1, the percentage of the most recent (1971-76) births with a report of 0 ounces is 37 percent compared with a report of 0 ounces for 53 percent of the earliest (before 1954) births reported in the same survey. The extent of increase in reports of 0 ounces with increasing length of recall is more modest in MFLS-2.

Greater time since event is also more likely to result in respondents providing less exact reports about event details, such as dates or durations. Responses to all MFLS duration questions were recorded in units of days, weeks, or months (and sometimes also hours or years)—whichever was used (or was closest to that used) by the respondent. Table 5 shows the percentage distribution of units of time used in reporting duration of breastfeeding by birth cohort for MFLS-1 and the MFLS-2 New Sample. In the MFLS-1 sample, the most precise time unit (days) was used more often for the most recent birth cohort (1971-76) than for any other cohort. The grossest time unit (months) was less often used for the 1971-76 birth cohort compared with other cohorts. In MFLS-2, two time units were added (hours and years), though again we see an association between year of child's birth cohort and use of gross time units. The percentage of breastfeeding durations reported in years increases twofold from the 1983-88 birth cohort to the pre-1965 birth cohort, while the percentage reporting in days drops from 10 percent to 5 percent. In the matched sample of births, reports on breastfeeding duration in MFLS-1 that used days or weeks are generally reported in grosser time units in MFLS-2 (see Table 6). Change to a finer time unit in MFLS-2 occurred in very few instances.

Table 5
Percent Distribution of Time Units Used to Report Duration of Breastfeeding in MFLS-1 and MFLS-2 New Sample, by Child's Year of Birth

MFLS-1						
Year of Child's Birth	Days	Weeks	Months	(n)		
1971-76	14.1	3.3	32.6	1,042		
1965-70	7.7	2.0	90.3	1,217		
1959-64	7.5	1.3	91.1	1,114		
1954-58	3.1	1.6	90.4	695		
Before 1954	7.0	3.1	39.7	556		
Total	9.1	2.2	38.7	4,624		

MFLS-2						
Year of Child's Birth	Hours	Days	Weeks	Months	Years	(n)
1983-88	0.5	10.2	8.3	70.5	10.6	1,454
1977-82	0.1	10.5	7.6	66.6	15.3	1,494
1971-76	0.2	6.4	5.5	71.6	16.3	969
1965-70	0.2	6.1	3.3	70.8	19.6	541
Before 1965	0.0	5.0	2.5	71.7	20.8	283
Total	0.2	8.7	6.6	59.6	14.9	4,741

Samples restricted to those who reported a completed duration of breastfeeding.

Heaping is another type of inexact report that is associated with time since event.¹³ MFLS data on post-partum amenorrhea (the interval after a pregnancy before menstruation returns) is subject to heaping on responses that are multiples of six months (Haaga 1986; Sine and Peterson 1993).¹⁴ Moreover, Haaga's results show that heap-

13. Heaping need not always be indicative of poor data quality. For example, if health professionals recommend that women breastfeed for 12 months or there is a norm to wean babies around their first birthday, one would expect to find a corresponding peak in reporting breastfeeding duration. Haaga (1986) examined this issue using two approaches. First, he searched the anthropological literature for known norms in Malaysia with respect to breastfeeding; he did not find any such evidence. Second, he compared the distribution of retrospective responses to current status reports. If heaping reflects socially regulated behavior rather than reporting patterns the same distribution of responses should be seen in retrospective and contemporaneous data. This was not the case.

14. This has also been found for data from other studies on breastfeeding and post-partum amenorrhea (Knodel and Debavalya 1980; Diamond, McDonald, and Shah 1986).

Table 6
Percent Distribution of Time Units of Breastfeeding Duration Reported in MFLS-2 by Time Unit Reported in MFLS-1, for Matched Sample of Births

Time Unit Used in MFLS-1	Time Unit Used in MFLS-2					(n)
	Hours	Days	Weeks	Months	Years	
Days	1.5	35.2	4.1	52.6	6.7	270
Weeks	0.0	2.1	14.6	79.2	4.2	48
Months	0.2	1.8	1.1	70.9	26.0	3,034
(n)	10	151	51	2,331	809	3,352

◻ = Finer time unit in MFLS-2

◼ = Grosser time unit in MFLS-2

Samples restricted to those who reported a completed duration of breastfeeding in each survey. Hours and years codes not used in MFLS-1.

ing is stronger for pregnancies that occurred further back in time suggesting that as the duration of the recall period increases, respondents are increasingly estimating the duration rather than providing an exact report.

Klerman (1995) undertook a detailed comparison of MFLS-1 and MFLS-2 reports on duration of breastfeeding for infants born before 1976 who breastfed. Figure 2 shows the distribution of MFLS-1 responses for breastfeeding duration for the three most common MFLS-2 responses (1, 12, and 24 months representing 10 percent, 18 percent, and 13 percent of the MFLS-2 responses respectively). Consistent responses between the two waves were rare—53 percent for 1-month responses in MFLS-2, 26 percent for 12-month responses, and 36 percent for 24-month responses. More problematically, changes in responses were usually to non-adjacent categories, and in some cases these changes were from one heaping point to another heaping point (for example, 24 months to 12 months). A close examination of Figure 2 reveals that MFLS-2 reports tend to be of shorter duration than MFLS-1 reports. For example, among MFLS-2 reports of 12 months the most frequently matched MFLS-1 reports were, in descending order, 12, 24, 18, 10, and 16 months. The analytic implication of such misreporting is that reports on durations for earlier births will be underestimated, and will bias upward the trends in duration of breastfeeding, and bias downwards the association between breastfeeding and health outcomes, such as child survival, the further back in time one goes.

In some cases, digit preference does not bias the aggregate report. We analyzed changes in reported birthweights between MFLS-1 and MFLS-2 for the sample of matched births and found very strong digit preference for (in descending order) 0 ounces, eight ounces, and the one to seven ounce category. There is an apparent aversion towards reporting nine to 15 ounces. Table 7 shows the percentage distributions of MFLS-2 birthweights for these four categories of MFLS-1 birthweights. The cells are shaded in accordance with the level of agreement between the two reports.

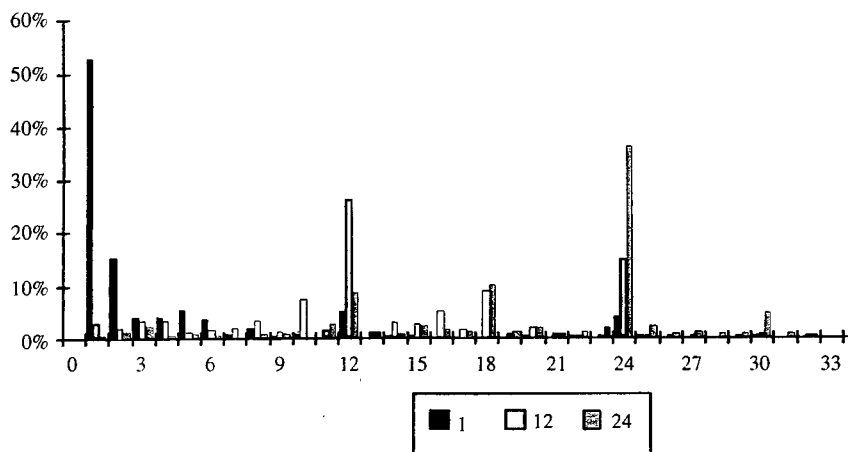


Figure 2

Distribution of MFLS-1 responses for MFLS-2 responses of one, 12, and 24 months, from subsample with two retrospective reports (Klerman 1995).

Reproduced with permission from the Society for the Study of Social Biology.

The preference for 0 ounces (691/1705 or 41 percent of all MFLS-2 reports) and eight ounces (360/1705 or 21 percent of all MFLS-2 reports) is strong. Birthweights involving 0 or eight ounces in MFLS-1 are the most likely to be reported consistently over time—in 40 percent and 24 percent of cases, respectively, the same number of ounces was reported in MFLS-2. Among reports that differ between MFLS-1 and MFLS-2, birthweight ounces in MFLS-2 tend to gravitate toward the preferred digits of 0 and eight ounces and the one to seven ounce category, and away from nine to 15 ounces (even among persons who started with nine to 15 ounces). This preference for a rounded answer is so strong that a higher percentage of reports involving 0 ounces in MFLS-1 move down (10 percent) or up (9 percent) exactly one pound than move up (7.1 percent) or down (4.2 percent) by half a pound. Movers from 0 ounces are five times as likely to move up a few ounces to one to seven ounces, as they are to move down a few ounces to nine to 15 ounces (9.8 percent versus 2.0 percent). The aversion for nine to 15 ounces is especially evident among MFLS-1 reports involving eight ounces: 21.7 percent of the corresponding MFLS-2 reports were down to one to seven category compared with 2.5 percent of reports which moved up a few ounces to the nine to 15 category. Finally, this table illustrates a tendency for MFLS-2 reports to round down to the nearest whole number of pounds rather than to round up. Among MFLS-1 reports involving eight ounces, 18.1 percent of corresponding MFLS-2 reports were rounded down to nearest whole pound compared with 10.0 percent which (more correctly) rounded up to the nearest whole pound. Despite the digit preferences and tendency to round down rather than up, the net result is that matched reports of birthweights are highly consistent overall (as discussed above).

Table 7
Cross Tabulation of Exact Birthweight Reported in MFLS-1 Against Exact Birthweight Reported in MFLS-2, for Matched Sample of Births

MFLS-1	MFLS-2										Some other answer	Total (n)	
	MFLS-1 # lbs. - 1 lb.					MFLS-1 # lbs. + 1 lb.							
	0	1-7	8	9-15	0	1-7	8	9-15	0	1-7			8
0	10.0	4.3	4.2	2.0	40.2	9.8	7.1	1.3	9.0	2.5	2.5	8.1	100% (691)
1-7	←	11.5	→	1.3	15.2	40.6	9.7	3.5	7.5	←	4.0	→	100% (453)
8	←	8.4	→	1.1	18.1	21.7	23.9	2.5	10.0	←	5.0	→	100% (360)
9-15	←	←	6.0	→	14.4	13.9	9.5	34.3	8.0	←	1.0	→	100% (201)

Based on births where exact, rather than approximate, birthweight reported in MFLS-1 and MFLS-2. The column subheadings under MFLS-2 refer to groups where the MFLS-1 and MFLS-2 reports are within one pound and differentiate these reports as to whether the difference is negative (MFLS-1 # lbs. - 1 lb.), the same or zero, or positive (MFLS-1 # lbs. + 1 lb.).

Key:

←	→	←	→
Rounded down more than 1 pound	Rounded down within a half pound	Basic agreement	Rounded up within a half pound
			Rounded up more than 1 pound

D. Salience of Events and Recall

As found elsewhere in the literature, analyses of MFLS indicate that salience of an event is correlated with the likelihood that the event is reported (and reported correctly). In their comparison of MFLS-1 and MFLS-2 migration reports, Smith and Thomas (1997) tested the hypothesis that moves that were (believed to be) more salient to the respondent were better remembered. Several possible measures of salience were identified: first move, most recent move, moves involving longer time at the new place, moves involving greater distance, and moves linked to other landmark events (such as marriage or starting a new job). A multivariate model was estimated to identify which demographic factors and measures of salience predicted reporting a move in MFLS-2, conditional on the move being reported in MFLS-1. For both men and women, duration at new residence was a major predictor of a move being reported in both waves. Short-term moves (that is, moves that involve staying at the destination for less than six months), which may involve circular migration, were quickly forgotten. About one-third of short-term moves reported by men in MFLS-1 were also reported in MFLS-2. This percentage increased to nearly 80 percent for longer term moves that lasted at least 20 years. Among women, distance of the move was positively related to reporting a move in both waves.

Kuate Defo and DaVanzo (1996) evaluated the reliability of reported reasons for no or short breastfeeding in the matched sample. Reliability is greatest for reasons linked to specific (and salient) events, such as mother returning to work or child dying.

E. Social Sensitivity of Events and Recall

As with most surveys, abortions appear to be under-reported by MFLS respondents (Haaga 1988; Sine and Peterson 1993; Panis and Lillard 1994). It is not always the case, however, that the social sensitivity of an event leads to under-reporting. In some cases, respondents may reconstruct memories—based on new social mores—in such a way that past events are overreported. This phenomenon, referred to in the literature as “stability bias,” occurs when respondents infer previous behavior based on current behavior and can lead to an underestimate of changes in behavior. MFLS-1 data indicate a long-term downward trend in the proportion of infants breastfed, from 92 percent in 1958 to 78 percent in 1977 (Haaga 1986), but suggested that the general trend had reversed during the 1970s. In particular, breastfeeding rates increased significantly from 1970–74 (75 percent) to 1975–77 (79 percent). In an extension of this analysis, DaVanzo et al. (1994) used MFLS-2 data to examine whether the upward trend evident in the mid-1970s was real or a data artifact. Figure 3, from their study, shows the percentage of Malaysian infants breastfed in the MFLS-1 and in the MFLS-2 New Sample by year of birth of infants from 1955 to 1988. The same rise in breastfeeding from 1975 on that is observed in MFLS-1 is also apparent in MFLS-2 after 1976. However, the percentage of breastfed infants in the MFLS-2 New Sample is consistently and markedly higher (about three to ten percentage points) than the percentage derived from the MFLS-1 (DaVanzo et al. 1994). Although the vast majority (93 percent) of Panel respondents provided the same response in both surveys about whether they breastfed each child, on average

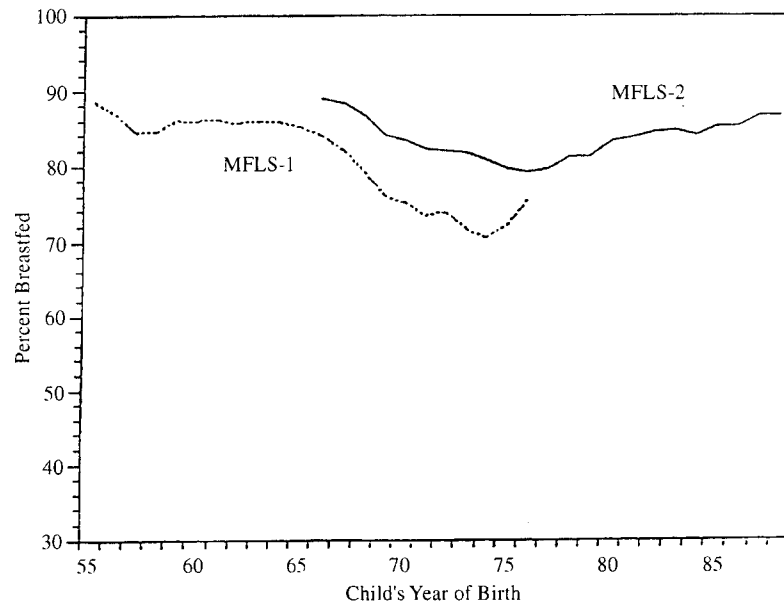


Figure 3

Trends in the percent of Malaysian infants ever breastfed, MFLS-1 and MFLS-2 New Sample (DaVanzo et al. 1994).

Reproduced with permission from RAND.

they were significantly more likely in 1988 to have reported breastfeeding for pre-1976 births (87.2 percent) than they were to have reported this for the same births in 1976 (84.3 percent) (DaVanzo et al. 1994).

One explanation for why women reported higher levels of breastfeeding relates to "stability bias." In the absence of a clear memory about one's behavior in the past, the respondent may look to current social norms of what it means (in this example) to be a good mother and infer her previous behavior based on these current norms. In the late 1980s, when breastfeeding was the socially normative form of infant feeding, if a woman was unsure about whether she breastfed a child born in the past, she may have considered herself a "good mother" according to contemporary standards, and assumed that she had breastfed her child. A second explanation is that social desirability led women to consciously over-report breastfeeding when it was more popular (DaVanzo et al. 1994).

A third alternative explanation for the change in level of breastfeeding relates to the change (described above) in the way that the two surveys qualified the main breastfeeding question to identify less successful attempts to breastfed (DaVanzo et al. 1994). This explanation may account for up to a quarter of the inconsistent reports on breastfeeding. Of the MFLS-1 non-breastfed children who were reported to have been breastfed according to MFLS-2, 13 percent reported a duration of three days or less, and 25 percent reported a duration of one week or less (DaVanzo et al. 1994).

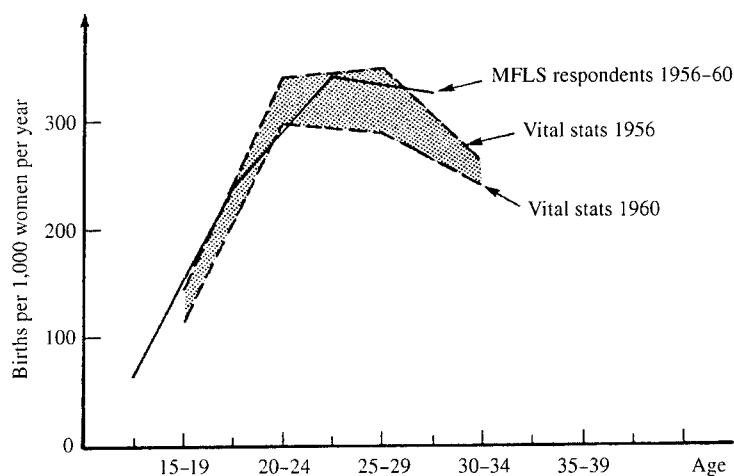


Figure 4a
Age-specific fertility rates, 1956-60—Comparison of MFLS retrospective data with contemporary vital statistics (Haaga 1986).

However, 75 percent of these children were reported to have been breastfed for longer than a week and the change in the question qualifier should not have affected these reports.

F. Telescoping

Several researchers have explored the presence of telescoping in the MFLS with mixed results. Haaga (1986) was interested in the extent to which MFLS-1 respondents recalled births as having occurred more recently than they really did. He did this by comparing the age-specific marital fertility rates implied by the MFLS-1 data for two periods, 1956-60 and 1961-65 with contemporaneously reported vital statistics for 1956, 1960, 1961, and 1965 (see Figures 4a and 4b). Over almost the entire age-range, the MFLS age-specific marital fertility rates were bounded by the vital statistic rates for the beginning and ending years of each period. Also, the age patterning of fertility rates implied by the MFLS-1 was very similar to the vital statistic rates. If women were systematically misreporting births as having occurred more recently than they actually did, MFLS-1 age-specific fertility rates should be higher than the fertility rates implied by the vital statistics in the most recent period. Sine and Peterson (1993) replicated these comparisons using data from the MFLS-2 New Sample and also concluded that reported dates of births were not telescoped. The MFLS data may be “unusually accurate for retrospective reports” of births because Malaysian citizens are required to keep identification cards on which children’s date of birth is recorded (and often these were consulted during the MFLS interview) (Haaga 1986).

In contrast, Smith and Thomas (1997) concluded that there is some telescoping

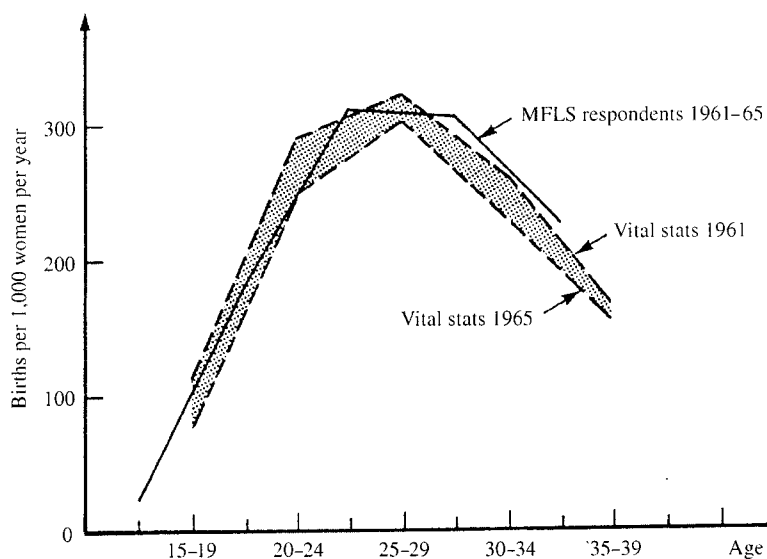


Figure 4b

Age-specific fertility rates, 1961-65—Comparison of MFLS retrospective data with contemporary vital statistics (Haaga 1986).

Reproduced with permission from the American Statistical Association.

of reported dates of moves. They considered whether, among the matched pre-1976 moves, the reported date of the move in MFLS-2 was more recent than the date reported in MFLS-1. Only reports with a complete (month and year) date were considered. Moves were divided into three groups: those where the reported dates of the moves were exactly the same (to the month and year) in MFLS-1 and MFLS-2, those where the MFLS-1 date preceded the MFLS-2 date, and moves where the MFLS-2 date preceded the MFLS-1 date. Table 8 shows that for 11 percent of all men's moves and 8 percent of all women's moves (including intra-district moves), the reported dates were exactly the same in the two reports. A slightly higher percentage of all moves fell into the category where MFLS-1 date preceded the MFLS-2 date, indicating that, on average, moves were reported as occurring more recently in MFLS-2 than they were reported in MFLS-1. The mean difference for men is six months and for women it is eight months for inter-district moves and five months for intra-district moves. To put these differences in perspective, the mean years since moves for men was 16.9 years, for women's inter-district moves 14.0 years, and for women's intra-district moves 12.0 years. Consistent with telescoping, however, is that the discrepancy in the date of the moves was significantly larger for moves when MFLS-1 date preceded the MFLS-2 date than when the MFLS-2 date preceded the MFLS-1 date.

Although the net effect of telescoping in the Smith-Thomas study is small, telescoping errors seem to be concentrated among the highly salient moves, consistent with the accessibility principle. Smith and Thomas tested the hypothesis that highly

Table 8
*Percent Distribution of Date of Reported Moves
 in Matched Sample*

	Men	Women
MFLS-1 month and year same as MFLS-2 month and year	11	8
MFLS-1 date precedes MFLS-2 date	45	48
MFLS-2 date precedes MFLS-1 date	44	44
	100	100

Based on matched moves where MFLS-1 and MFLS-2 reports have complete (month and year) date reported. Adapted from Smith and Thomas (1997).

salient (or “accessible”) memories are not only more likely to be remembered, but that they are also recalled as having occurred more recently than they really occurred. They estimated a model of time discrepancy between matched moves which included an interaction between a control variable for whether the MFLS-1 reported date of the move preceded the MFLS-2 reported data with each of three salient events—marriage and birth of a child (for females) and a job change (for males). Among men, moves that were coincident with a job change were more likely to be characterized by the date reported in MFLS-2 being more recent than the date reported in MFLS-1. Among women, moves that co-occurred with marriage were most often telescoped (that is, MFLS-1 date precedes the date reported in MFLS-2). If we assume that the MFLS-1 reports are more accurate than the MFLS-2 reports, the paradoxical conclusion from this analysis is that less salient events are forgotten more easily, but, conditional on being remembered, dates of less salient events may be more reliably reported.

We can think of two possible explanations for why telescoping was not apparent in the aggregate comparison of MFLS-1 births with contemporaneous vital statistic rates but is slightly apparent in the comparison of matched reported date of moves at two points of time. First, telescoping may be more of an issue for moves than for birth dates, possibly because of the relative salience of the two types of events: birth dates of children may be so well recalled by mothers that there is little reporting error, and hence forward telescoping bias. Second, the effects of telescoping are so subtle that telescoping is only apparent if the respondent’s reports at two points in time about the same event are compared.

G. Respondent Characteristics and Reporting Bias

In the MFLS, respondent characteristics are more strongly and consistently related to reporting bias for retrospective events across a broader range of topics than is found in the literature. For instance, Haaga (1988) estimated multivariate models predicting the likelihood of providing an inconsistent report across MFLS-1 inter-

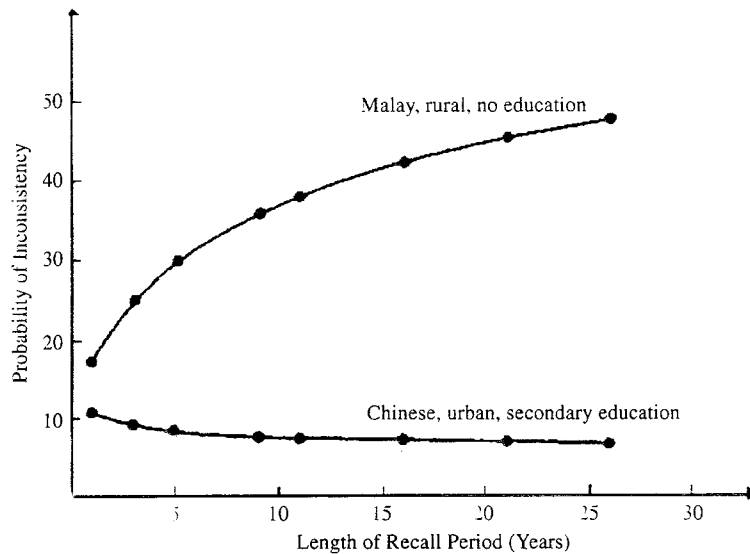


Figure 5a

Predicted probability of inconsistent reporting of first food, by length of recall period (Haaga 1988).

Reproduced with permission from the Population Association of America.

view rounds on type of first supplementary food provided to an infant and the likelihood of reporting a heaped amount for the duration of breastfeeding. The predictor variables included education, ethnic group, rural residence, and the logarithm of the length of recall period between the birth in question and date of interview. Most of the decline in data quality occurred within a 10-year recall period. There were significant interactions between length of recall period and ethnicity for inconsistent reporting of type of first food but not for heaping on breastfeeding duration. Chinese women show a very flat curve associated with inconsistent reporting of first food compared with Malay women. It may be that type of first food given to infants is a more salient topic to Chinese women, because they breastfeed for a shorter duration and hence introduce foods earlier. Chinese and Malay women have the same-shaped curve for heaping on breastfeeding duration, though the level of heaping among Chinese women is considerably lower than for Malay women. Figures 5a and 5b show the predicted probabilities of an inconsistent report on supplementary milk and heaped value for breastfeeding duration graphed against length (in years) of the recall period for two groups of respondents: Chinese urban residents with secondary education and Malay rural residents with no education. An important finding was that the same respondent characteristics (ethnicity, rural residence, and no education) were associated with both types of inaccurate data.

Other studies based on MFLS data indicate that these same respondent characteristics are predictive of other data quality problems (summarized in Table 9). In a multivariate model, less educated women, and ethnic Malays and Indians were less

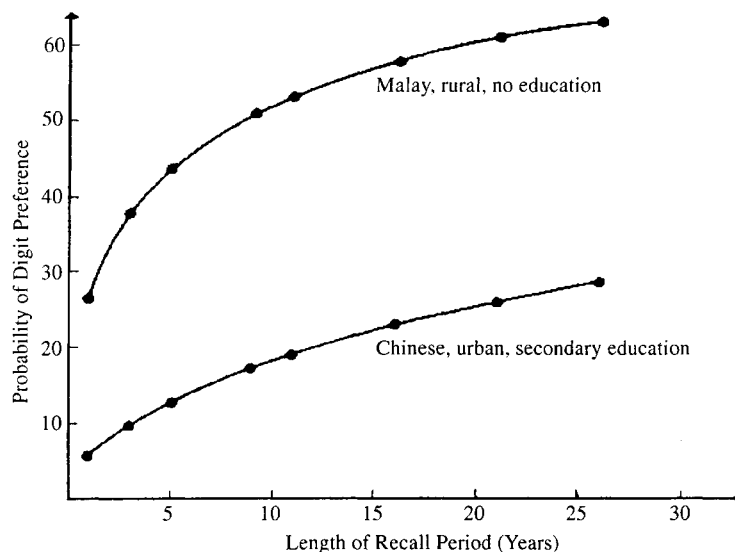


Figure 5b

Predicted probability of digit preference in breastfeeding duration, by length of recall period (Haaga 1988).

Reproduced with permission from the Population Association of America.

likely to provide complete or exact birth dates of children (that is, where both month and year are specified) (Haaga 1986). Approximate (rather than exact) birthweights were more likely to be reported for births to mothers who were less educated, resided in rural areas, and were of Malay ethnicity (DaVanzo, Habicht, and Butz 1984). Not only are these characteristics associated with quality of reports about birth and infant-related outcomes, but they also appear as correlates of consistent reporting of migration events. Education is a strong and significant correlate of reporting a move in both MFLS-1 and MFLS-2 and is associated with fewer inconsistencies in dating of moves (Smith and Thomas 1997). Smith and Thomas also found that the interviewer's assessment of the reliability of a respondent's answers to items in the migration module was predictive of consistent reporting. Although there were no ethnic differences in consistent reporting of moves among men, among women, Indians and Chinese provided more consistent reports on their moves than Malays (Smith and Thomas 1997). Panis and Lillard (1994) found that the most educated women were more likely to report miscarriages, probably reflecting underreporting by less educated women. Finally, education and the interviewer's subjective impression of the respondent's quality of reporting were negatively related to the residual variance in a model of earnings (Lillard and Kilburn 1995).

We can think of two reasons why we find strong and consistent predictors of poorer data quality across topics in contrast to much of the previous literature. First, with the exception of the two studies based on the NHEFS, previous work examined data quality only for short-term retrospectives. It may be that the extent of data

Table 9
Summary of Multivariate Correlates of Reporting Error in MFLS, by Study and Outcome Variable

Study	Outcome Variable	Respondent characteristic correlated with outcome			
		Interviewer assessment	Education	Urban/rural residence	Ethnicity
Haaga 1988	Inconsistent report on supplementary food		X	X	X
	Heaping on breastfeeding duration		X	X	X
Kuate Defo and DaVanzo 1996	Reasons for no or short breastfeeding		X	X	X
Haaga 1986	Complete/exact birth date		X		X
DaVanzo et al. 1984	Exact (versus approximate) birthweight		X	X	X
Smith and Thomas 1997	Consistent dating of moves	X	X		X
Lillard and Kilburn 1995	Variance of error term associated with earnings	X	X		

quality problems in short-term retrospectives is not large enough to be consistently detectable. Second, socioeconomic indicators may be particularly important predictors of accuracy of recall data in situations where the population's education levels are more variable.

V. Summary of Findings and Implications for Data Collection

We now summarize data quality patterns that emerge across analyses of the MFLS and discuss their implications for data collection.

1. Data Quality Deteriorates with the Length of the Recall Period

Consistent with research based on shorter retrospective histories, we find evidence that while overall the quality of long-term MFLS retrospective reports is high, some systematic problems emerge as the length of the recall period increases. Some events are forgotten and details about events are blurred (resulting in, for example, incomplete or inexact reports).

2. The More Salient an Event is to the Respondent, the More Accurate its Report

Among men and women, moves that are associated with highly salient life events (for example, new job for men, marriage for women) are more likely to be consistently reported 12 years apart.

3. Certain Events May be Subject to Telescoping

There is very slight forward telescoping of moves in MFLS-2 compared with MFLS-1 reports on the same moves. There is no evidence of telescoping of births.

4. Events that are Socially Undesirable are More Poorly Reported

Previous work has focused especially on under-reporting of especially socially sensitive topics like abortion and illegal drug use. Analyses based on the MFLS suggest that reports about less socially sensitive behaviors, like breastfeeding, may also be responsive to social norms.

5. Socioeconomic Characteristics Predict Quality of Retrospective Reports

An important contribution of the MFLS analyses to the literature on data quality is the recognition that a correlate of data quality (for example, education, ethnicity) in retrospective reports on one topic (for example, details about a birth) is likely to be predictive of quality issues related to other topics (for example, migration histories) and that interviewers' impressions of the quality of respondents' answer correlate with quantitative measures of the quality of the data.

A. Implications for Data Collection

The findings of our review of studies examining data quality in the MFLS lead to some suggestions for survey design that may help mitigate recall error in retrospective surveys.

1. Provide Benchmarking (Preloading Information)

In panel surveys, respondents' recall may be improved by reminding them of their situation at the time of the previous wave and bringing them forward from that point. This also reduces interview time, relative to administration of a complete retrospective questionnaire because fewer events need to be reported. Many large, national surveys, including the Panel Survey on Income Dynamics and HRS take advantage of computer technology to incorporate benchmarking (or preloading responses from prior interviews) in reinterviews.

A caveat applies if the question involves some concept, such as a health condition, about which the general public becomes more educated about with time. For example, a respondent may have indicated lack of hypertension in the baseline, only to realize later that he or she did in fact suffer from the condition at that time but did not know its proper term. The survey designer has to allow for instances where this could be the case. Also, benchmarking is not costless; the interviewer has to have access to earlier datafiles and has to have this information readily available when revisiting the respondent. Another caveat is that many retrospective histories begin in the present and work back in time; for benchmarking to work easily, the retrospective history would have to begin at the last interview and work forwards in time.

A final caveat is that the exclusive use of benchmarking in follow-up interviews precludes the opportunity to identify reporting error by comparing overlapping retrospectives. Moreover, it is incorrect to assume that initial or earlier reports are always the correct ones. There is some evidence to suggest that people may be more likely to report sensitive events (for example, use of alcohol) when the event occurred further in the past. For methodological purposes, it would be valuable to collect replicate life histories for a subsample so that questions that elicit the greatest inconsistent reporting can be identified, as can respondents who provide inconsistent reporting. Replicate reports provide important insights into ways to improve data collection efforts (such as question wording) and flag variables that are particularly prone to reporting bias. We also recommend that more "gold standards" be sought as a means of further evaluating the quality of data. For retrospective surveys like the MFLS, possible "gold standards" include school records or clinic records on contraceptive use, prenatal care, or children's immunizations.

2. Ask Alternative Questions if Respondent is Unable to Provide a Precise Response

Data quality for events in the more distant past may be improved by asking alternative questions when necessary. For example, if a person does not remember an exact date, he or she may be able to provide the age at the time of the event or the time of the year (early, middle, late) it occurred. Other examples of follow-up questions:

in the MFLS include asking for an approximate birthweight when an exact birthweight could not be recalled. When this approach was used in the MFLS (for the birthweight and birth date), it always elicited a response. An alternative method that has been used successfully in HRS is to obtain bracketed information when an exact response is not provided. The survey designer needs to confirm that the distributions of the approximate responses are consistent with that of the exact responses or with external data sources. DaVanzo, Habicht, and Butz (1984) did this for approximate responses for birthweights, which were related to child survival in the expected way.

3. Use a Calendar

Responses may improve by using a calendar on which major national events (key holidays, national events, and natural disasters) are pre-printed and on which salient life events are recorded. In particular, this may improve internal consistency and sequencing of events. For example, did the move take part before or after the marriage? Did the discontinuation of contraception occur before or after the death of a child? Did the respondent re-enter the workforce before or after the onset of the economic crisis? Use of a calendar has been shown to improve data quality (Becker and Sosa 1997).

4. First Ask Summary Questions

Details on sensitive areas may be better retrieved if the questions are preceded by one or more summary questions. For example, MFLS-2 included "Brass" questions that inquired about the total number of pregnancies and living and deceased children. The interviewer then followed up with detailed questions about their outcomes, pregnancy by pregnancy, and prompted the respondent if the totals did not match. Peterson (1993) found newly reported births in the MFLS-2 for the period prior to the MFLS-1 interview date, which she concludes, may have resulted from the use of the Brass questions in MFLS-2.

5. Train Interviewers to Prompt

A well-trained interviewer may be able to elicit information on underreported events. MFLS interviewers, for example, were instructed to inquire about a miscarriage, induced abortion, missed episodes of contraceptive use, or spousal separation if the period between two births was more than four years. For reported durations or amounts that may be subject to heaping at critical values, probing may help correct classification. For example, MFLS interviewers were instructed to probe if a duration was reported as "about one year" or "about six months" for whether the duration was more likely to have been shorter or longer than the reported value. In MFLS-2, a probe was used to try to determine how and when each pregnancy ended to increase reporting of miscarriages and induced abortions. If the calendar method is not used, interviewers should be trained to cross-reference events in different life areas (as the MFLS did).

6. Vary Recall Period with Saliency

Since more salient life events are remembered better, one may restrict the recall period for common, nonsalient events, as Radloff (1983) recommends for short-distance, return migration events. Radloff recommended that short-distance and short-term migration be collected for the preceding twelve months only, while more salient moves, such as noncircular moves or those involving crossing an administrative boundary, be collected for longer recall periods.

VI. Conclusion

Although the theoretical and qualitative work on recall error has identified a number of likely problems with retrospective reports, few national surveys lend themselves to empirical measurement of errors in retrospective reports. Ideal assessment of retrospective reports requires the availability of a "gold standard" against which to compare retrospective reports or replicate reports from the same respondent about the same event at the same point in time. A "gold standard" is rarely available, and if available, very costly and time-consuming to link with survey reports. Replicate reports are rare in national panel studies because most survey designers use benchmarking to minimize the possibility that respondents will contradict earlier reports. The Malaysian Family Life Surveys (MFLS) provide a unique opportunity to compare replicate retrospective reports for events that in some cases occurred several decades earlier.

We find that the quality of the long retrospective histories in the MFLS is quite high, across a range of topics. Quality is highest at the aggregate, somewhat lower when reports provided by the same respondent about the same event from two waves are matched. We find the most consistent degradation in data quality with respect to details about events rather than the occurrence of the events themselves. Our review of existing studies and the analyses we present here suggests that long-term retrospective histories provide nearly as good quality reports as provided by short-term retrospective histories as indicated in the literature. Nevertheless, researchers should pay close attention to potentially systematic misreporting of events and their associated details during the data collection.

References

- Arnold, Fred. 1990. "Assessment of the Quality of Birth History Data in the Demographic and Health Surveys." In *An Assessment of DHS-I Data Quality*, DHS Methodological Reports, No. 1. Columbia, Maryland: Institute for Resource Development/Macro Systems, Inc.
- Becker, Stan, and Doris Sosa. 1997. "An Experiment Using a Month-by-Month Calendar in a Family Planning Survey in Costa Rica." *Studies in Family Planning* 23(6):386-91.
- Beckett, Megan, Julie DaVanzo, Narayan Sastry, Constantijn Panis, Christine Peterson. 1998. "The Quality of Retrospective Reports in the Malaysian Family Life Survey: Implications for Survey Design and Data Analysis." presented at the Conference on Data

- Quality Issues in Longitudinal Surveys at the Institute of Social Research. Michigan: University of Michigan, October 1998.
- Beckett, Megan, Maxine Weinstein, Noreen Goldman, and Lin Yu-Hsuan. 2000. "Do Health Interview Surveys Yield Reliable Data on Chronic Illness among Older Respondents." *American Journal of Epidemiology* 151:315–23.
- Bowman, Jennifer Ann, Rob Sanson-Fisher, and Sally Redman. 1997. "The Accuracy of Self-Reported Pap Smear Utilization." *Social Science and Medicine* 44:969–76.
- Brown, N. R., L. J. Rips, and S.K. Shevell. 1986. "The Subjective Dates of Natural Events in Very Long-Term Memory." *Cognitive Psychology* 17:139–77.
- Collins, L. M., J. W. Graham, W. B. Hansen, and C. A. Johnson. 1985. "Agreement between Retrospective Accounts of Substance Use and Earlier Reported Substance Use." *Applied Psychological Measurement* 9:301–309.
- DaVanzo, Julie, and William P. Butz. 1978. "The Malaysian Family Life Survey." Summary Report R-2351-AID. Santa Monica: RAND.
- DaVanzo, Julie, William P. Butz, D. Z. Fernandez, R. Jones, and N. Spoelstra. 1978. "The Malaysian Family Life Survey: Appendix A, Questionnaires, and Interviewer Instructions." Report R-2351/1-AID. Santa Monica: RAND.
- DaVanzo, Julie, Jean-Pierre Habicht, and William P. Butz. 1984. "Assessing Socioeconomic Correlates of Birthweight in Peninsular Malaysia: Ethnic Differences and Changes Over Time." *Social Science and Medicine* 18(5):384–404.
- DaVanzo, Julie, Jeffrey Sine, Christine Peterson, and John Haaga. 1994. "Reversal of the Decline in Breastfeeding in Peninsular Malaysia? Ethnic and Educational Differentials and Data Quality Issues." *Social Biology* 41(1–2):61–77.
- DaVanzo, Julie, John G. Haaga, Tey Nai Peng, Ellen H. Starbird, Christine E. Peterson. 1993. "The Second Malaysian Family Life Survey: Survey Instruments." Report MR-107-NICHD/NIA. Santa Monica, RAND.
- Diamond, I., I. McDonald, and I. Shah. 1986. "Proportional Hazards Models for Current Status Data: Application in the Study of Differentials in Age at Weaning in Pakistan." *Demography* 23:607–20.
- Dodge, R. 1970. "Victim Recall Pretest." unpublished memorandum, Washington D.C.: U.S. Bureau of the Census. [Cited in R. Groves 1989]
- Dufour, M., J. Colliver, M. B. Grigson, F. Stinson. 1990. "Use of Alcohol and Tobacco." In *Health Status and Well-Being of the Elderly: National Health and Nutrition Examination Survey-I Epidemiologic Follow-Up Study*, eds. J. C. Coroni-Huntley, R. R. Huntley, & J. J. Feldman. New York: Oxford University Press.
- Ebbinghaus, H. 1964. "Memory: A Contribution to Experimental Psychology." New York: Russell Sage Foundation.
- Fu, H., J. E. Darroch, S. K. Henshaw, and E. Kolb. 1998. "Measuring the Extent of Abortion Underreporting in the 1995 National Survey of Family Growth." *Family Planning Perspectives*, 30(3):128–33, 138.
- Groves, R. 1989. "Survey Costs and Survey Errors." New York: John Wiley and Sons.
- Haaga, John G. 1986. "The Accuracy of Retrospective Data from the Malaysian Family Life Survey." Report N-2157-AID. Santa Monica: RAND.
- . 1988. "Reliability of Retrospective Survey Data on Infant Feeding." *Demography* 25(2):307–14.
- Hahn, Robert A., Elaine Eaker, and Henry Rolka. 1997. "Reliability of Reported Age at Menopause." *American Journal of Epidemiology* 146:771–75.
- Herzog, A. R., and W. L. Rodgers. 1989. "Age Differences in Memory Performance and Memory Ratings in a Sample Survey." *Psychology of Aging* 4:173–82.
- Huttenlocher, J., L. V. Hedges, and Norman M. Bradburn. 1990. "Reports of Elapsed Time: Bounding and Rounding Processes in Estimation." *Journal of Experimental Psychology, Learning, Memory, and Cognition* 16:196–213.

- Klerman, Jacob A. 1995. "Insights into Heaping from Retrospective Breastfeeding Data." In *Proceedings, International Conference on Survey Measurement and Process Quality*. Bristol, England, April 1995.
- Knodel, J., and N. Debavalya. 1980. "Breastfeeding in Thailand: Trends and Differentials, 1969-1979." *Studies in Family Planning* 11:355-77.
- Kuate Defo, Bartolomey, and Julie DaVanzo. 1996. "Reliability of Reports of Reasons for No or Short Breastfeeding." Report DRU-1305-NICHD, Labor and Population Program Working Paper 96-03. Santa Monica: RAND.
- Landis, J., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33:159-74.
- Lillard, Lee A., and Rebecca Kilburn. 1995. "Intergenerational Earnings Links: Sons and Daughters." Report DRU-1125-NIA, Labor and Population Program Working Paper Series 95-17. Santa Monica: RAND.
- Marquis, K. H., C. F. Cannell, and A. Laurent. 1972. *Reporting Health Events in Household Interviews: Effects of Reinforcement, Question Length, and Reinterviews*. Rockville, Md: National Center for Health Statistics. (Vital and Health Statistics, Series 2: Data Evaluation and Methods Research, no. 45) (DHEW publication no. (HSM) 72-1028).
- Mathiowetz, Nancy A. 1999. "Measurement Error in Surveys of the Low Income Population." Presented at the Workshop on Data Collection for Low Income and Welfare Populations, National Academy of Sciences, November 1999.
- Panis, Constantijn W. A., and Lee A. Lillard. 1994. "Health Inputs and Child Mortality: Malaysia." *Journal of Health Economics* 13:455-489.
- Peterson, Christine E. 1993. "Long Term Recall of Fertility-Related Events: The MFLS Experience." Paper presented at the Annual Meeting of the Population Association of America, Cincinnati, Oh.
- Radloff, Scott R. 1983. "Detecting Migration: An Exploration of Measurement Issues Using the Malaysian Family Life Survey." Report N-1927-AID. Santa Monica, RAND.
- Rubin, D. C., and A. D. Baddeley. 1989. "Telescoping Is Not Time Compression: A Model of the Dating of Autobiographical Events." *Memory and Cognition* 17:653-61.
- Sine, Jeffrey. 1993. "Data Quality in Retrospective Surveys." In *Proceedings of the Seminar on the Second Malaysian Family Life Survey*. Kuala Lumpur, Malaysia, October 1991. Report CF-109-NICHD/NIA/WFHF. Santa Monica: RAND.
- Sine, Jeffrey, and Christine E. Peterson. 1993. "The Second Malaysian Family Life Survey: Quality of Retrospective Data for the MFLS-2 New Sample." Report MR-110-NICHD. Santa Monica: RAND.
- Smith, James P. and Duncan Thomas. 1997. "Migration in Retrospect: Remembrances of Things Past." Report DRU-1628-NICHD. Santa Monica: RAND.
- Strack, F., and L. L. Martin. 1987. "Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys." In *Social Information Processing and Survey Methodology*, eds. H. J. Hippler, Norbert Schwarz, and Seymour Sudman, 123-48. 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, CA: Jossey-Bass Publishers.
- Thompson, C. P., J. J. Skowronski, and D. J. Lee. 1988. "Telescoping in Dating Naturally Occurring Events." *Memory and Cognition* 16:461-68.
- Tourangeau, R. 1984. "Cognitive Science and Survey Methods: A Cognitive Perspective." In *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, ed. T. Jabine, M. Straf, J. Tanur, and R. Tourangeau, 73-100. Washington, D.C.: National Academy Press.
- Tversky, A., and D. Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5:207-32.

- Upchurch, Dawn M., Lee A. Lillard, Carol S. Aneshensel, and Fang Li. 2000. "Inconsistencies in Reporting the Occurrence and Timing of Sexual Initiation among Adolescents." Working Paper. Westwood, CA: UCLA School of Public Health.
- Warnecke, Richard B., Seymour Sudman, Timothy P. Johnson, Diane O'Rourke, Andrew M. Davis, and Jared B. Jobe. 1997. "Cognitive Aspects of Recalling and Reporting Health-related Events: Papanicolaou Smears, Clinical Breast Examinations, and Mammograms." *American Journal of Epidemiology* 146:982-92.
- Withey, S. B. 1954. "Reliability of Recall of Income." *Public Opinion Quarterly* 18:31-34.
- Wu, Lawrence L., Steven P. Martin, and Daniel A. Long. 1998. "Comparing Data Quality of Fertility and First Sexual Intercourse Histories." Paper presented at the conference on Data Quality in Longitudinal Surveys, Institute for Social Research. Michigan: University of Michigan.

RAND
1700 Main Street, Santa Monica, CA 90407
Labor and Population Program Reprint Series

2001

Series Number	Order Number	Title	Author
01-07	RP-964	<i>Bargaining Power Within Couples and Use of Prenatal and Delivery Care in Indonesia</i> [In <i>Studies in Family Planning</i> , Vol. 32, No. 2, June 2001, pp. 130-146]	Kathleen Beegle Elizabeth Frankenberg Duncan Thomas
01-06	RP-966	<i>Comments on: Explaining Recent Declines in Food Stamp Program Participation</i> [In <i>Brookings – Wharton Papers on Urban Affairs</i> 2001, W.G. Gale and J.R. Pack (eds.), 2001, pp. 236-241]	Robert F. Schoeni
01-05	RP-956	<i>The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country</i> [In <i>The Journal of Human Resources</i> , Vol. 36, No. 3, 2001, pp. 593-625]	Megan Beckett Julie DaVanzo Narayan Sastry Constantijn Panis Christine Peterson
01-04	RP-953	<i>Why Is Wealth Inequality Rising?</i> [In <i>The Causes and Consequences of Increasing Inequality</i> , Finis Welch (ed), 2001, pp. 83-115]	James P. Smith
01-03	RP-945	<i>Commentary: Methodological Biases in Estimating the Burden of Out-of-Pocket Expenses</i> [In <i>Health Services Research</i> , Vol. 35(6), 2001, pp. 1357-1365]	Dana P. Goldman James P. Smith
01-02	RP-936	<i>Race and Ethnicity in the Labor Market: Trends Over the Short and Long Term</i> [In <i>America Becoming: Racial Trends and Their Consequences</i> , Vol. II, Neil J. Smelser, et al. (eds), 2001, pp. 52-97]	James P. Smith
01-01	RP-923	<i>The Importance of International Demographic Research for the United States</i> [In <i>Population Research and Policy Review</i> , Vol. 19, 2000, pp. 199-232]	Narayan Sastry

2000

Series Number	Order Number	Title	Author
00-08	RP-924	<i>Forecasting Enrollments for Immigrant Entry-port School Districts</i> [In <i>Demography</i> , Vol. 37, No. 4, November 2000, pp. 499-510]	Peter A. Morrison
00-07	RP-916	<i>Trends in Long-run Versus Cross-section Earnings Inequality in the 1970s and 1980s</i> [In <i>Review of Income and Wealth</i> , Series 44, No. 2, June 1998, pp. 199-213]	Jeremy Arkes
00-06	RP-884	<i>Converging Health Inequalities in Later Life—an Artifact of Mortality Selection?</i> [In <i>Journal of Health and Social Behavior</i> , Vol. 41, March 2000, pp. 106-119]	Megan Beckett
00-05	RP-882	<i>Do Health Interview Surveys Yield Reliable Data on Chronic Illness Among Older Respondents</i> [In <i>Journal of Epidemiology</i> , Vol. 151, No. 3, 2000, pp. 315-323]	Megan Beckett Maxine Weinstein Noreen Goldman Lin Yu-Hsuan
00-04	RP-891	<i>Wives' Earnings and the Level and Distribution of Married Couples' Earnings in Developed Countries</i> [In <i>Journal of Income Distribution</i> , Vol. 8, No. 1, 1998, pp. 45-61]	Maria Cancian Robert F. Schoeni
00-03	RP-881	<i>Does Head Start Help Hispanic Children?</i> [In <i>Journal of Public Economics</i> , Vol. 74, No. 2, 1999, pp. 235-262]	Janet Currie Duncan Thomas
00-02	RP-804	<i>Job Continuity Among New Mothers</i> [In <i>Demography</i> , Vol. 36, No. 2, May 1999, pp. 145-155]	Jacob Alex Klerman Arleen Leibowitz
00-01	RP 887	<i>The Role of Education in Explaining and Forecasting Trends in Functional Limitations Among Older Americans</i> [In <i>Demography</i> , Vol. 36, No. 4, November 1999, pp. 461-473]	Vicki A. Freedman Linda G. Martin

1999

Series Number	Order Number	Title	Author
99-10	RP-865	<i>The Intergenerational Transmission of "Intelligence": Down the Slippery Slopes of 'The Bell Curve'</i> [In <i>Industrial Relations</i> , Vol. 38, No. 3., 1999, pp. 297-330]	Janet Currie Duncan Thomas
99-09	RP-843	<i>Parent-Child Coresidence and Quasi-Coresidence in Peninsular Malaysia</i> [In <i>Southeast Asian Journal of Social Science</i> , Vol. 27, No. 2., 1999, pp. 43-64]	ChorSwang Ngin Julie DaVanzo
99-08	RP-833	<i>The Measurement and Structure of Household Wealth</i> [In <i>Labour Economics</i> , Vol. 6, 1999, pp. 253-273]	F. Thomas Juster James P. Smith Frank Stafford
99-07	RP-800	<i>US Abortion Policy and Fertility</i> [In <i>The American Economic Review</i> , Vol. 89, No. 2, May 1999, pp. 261-264]	Jacob Alex Klerman
99-06	RP-802	<i>Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status</i> [In <i>Journal of Economic Perspectives</i> , Vol. 13, No. 2, Spring 1999, pp. 145-166]	James P. Smith
99-05	RP-792	<i>Implications of Population Aging for Geriatric Health</i> [In <i>Reichel's Care of the Elderly: Clinical Aspects of Aging</i> , Gallo, J.J., et al (eds.), Philadelphia, pp. 659-669, 1999]	Emily M. Agree Vicki A. Freedman
99-04	RP-788	<i>The Associations between Self-rated Vision and Hearing and Functional Status in Middle Age</i> [In <i>Ophthalmology</i> , Vol. 106, No. 2, February 1999, pp. 401-405]	Paul P. Lee James P. Smith Raynard S. Kington
99-03	RP-779	<i>The Relationship Between Infant and Child Mortality and Subsequent Fertility in Indonesia: 1971-1991</i> [In <i>From Death to Birth: Mortality Decline and Reproductive Change</i> , 1998, pp. 316-338]	Elizabeth Frankenberg
99-02	RP 764	<i>Long-Term Admissions to Home Health Agencies: A Life Table Analysis</i> [In <i>The Gerontologist</i> , Vol. 39, No. 1, February 1999, pp. 16-24]	Vicki A. Freedman
99-01	RP 757	<i>Fertility, Education, and Resources in South Africa</i> [In <i>Critical Perspectives on Schooling and Fertility in the Developing World</i> , 1999, pp. 138-180]	Duncan Thomas

To order a reprint, free of charge, please call RAND Distribution Services at 310.451.7002.
For reprints prior to 1999, please see our web page: <http://www.rand.org/labor/>

RP-956